The Audio Pages

Elliott Sound Products — Audio Articles

Page Last Updated - 11 May 2006

Feature Articles	Updated
Active Vs. Passive Crossovers - An explanation for the profound audible differences	11 Jan 04
Benefits of Bi-Amplification - Not quite magic, but close (Part I)	02 May 05
Benefits of Bi-Amplification - Crossovers, tri-amping, etc. (Part II)	02 May 05
Counterfeit (Fake!) Semiconductors - Don't *you* be caught with fake, below spec transistors when you build a project!	21 Feb 06
Education - Customised projects for schools, technical colleges and universities	30 Apr 03
The State of Manufacturing - This article should be a mandatory read for all businessmen and politicians, a real wake-up call !	03 Sep 02
Phase Angle Vs. Transistor Dissipation - A simplified approach to understanding Safe Operating Area	19 Mar 05
Sound Impairment Monitor - Is This The Answer? - A new method to monitor distortion in real time	23 Sep 00
Transistor Safe Operating Area - What is SOA, and how does it affect amplifier designs	14 Aug 03

Cables	
The Truth About Cables, Interconnects and Audio in General - A "must read" article	20 Sep 02
More Truth About Cables - A white paper discussing speaker cables and interconnects *	22 Jul 01
Cable Impedance - Characteristic impedance, high capacitance designs and amp stability	17 Oct 03

Amplification	

Amplifier Sound - What causes the perceived differences between amps? Is it real or imaginary? *09 Apr 00*

Class- A - a discussion of the benefits of Class-A amplifiers *02 Apr 05*

Clipping Behaviour - One of the lesser known effects of amplifier clipping, and the damage it can do. *15 Oct 04*

Distortion & Negative Feedback - A detailed look at the way negative feedback reduces distortion new *06 May 06*

Musical Instrument Amplifiers - Guitar, Bass and Keyboard *04 Feb 01*

Short Circuit Protection - Testing amplifiers to the limits * *22 Aug 00*

Soft Clipping - Is adding a soft-clip circuit worth the effort? *15 Apr 06*

The Need For Speed - Very high speed opamps from Texas Instructions *17 Aug 01*

PWM Amplifiers - Theory and design concepts behind these 'new' amplifiers *25 Jun 05*

Valve Amplifiers - do they really sound different - Includes a review of one of my valve amps *28 Nov 99*

General Information

A-B Switch Box for Amplifier Comparisons - Something to offend everyone, or a great tool? * *15 Aug 00*

ABX Comparator - A true ABX system that does not need a PC or microcontroller * *11 Aug 02*

AM Radio - Approaches to high fidelity AM broadcast band reception * *18 Feb 06*

Capacitors - What are the influences that determine the so-called 'sound' of capacitors. Covers dielectrics, dissipation factor, etc. *25 Oct 05*

CD, SACD and DVD-A -A Comparison of the formats in a discussion between Niklas Ladberg and Ing. Öhman (*very* interesting) * *28 Aug 02*

Compression In Audio - Ever wondered why some music sounds flat and lifeless, even though it's loud? *21 Dec 01*

Earthing Your Hi-Fi - Tricks and Techniques - How to eliminate hum without affecting safety *30 Dec 99*

Build Your Own Heatsink? - Yes, it can be done. If you have the tools, you can make a heatsink as big as you'll ever want. *11 Feb 06*

Impedance - What does it really mean? A general discussion on impedance, damping, etc. *24 Jan 05*

Microphones - An introduction to the different types, polar response and behaviour of microphones *12 Jan 06*

Why *Do* Tweeters Blow? - The real reason that tweeters (and other drivers) are destroyed by distortion	*28 Jul 01*
What is Hi-Fi - A contributed article on some of the subtle aspects of Hi-Fi *	*05 May 01*
What is Hi-Fi (Part II) - Continuation of 'What is Hi-Fi' explaining the idea of accuracy (amongst other things)	*29 Mar 04*

Beginners' Luck - The beginners' Guide to ...

Amplifier Basics - How Amps Work - How amplifiers do what they do	*06 Apr 05*
Amplifier Efficiency - Power Loss, DC voltages and current explained	*29 Jan 00*
Amplifier Troubleshooting and Repair Guide If it's broken, you may be able to fix it. This article is designed for those new to repairs	*27 Apr 03*
Bridging - How the "Trimode" bridging system works in car power amplifiers	*20 Oct 02*
Bulletin Boards - A guide to etiquette when posting at your favourite forum *	*09 Apr 01*
Designing With Opamps I - What they are, what they do, and how they do it	*03 May 00*
Designing With Opamps II - Audio applications, filters, crossovers, etc.	*08 May 00*
Designing With Opamps III - More Filters, Specifications, Comparators, etc. new	*11 May 06*
Electronics - Part 1 Passive components - resistors, capacitors and inductors explained	13 Mar 01
Electronics - Part 2 The resistor 'E' series (e.g. E12, E24, etc), Zener Diodes	05 Mar 04
Meters, Multipliers & Shunts - How to use analogue and digital panel meters new	*06 May 06*
Potentiometers. - what they are and how to use them in practical circuits.	*22 Jan 03*
Power - An explanation of the derivation of PMPO in amplifiers. Light reading, really	*19 Jun 00*
Power Supply Wiring There are a lot of considerations for the wiring of a PSU - read about it here	*07 Jun 03*
Tools - The first in a series on mechanical aspects of your hobby *	*21 Mar 01*
Transformers - Part 1 The basics of transformers, what they do, and how they do it.	*05 May 01*
Transformers - Part 2 The details of magnetic circuits, windings, core types and losses	*03 Apr 01*
Transformers - Part 3 Designing and analysing transformers - not at all for the faint-hearted, but all very useful information.	*16 Jan 06*

Voltage Dividers & Attenuators - A short article describing how to design these essential circuits	*21 Dec 02*

Design Guidelines

Amplifier Design - ESP's contribution to the everlasting debate	*09 Jun 01*
Amplifier Design With HEXFETs - Getting good performance from vertical MOSFETs *is* possible (Updated)	*13 Mar 06*
Balanced Driver with Floating Output - An in depth discussion of the design process *	*14 Apr 02*
Current Sources, Sinks & Mirrors - A discussion on these much misunderstood circuits	*19 Feb 01*
Heatsinks - An article on heatsink selection, transistor mounting and thermal transfer principles	*17 Sep 03*
Noise in Audio Amplifiers - An explanation of noise figures, and how circuit noise is calculated	*29 Jan 00*
Opamp Alternatives - A sample of the circuits that were in common use before opamps 'took over'	*28 Apr 05*
Power Supply Design - How to design linear power supplies - transformers, rectifiers and filters	*03 Mar 01*
Transistor Matching Tests - Matching Bipolars or MOSFETs for critical amplifier applications	*08 Dec 01*
Variable Impedance Amplifiers - How you can control the output impedance of a power amp	*22 Apr 00*

Loudspeaker and Crossover Design

Baffle Step Compensation - What it is and why we need to compensate for it	*08 Dec 01*
Compliance Scaling - Fitting just about any driver to just about any alignment	*30 Jun 05*
Derived (Subtractive) Crossovers - An examination of these networks, and why they are unsuitable for serious use	*21 Sep 05*
Doppler Distortion in Loudspeakers - Is the effect real or imaginary when applied to loudspeakers? The results may surprise you	*23 Aug 04*
Impedance - Effects Of Source Impedance on Loudspeaker Drivers	*22 Jul 01*
The Linkwitz Transform Circuit - What it is, what it does and how it does it *	*16 Sep 02*
NTM™ Crossovers - This is a topic that has virtually zero technical information on the Net ... until now.	*15 Sep 05*

	Loudspeakers (Part I) - see the production of my latest triamped speaker boxes unfold	*14 Jul 01*
	Loudspeakers (Part II) - The electronics that drive the 3-way system are described	*18 Nov 01*
	Passive Crossovers - The Design of High Quality Passive Crossover Networks	*05 Nov 01*
	Phase, Time and Distortion in Loudspeakers - Is "Time Alignment" worth the effort?	*30 Jul 02*
	Serial vs. Parallel Crossover Networks - Are there any *real* differences?	*14 Aug 03*
	Phase Correction - Myth or Magic? - Do all-pass filters really work to correct loudspeaker phase anomalies?	*12 Mar 04*
	Thiele Small Parameters - Determining the Thiele/Small parameters for a loudspeaker driver (Updated)	*12 Jan 06*
	Satellites & Subwoofers - Description of the little known QB5 loudspeaker alignment. Maximum SPL from 2-way bookshelf speakers. *	*28 Aug 04*
	Subwoofer Conundrum - Some tips and hints on subwoofer selection, efficiency and placement	*12 Mar 04*

Explanation of Symbols ...

new	The article is new - that must come as a surprise Less than 2 months old (typically).
RED	Dates in the update column indicate a new or updated (within 2 months) article.
	Indicates an article that may include humour - caution is advised
*	Contributed article (see copyright notice on article page)

Main Index

Please refer to Copyright information in the individual articles. All articles are Copyright © 1998 ... 2006 Rod Elliott, unless otherwise stated.

My Thanks

I must express my great appreciation to Geoff Moss from the UK, who has provided me with information about spelling errors, inconsistencies, and general errors in many of these articles. Geoff has done the thing that I cannot do - looked at the articles with detachment, and has spotted and brought to my attention several howling blunders as well as the general editorial help.

Geoff, my sincere thanks and appreciation for what has been many hours of work. This

short acknowledgement is scant reward, but I really appreciate the help and assistance.

In addition, I would like to thank the many readers who have pointed out various additional errors and omissions - Your help in ensuring that the information on these pages is as accurate as possible is appreciated, *always!*

Although it goes without saying (but I will say it anyway :-) the contributed articles and projects from all over the world have added considerably to the overall value and range of topics covered in these pages, and on behalf of all the readers of The Audio Pages, I thank you for you efforts. Writing a clear and descriptive technical article is not an easy task, but your perseverance is acknowledged, and I am grateful that you have taken the time to share your knowledge.

May the flow of input continue - for anyone who would like to contribute, please do not hesitate to contact me.

Upcoming Articles

At any given time, there are usually articles in production either by myself or contributors. It is no longer possible (or sensible) to leave hints as to what may or may not be coming, since it is so easy to get sidetracked by something new that comes along. Feel free to offer suggestions, or offer your own contribution.

Total Visitors since Jan 2001 - 5322194

Last significant page update 11 Feb 2006 - rearranged each section into alphabetical order, added DIY heatsinks

The Audio Pages

Elliott Sound Products **Active Vs. Passive Crossovers**

Copyright © 2004 - Rod Elliott (ESP)
Page Created 11 Jan 2004

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 How Many Ohms?
  - 2.1 The Crossover - a Different View
  - 2.2 The Active Solution
  - 2.3 Other Network Orders
  - 2.3 Zobel (and Other) Networks
- 3.0 Conclusion

1.0 - Introduction

In the article Benefits of Biamping, I discussed the many advantages that are to be had by using separate amplifiers for bass and mid+high. There is also a section devoted to tri-amping (for a typical 3-way system). Essentially, the ideal arrangement is to use a separate amplifier for each loudspeaker driver in the system. Although there are still many who consider this to be overkill, the advantages are so compelling that there is no reason not to adopt this approach as a matter of course.

Of course, if the speaker arrangement uses two drivers in parallel (for example the well known MTM or D'Appolito topology), a single amp may drive both mid-woofers - dual amplifiers will usually not give any major benefit in this setup.

One area of the original article was not covered in sufficient detail - driver control. While I firmly believe that the ideal situation is to damp a resonant body at the source, this is not always feasible or even possible. There is also the occasional driver that simply cannot be controlled from "an ohm away" - i.e. it may require that the source (amplifier) is hard-wired to the driver, with an absolute minimum of resistance or impedance between the two. Some compression drivers (for horn speakers) are an example, where even a few hundred milliohms may allow the driver to do "its own thing" rather than

faithfully reproduce the applied signal.

Driver control (AKA "damping factor" - somewhat erroneously IMO) is a much touted parameter, and is considered important by the majority of hi-fi enthusiasts/ audiophiles. Indeed, even where a defined amplifier output impedance is used (such as 4 ohms, as used in my own system), this is done to provide a specific loading to the voice coil motor to control the back-emf that is developed in any electromagnetic loudspeaker driver. The most commonly sought after figure is zero ohms, implying an infinite "damping factor", but the laws of physics conspire to make this unrealistic.

However, a damping factor of (say) 100 or more is easily achieved, even with typical loudspeaker cables and amplifiers ... or is it?

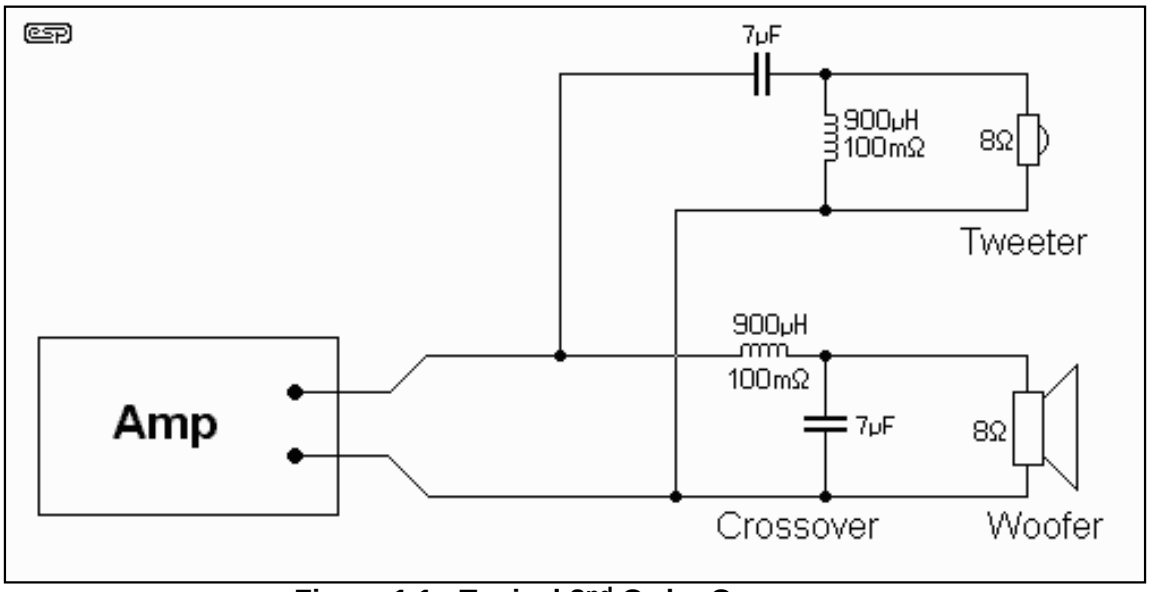


Figure 1.1 - Typical 2nd Order Crossover

There is no attempt on my part to add impedance compensation networks, notch filters, or any of the other typical additions to the circuit, and for convenience I have used purely resistive "speaker drivers". My one concession to a conventional design here is that I included 100mΩ resistance in the inductors. Any additional circuitry will affect the impedance seen by the driver - in some cases it will introduce an advantage, in others a disadvantage. I shall leave it to the reader to determine the specific differences (with a little guidance, of course).

One may rest assured though, that the performance changes due to extra circuitry will only modify the performance to a marginal degree - the primary issues remain unchanged.

2.0 - How Many Ohms?

For the sake of this discussion, we will assume a perfect amplifier, with an output impedance of zero ohms, and zero ohm speaker cables. I know this is unrealistic, but it shows the real situation very clearly - "real" components will be worse - always!

The exact parameter we will examine is the impedance "seen" by the loudspeaker driver, over a range of frequencies from well below the crossover frequency, to well above. Conventional logic indicates that

this should be as low as possible over the entire frequency range. There has been a concerted effort by amp makers to ensure that their products output impedances are as low as possible to satisfy this requirement. Valve (tube) amps are naturally different in this respect, although that is not part of this discussion.

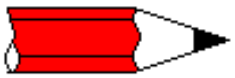
Figure 1.1 shows the crossover connections used, and the circuit is a conventional 2-way, 2nd order (12dB/octave) Butterworth type. Note that *all* versions of crossover will have similar response characteristics, although there are significant differences that will be looked at a little later in this article.

2.1 - The Crossover - a Different View

To see exactly what happens (and why), we need to redraw the crossover network so that it can be examined from the loudspeaker driver's perspective. As shown below, we see each section of the crossover (high and low pass). This is exactly the same crossover as shown in Figure 1.1, but redrawn. Since we are looking only at the damping, the amplifier is irrelevant and has been removed from the picture. It is assumed (along with speaker leads) to have zero impedance.

Remember that for this exercise, we are looking at the impedance seen by the loudspeaker, as this has a direct effect on the ability of the amplifier to damp the back-EMF (Electro-Motive Force) from the motor assembly. The back-EMF is produced whenever the cone is moved by a current, or the current is removed or changes direction. Inertia of the cone and suspension means that it cannot move or stop instantly, so there will be "overshoot" and "undershoot" caused by the cone continuing to move after the applied current has stopped.

A simple demonstration can be done to show that the speaker does indeed "generate" a voltage and current. Take a small speaker (not a tweeter), and connect it to an unused input on your preamp. Advance the gain of the amp slowly, whilst gently tapping on the cone. "Thump, thump" says your hi-fi. You can even speak into the loudspeaker, and it will act as a microphone.



Be careful - if you increase the gain too far, you may get acoustic feedback - potentially at **very** high volume levels. This will do little for your hearing, and may also damage loudspeakers. Make sure you keep the "microphone" as far from the speakers as possible to minimise the likelihood of feedback.

This simple test shows that loudspeakers do indeed generate a signal, and it is this signal that the amplifier is meant to absorb, by means of "damping factor". A back-EMF signal is generated every time your amplifier sends a signal that causes the cone to move - namely, all the time when you are listening to music (or home theatre). It is this signal that we will investigate in this article, and no other parameters. All dynamic (electro-magnetic) loudspeaker drivers do this - bar none. It should be obvious that if you short circuit the speaker that you used as a microphone, then you will hear no sound from it - this is maximum damping factor, and is what is *meant* to happen with your loudspeakers.

The phenomenon that you have experienced by using a loudspeaker as a microphone also happens with your real speakers! The woofer *will* produce a signal that is picked up by the midrange or tweeter, which in turn *will* generate a signal. This signal *should* be dissipated entirely by the amplifier to prevent (as far as is possible) the cone moving in sympathy with the soundwaves. As we shall see, this cannot happen

as it should with a passive crossover!

Even electrostatic drivers will do the same thing (although by a different mechanism entirely), but their mode of operation is such that the generated signal is of extremely small amplitude (perhaps a few millivolts at the very most). We shall not concern ourselves with this.

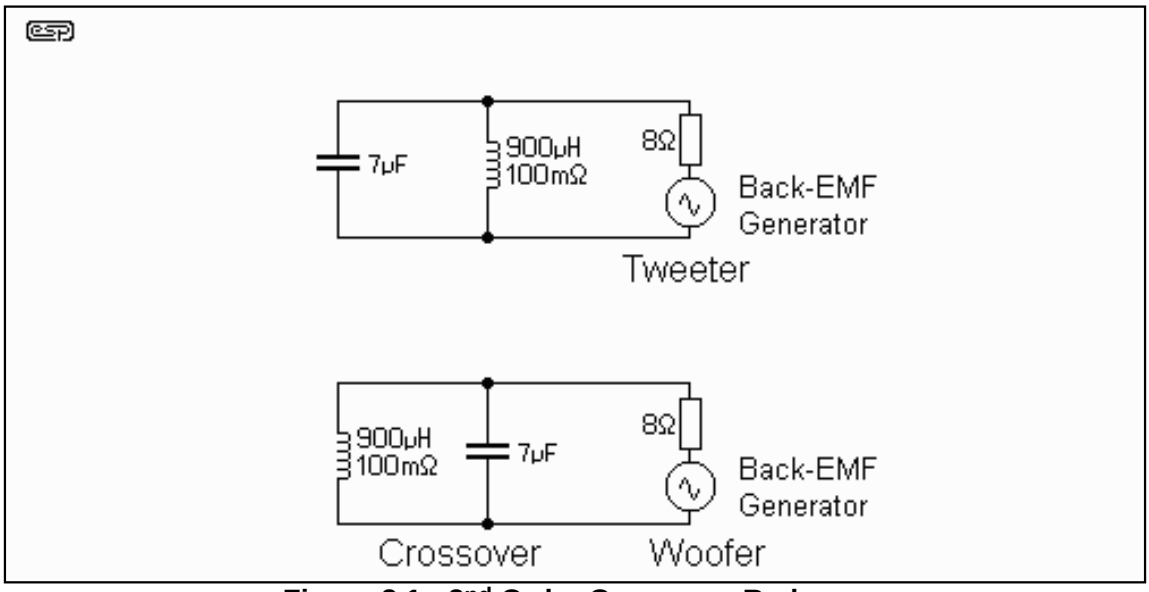


Figure 2.1 - 2nd Order Crossover Redrawn

The generators in series with each "driver" in the above diagram are to simulate the back-EMF from real-world drivers, and this is exactly the equivalent circuit that exists in reality. The only difference is that I used 8Ω resistors rather than the complex impedance of real drivers. This changes nothing, but makes the following graphs more comprehensible, without the wild fluctuations that would only confuse the issue.

Figure 2.1 shows the crossover network as it is seen by the loudspeaker. The amplifier and speaker leads no longer exist, as they were assumed from the beginning to have zero impedance. The crossover now appears as a simple parallel LC network, with resonance tuned to the crossover frequency. For those who know what this means, the implication is obvious. For the remainder, we have a parallel tuned circuit, and with ideal components (no losses), its impedance is infinite at resonance! That means that at resonance, there is no damping whatsoever, and the "damping factor" is ... zero!

But wait - there's more! Let's look at the impedance of this network over a couple of octaves below and above the crossover frequency. This gives a more balanced perspective, and we can determine the effective damping over a sensible range. The damping within the stop band (i.e. the band of frequencies the crossover network section rejects) is not so important, as the signal applied to the driver is minimal anyway. There is still the potential for considerable energy within the first octave above the crossover frequency (X_f) for a low pass section, and an octave below for a high pass section, so this is still of some importance.

Figure 2.2 shows the impedance curve of a 2kHz XO, looking backwards from the loudspeaker into its crossover section. I have shown the low and high pass sections here, but they cannot be separated because they are identical (as one would expect, since the inductance and capacitance are in parallel in both cases). There may be small differences with real components having some tolerance, but they do

not affect the picture significantly.

Indeed, even adding a 1Ω resistor in series with the short circuit shown (and thereby introducing some real-world losses into the parallel networks), there is only a small change. Naturally, it does not improve the situation.

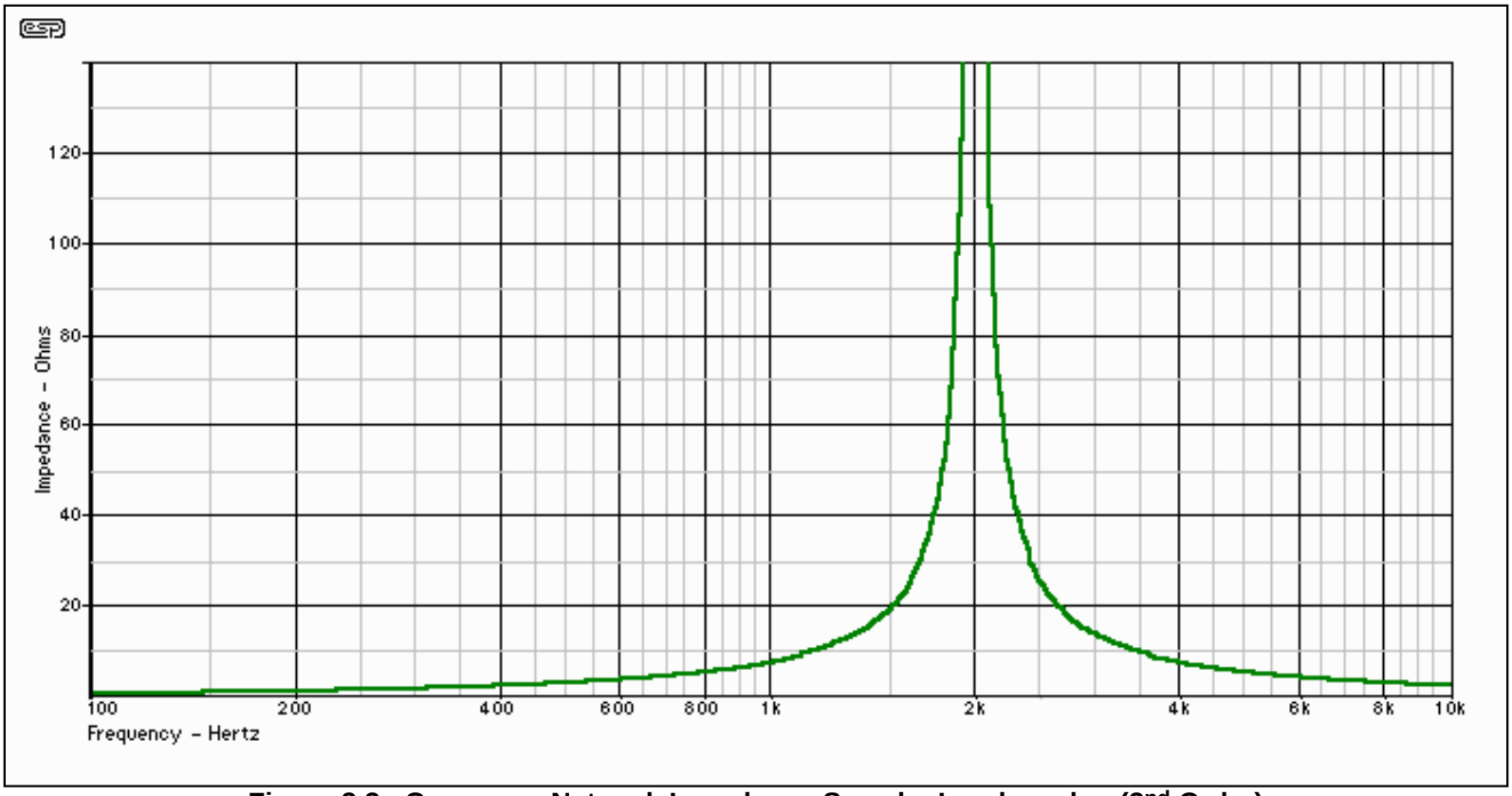


Figure 2.2 - Crossover Network Impedance Seen by Loudspeaker (2nd Order)

The high pass filter impedance response is shown in red, and the low pass in green, although only one is visible since they are perfectly overlayed.

At one octave, the impedance is nominally the same as the design impedance, so for an 8Ω speaker, the network impedance is also 8Ω one octave above and below X_f. This means that the driver sees a damping factor (DF) of one! And this with a perfect amplifier, and superconducting speaker leads. This is not only unexpected, but is potentially quite unsatisfactory, as there is little to damp the loudspeaker back-EMF, so allowing perhaps significant overshoot and undershoot, with inevitable "smearing" in the time domain. Transients will not be right, as the loudspeaker is still able to contribute a significant amount of its own "signature" to the reproduced sound.

What about moving further away from X_f ? Well, things improve, but not as much as you might expect or desire. At 2 octaves (500Hz and 8kHz), the parallel tuned circuit has an impedance of 3 ohms, so the DF is now ...

$$DF = Z_{speaker} / Z_{source} = 8 / 3 = 2.66$$

This is a far cry from the DF of between perhaps 50 to several hundred presented by the amplifier, and

for many drivers may be unsatisfactory. Even at one decade (200Hz or 20kHz (i.e. 3.16 octaves either side of the XO frequency of 2kHz), the impedance is still 1.17 Ohms, giving a DF of only 6.8 - in a 3-way system, it is probable that the low-mid XO will be close by the 200Hz figure, and this will introduce even more problems!

2.2 - The Active Solution

With an active crossover, the amplifier is connected directly to the driver, and the only thing between them is the loudspeaker cable. The amplifier presents the maximum damping factor at all times, regardless of frequency, and is not affected by the crossover network, since that is also active, and located before the power amp.

The loudspeaker driver now has the maximum control that the amplifier can provide, across the entire frequency range - not just the crossover network's pass band. The difference in damping is quite obvious, and although some (very well behaved) drivers will show little improvement, the vast majority will be much better controlled, and this will show in an impulse measurement. Not at all uncommonly, it will also show up on a swept sinewave frequency response measurement as well, with the amplitude of peaks and dips generally reduced (albeit marginally in most cases).

Well apart from the other advantages of an active system, this is perhaps one of the most compelling reasons to use an active system rather than passive. Not only is it possible to achieve the maximum damping, but if it is determined that a particular driver is best suited to some defined impedance, this can be provided by the amplifier, and will be stable across the frequency range. In some cases, just a series resistor will be sufficient, and even though there will be some power loss, if it makes the driver behave the way it should, then any small power loss is a small price to pay.

In short, there is simply no comparison between the two systems. A passive XO will *always* add (usually) undesirable impedance to that seen by the driver(s), the impedance is frequency dependent, and ranges from perhaps an ohm or so to almost infinite. The potential for uncontrolled cone movement, intermodulation distortion and loss of performance is so great that it is impossible to determine in advance, but it is all negated in one fell swoop by using a fully active system.

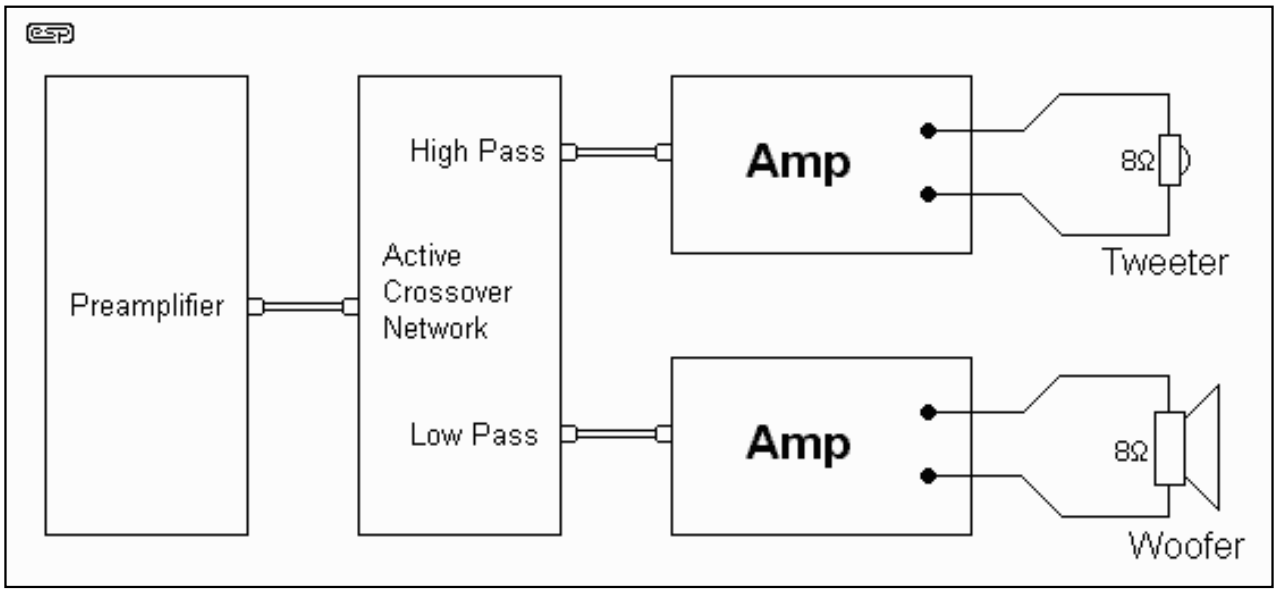


Figure 2.3 - Block Diagram of an Active 2-Way Loudspeaker System

Figure 2.3 shows the essential parts of an active 2-way system. This may be expanded to 3-way, and used with 3-way speakers, or 2-way speakers and stereo subs. Four-way systems - or more - are easily achieved. In contrast to a passive crossover (whether fully optimised or not), each driver has its own amplifier, and each amp has to reproduce less power, and over a narrower frequency range. This allows each amp to have an easier time with a less complex load, potentially reducing amplifier heating and overload - even at high listening levels. For a complete rundown of the other benefits, see The Benefits of Biamping (Not Quite Magic, But Close).

The important point here is that each driver has its own amplifier - there is nothing in between except for the cable, and amplifier control is maximised. The demands on the cable are also minimised (assuming that you believe this to be a critical component), and cheap speaker leads in an active system will provide far better performance than expensive leads with a passive crossover.

2.3 - Other Network Orders

First, let's look at a 1st order network, as this is the network of choice for many audiophiles. Notwithstanding any other problems it may have (due to the shallow rolloff slope), the impedance seen by the drivers is shown below.

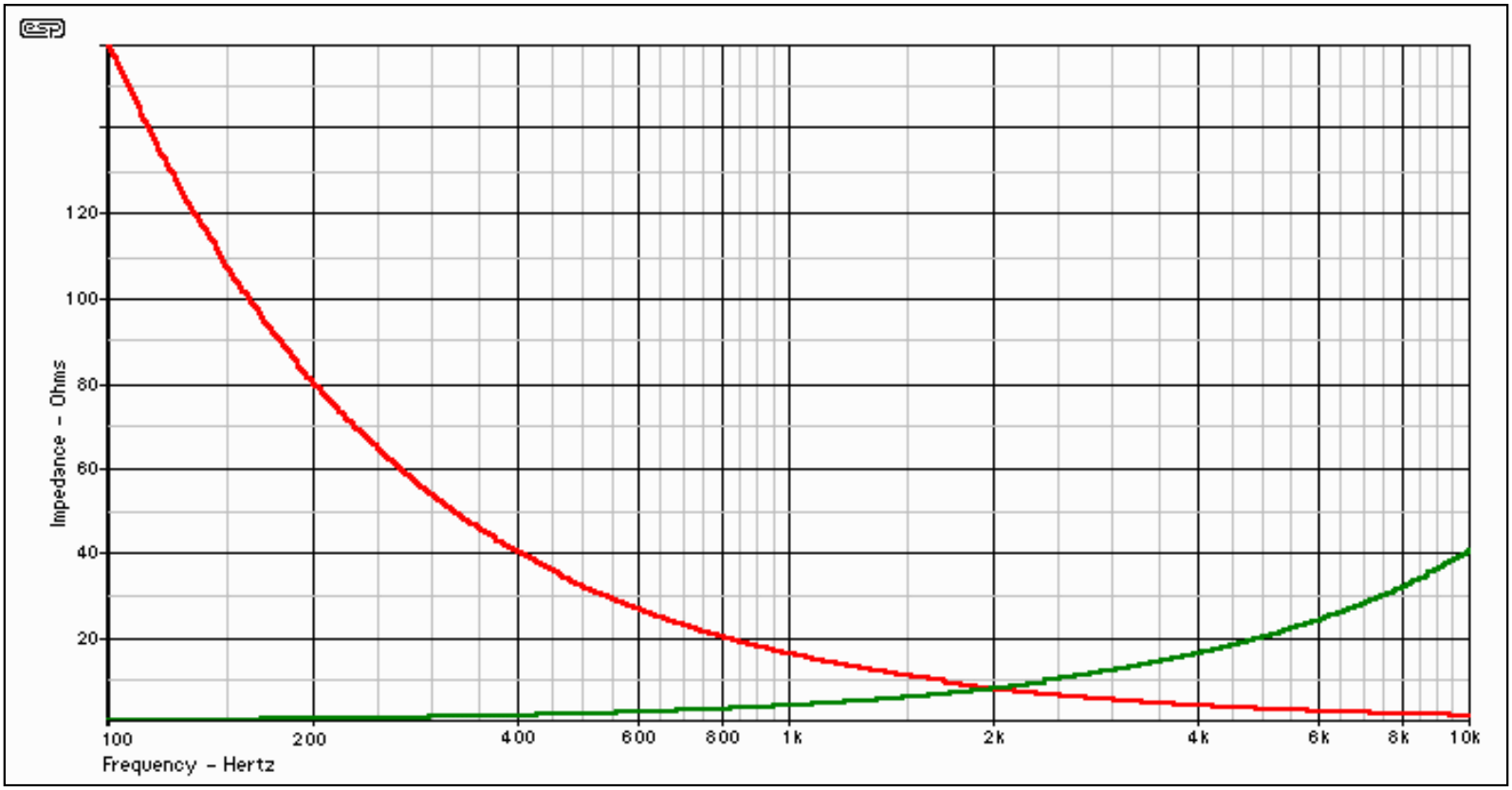


Figure 2.4 - Crossover Network Impedance Seen by Loudspeaker (1st Order)

Again, the high pass filter impedance response is shown in red, and the low pass in green. At the crossover frequency, the impedance is equal to the speaker design impedance, or 8Ω in this case. This provides a DF of 1 - significantly better than a second order filter, but still somewhat shy of ideal. At 1

octave below X_f, the low pass section shows 4Ω - still better than the second order which gave a DF of 2.66 at the same frequency. At one decade, impedance is around 800mΩ, again, an improvement over the second order filter. Unlike second order filters, a first order filter keeps increasing its impedance in the stop band, and at 1 octave above X_f stop band impedance is 16Ω, rising to 80Ω at one decade.

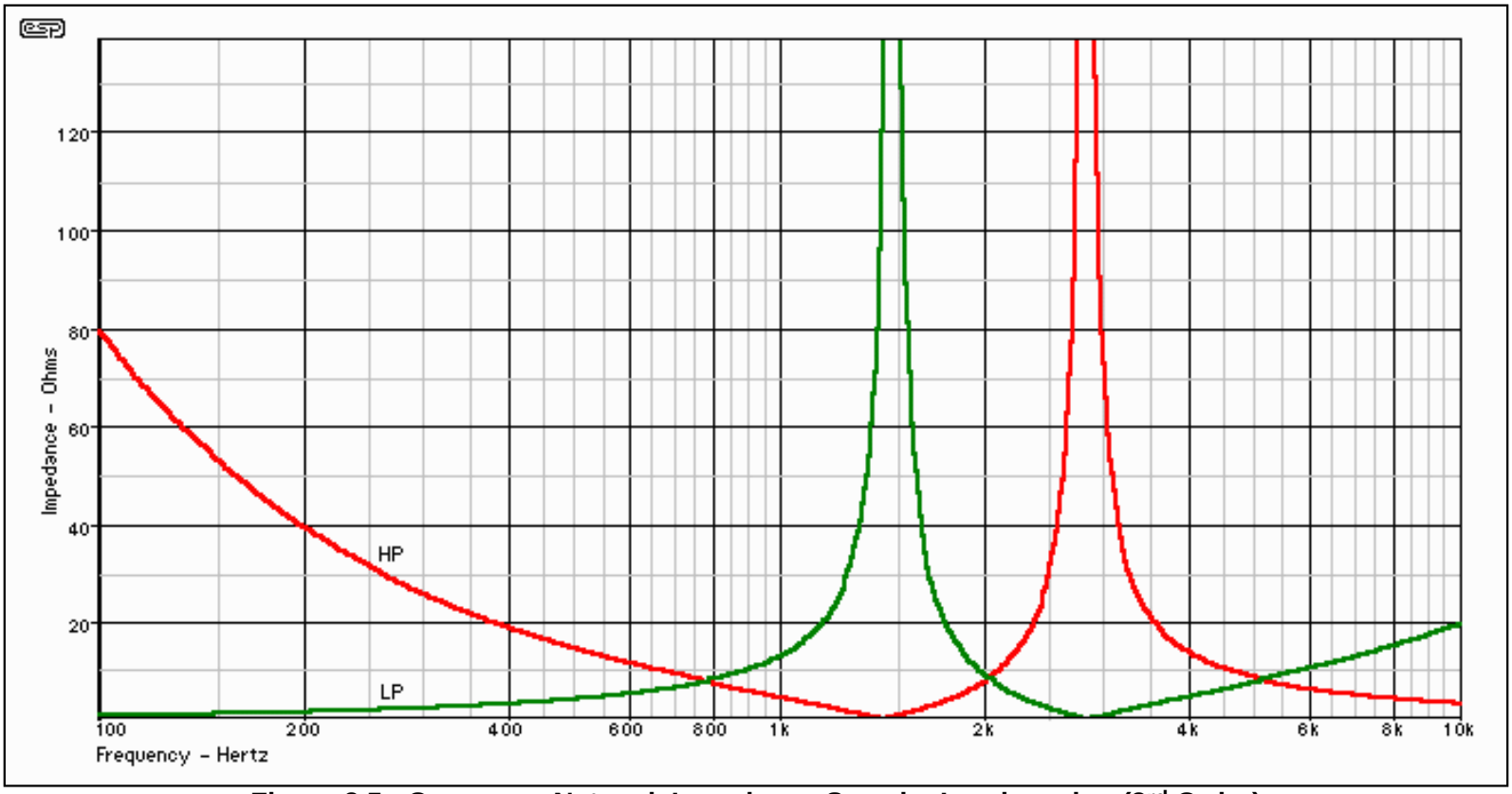


Figure 2.5 - Crossover Network Impedance Seen by Loudspeaker (3rd Order)

Here is where things get *really* interesting (colours as before). Note that there are two peaks of almost infinite impedance (the maximum is not shown, but will typically be in excess of 200Ω). These are at 1.46kHz (theoretically 1.414Khz in fact) for the low pass section, and 2.8kHz (2.828kHz) for the high pass. This means that there is a point where the loudspeaker driver sees almost infinite impedance to back-EMF, within the normal passband! That this will cause some unexpected results is fairly obvious, but it is unpredictable unless you know the drivers' behaviour at these frequencies, when driven by a high impedance source.

What of the other frequencies - 1 octave and one decade away from the crossover frequency? Because of the behaviour of the third order network, we need to look at X_f as well.

Impedance at X_f is 8.45Ω, and at 1 octave either side of that frequency, the impedance is about 13Ω. This gives a DF of 0.95 at X_f, and 0.6 at the one octave points (within the passband - impedance is much lower in the stopband, at around 4.7Ω). At one decade, passband impedance is 1.58Ω and stopband impedance is about 39Ω.

Interestingly, the impedance in the stop-band (again at those "magic" frequencies of 1.414kHz and 2.828kHz), the impedance seen by the drivers is extremely low, at only a few milli-ohms. This is clearly

visible in Figure 2.5 but is of little consequence in reality.

2.4 - Zobel (and Other) Networks

By adding (for example) a Zobel or series LC network to the tweeter to reduce the effect of the impedance peak at resonance, the impedance seen by the tweeter will be lower than indicated above. The woofer is (of course) unchanged by this, but again, a woofer Zobel used to equalise the rising impedance due to voice-coil inductance *will* have an effect.

Don't imagine for an instant that it will cure the problem, because it won't. The impedance to back-EMF will still be a great deal higher than you ever imagined, and damping factor will be around unity at best. There is no network that can be placed in parallel with the loudspeaker that will solve the problem, with the possible exception of a 0.1Ω resistor!

That would be a very bad idea indeed, unbelievably very bad in fact. Your amplifier would see almost a dead short, and even if the amp had an infinite current capacity, a 10W amp would be expected to produce 800 Watts (*all dissipated as heat in the 0.1Ω resistor!*). Don't even think about it!

But ... there **is** something you can do. It is called an active system, and you can at last obtain the genuine damping that any amplifier can produce, which is always going to be better than any passive crossover can provide.

What of bi-wiring or "passive biamping"? Not a sausage worth of difference, I'm afraid. Certainly, there may be some minor improvements in some cases, but they have nothing to do with driver damping. In my opinion, passive biamping (using two amps, but retaining the passive crossovers) is a waste of an amplifier. The same two amps with an active crossover (with the passive XO removed completely from the circuit) will outperform the passive biamp arrangement by such a margin that it's not even worth *considering* - let alone actually doing it.

3.0 - Conclusion

This aspect of active versus passive crossovers has received scant attention elsewhere, but it is very obvious that it is a major contributor to the audible difference between the two systems, even when all else seems equal. As has been shown, there is a major difference between the two types of speaker management, and this is probably the most significant (and important) distinction.

It must be understood that passive networks appear to be of sensibly low impedance from the amplifier's perspective, but behave entirely differently towards the driver's back-EMF. This seemingly contradictory situation is caused by the low output impedance of the amplifier, and this causes the impedance of the crossover filters to be asymmetrical (input-output vs. output-input).

Nothing here is magic, nor is it falsified or "tarted up" for the purposes of this article. Remember that I already stated that we would assume a zero impedance source for all tests. Note that every test shown here can be easily duplicated, using nothing more than a signal generator and a small amplifier wired in series with the loudspeaker (as depicted in Figure 2.1). You will not be able to measure impedance

directly, but the voltage obtained at the crossover's output terminal is directly related to the impedance.

It is very apparent that with a passive crossover, things are not as we would like them to be. Each variant has problems, and as with all things, a passive crossover is a compromise. IMO, this is not a compromise I am willing to make, as the performance is too unpredictable - this explains why so many passive designs require a considerable amount of tweaking before they sound their best - and may still disappoint the listener in critical listening sessions.

One of the great claims (which is completely true) for first order crossovers is that they have excellent transient response. This may well be true of the filter, but what of the loudspeaker? The degree of control offered is not good, although surprisingly (or not), it is better than second or third order filters. All passive filters will cause the amplifier to have a rather tenuous grip on the driver behaviour at best, and in extreme cases may allow a speaker to go "AWOL" at some frequencies.

In contrast, an active design minimises these problems. The driver is under the control of the amp to the maximum extent possible, regardless of frequency, passband, stopband, topology, order, etc. The use of high order (e.g. 24dB/octave Linkwitz-Riley) filters is seen by some audiophiles as a retrograde step, since transient performance is much worse than low-order filters. Be that as it may, the additional control that the amp has over the driver's behaviour *improves* the transient performance, and especially so at (or near) the crossover frequency - the most critical frequency point(s) in the design of any loudspeaker.

In this day and age, amplifiers and active crossovers can be built for (almost) peanuts - ok, not *great* amplifiers perhaps, but when used in an active system they can *still* outperform a megabuck top-of-the-line amp driving the same loudspeaker drivers through a passive crossover network.

The next step will be (of course) digital crossovers. I have one, and the ability to fine tune the network, apply delay to fully time align the respective loudspeakers, and the sheer control that each amplifier has over each connected driver, means that it is possible to make loudspeakers better than ever before. I use my digital for loudspeaker testing and development - my listening system uses an analogue 24dB/octave L-R crossover, and "time alignment" is achieved by reversing the phase of the tweeter. Not perfect, but it outperforms just about anything else I've heard. A full digital loudspeaker management system will be the next addition, and I already know just how good it will sound.

The days of the analogue active crossover are far from numbered, as this is still a good way for the budget-conscious audiophile to get the very best from available loudspeaker drivers, and with far fewer compromises than would be the case for a passive system. The overall cost will not be greatly higher for the DIY types, and the chances of success are improved beyond compare.

Not considered an active system yet? Do yourself a favour - it is extremely unlikely that you'll ever regret it.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 11 Jan 2004

Elliott Sound Products **Benefits of Bi-Amping (Not Quite Magic, But Close) - Part 1**

Rod Elliott - ESP
Last Updated 05 March 2006

Articles Index

Main Index

Introduction

Most people would tend to think that bi-amping a hi-fi system (or even a sound reinforcement system) is unnecessary, or only for the most esoteric of systems. This is not the case, as the following article will attempt to demonstrate. There are very real advantages to using bi-amplification instead of the standard arrangement we commonly use, where one power amplifier must drive all the loudspeakers in the enclosure, along with the typical passive crossover network which can - at times - have a mind of its own!

If you are in a position to spend $25,000 or more for a pair of speakers, then this is approaching the 'cost no object' arena, but the majority of people cannot afford such luxuries, and must settle for something a little more pedestrian. As a result, very few systems will be as good as they can be. Biamping is not a simple tweak, and is not to be taken lightly. Make no mistake though, its application *will* improve almost any loudspeaker available, with very few exceptions.

Note
Some of the terms used in the descriptions of various design configurations may be registered trade marks. These terms (where used) are not to be taken as a reference to any particular product, company or corporation - they are used only in their generic or common technical sense and infer no affiliation with any third party.

The following is a technical article, and is not an attempt to sell any product. It is informative and all ideas herein are a combination of common knowledge (and sense), and/or my own thoughts on the subject. No reference material is quoted, since none was used (other than computer simulation of various filter types to ensure that I am not speaking through my hat).

Some further reading from other ESP pages ...

- Passive Crossover Design
- Linkwitz-Riley Electronic Crossover
- Effects of Impedance on Loudspeaker Drivers
- Baffle Step Compensation
- Why *Do* Tweeters Blow When an Amp Distorts?

Table of Contents

- **1.0 - The Basics of Bi-Amplification**
  - The Most Common Question About Biamping
  - 1.1 - Terminology
  - 1.2 - Speaker Sensitivity
  - 1.3 - Power Distribution and SPL
  - 1.4 - Actual vs Effective Power
  - 1.5 - Separating The Signals
  - 1.6 - High Frequency Energy Content
  - 1.7 - Intermodulation Distortion
  - 1.8 - Passive Crossovers
  - 1.9 - Electronic Crossovers
  - 1.10 - Speaker Sensitivity
  - 1.11 - Phase Response
  - 1.12 - Phase-Coherent Electronic Crossover
  - 1.13 - Bi-Wiring Facts And Myths (and More on Passive Crossovers)
  - 1.14 - Midrange to High Frequency Crossovers
  - 1.15 - Summary
  - 1.16 - Adding up the Plusses
- **Biamping - Part 2**
- **Glossary of Terms**

- Copyright
- Update Info

1.0 - The Basics of Bi-Amplification

Bi-amping is a technique which uses one amplifier for the low frequencies, and another for mid and high frequencies. The choice of crossover frequency is not too critical, provided that the amplifier powers are properly balanced to achieve the maximum benefit, and the drivers used are operating well within their frequency and power limits.

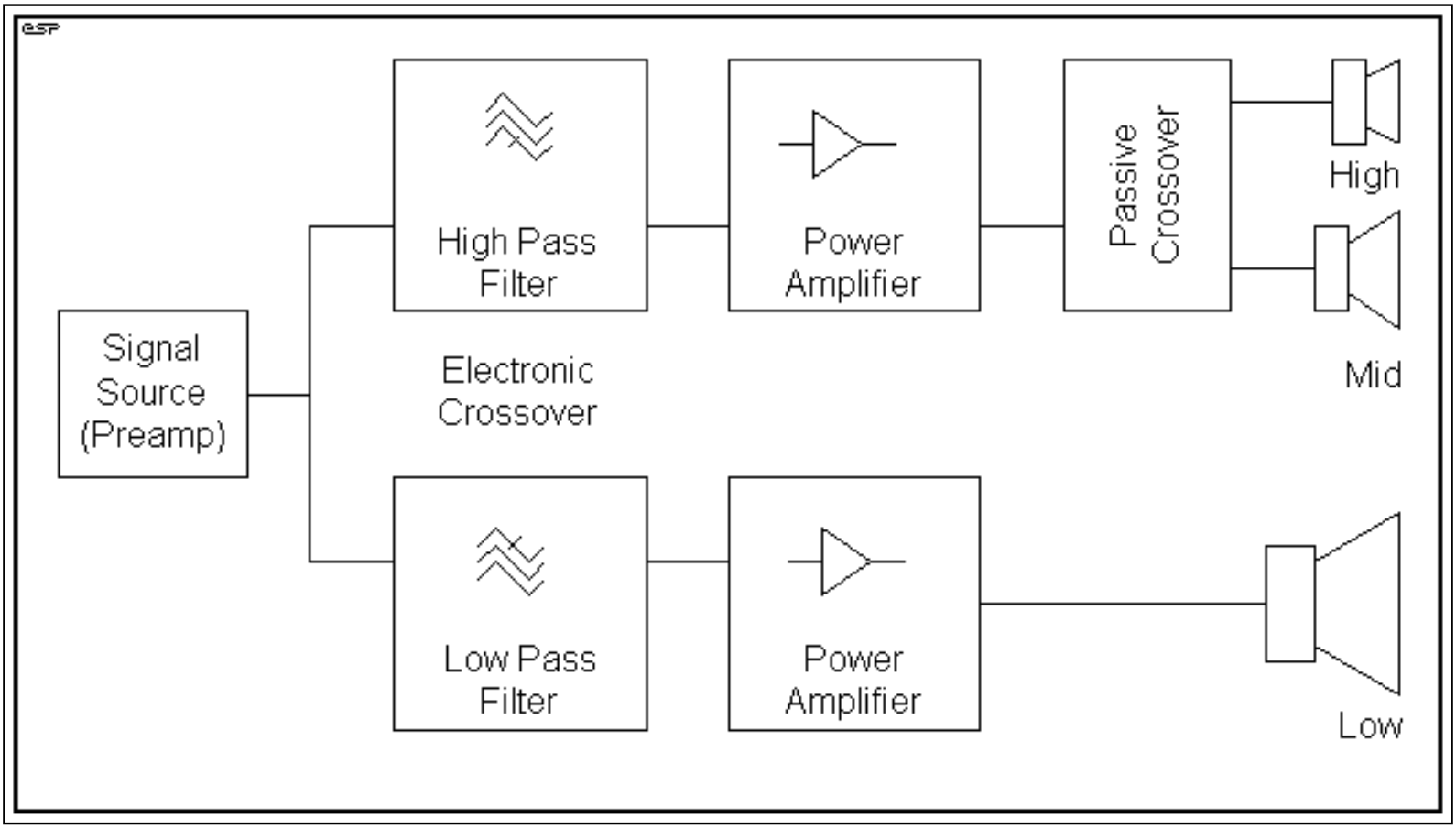


Figure 1 - Bi-Amplification Block Diagram

Figure 1 shows the basic concept in block diagram form. Only one channel of the stereo pair is shown, the remaining channel is identical. Note that the midrange to high frequency crossover retains a passive design - more on this later in the article.

In a simple form (using really simple electronic crossovers and little amps) biamping can be used even for computer speakers, clock radios and the like. The cost of the little amps is low, and the sonic improvement can be quite dramatic. I used to have a sub-woofer on my clock radio (really) and it actually sounded quite decent - at least insofar as a clock radio *can* sound decent.

As a solution to just about any amplifier-speaker combination, biamping has to be the way to go. At the highest or lowest ends of the audio equipment range, a biamped system will sound better than conventional passive crossovers, and one amp doing all the work.

The Most Common Question About Biamping

The most common question I get is ...

> "Do I need to disconnect the passive crossover in my speakers?"
> The answer is ... **Yes**, otherwise you are not really biamping at all.

Generally speaking, the mid to high section needs to be retained since a typical biamp setup will

only eliminate the bass to mid+high network. These sections are nearly always completely separate networks, although it may not seem like it when you first have a look at the board.

Equally important is the selection of the electronic crossover frequency. It must be the same as the original, within a few 10s of hertz. The only exception is where you might obtain information from the manufacturer of the speaker that allows the frequency to be modified. In general, I strongly suggest that you determine the original crossover frequency, and stay with it.

When the crossover is modified, make sure that you retain all the parts, along with the original connections. A drawing (including all component values) and photograph will be of great assistance when you want to restore the speakers to normal prior to selling them - it is unlikely that you will ever want to do this for your own use - not after you have enjoyed the benefits of biamping for any length of time.

Passive biamping (where two amplifiers are used in a bi-wiring connection) is, IMHO, a waste of money. Although there may be some moderate sonic benefits, they are not worth the expense of the extra amplifier.

1.1 - Terminology

In writing this article I have endeavoured to keep technical terms to a minimum. Unfortunately, this is quite impossible (for me anyway), so if you are not familiar with the terminology used, please refer to the Glossary of Terms, now in a separate page.

This includes some terms which are not in the body of the article, but are useful nonetheless since they may be encountered elsewhere.

1.2 - Speaker Sensitivity

Initially, let us look at an "ideal" situation, where the loudspeakers for low and midrange plus high frequencies have the same sensitivity (say 90dB / watt @ 1 metre). This means that in free space (without reflections from walls etc.), the speaker will provide an output of 90dB SPL (Sound Pressure Level) measured at a distance of 1 metre with an input power of 1 Watt. This will usually be measured with band limited noise, so the speaker's little peaks and dips will not overly influence the measurement.

The high frequency driver (tweeter) is of minimal interest at this point in the discussion, so will simply be lumped in with the midrange to give mid+high.

1.3 - Power Distribution and SPL

With typical program material (whatever that is), it has been determined that the "equal power" frequency between low and mid+high is between 250Hz and 350Hz. This is defined as the frequency where the bass and mid+high amplifier power requirements are equal. So with our 90dB/Watt/Metre speakers above we could assume that 100W amplifiers might be appropriate.

This will allow an absolute maximum of just over 110dB at one metre. You may think that is loud (you would probably be right, too), but this is the peak single-frequency SPL, and allows for transient signals - ensuring that at no time does the amplifier clip (cut-off the tops or bottoms of the waveform). This introduces distortion which quite apart from sounding awful, causes listener fatigue and places loudspeaker drivers at risk of damage.

The Speaker Damage popup has more information on this topic for those who are interested. In addition, it is suggested that you look at the article Why Do Tweeters Blow When Amplifiers Distort? for further details

The actual (averaged) SPL at one metre will be somewhere in the vicinity of 90 to 100dB, depending upon the program material. The average SPL at the listening position cannot be determined without complete analysis of the room's acoustics (for a typical room you will lose another 6 to 10dB), so for simplicity we will use the 1 metre SPL as a reference value.

Thanks to a reader, here is a small table that shows the power distribution at different crossover frequencies. The table came from a loudspeaker manual "LOUDSPEAKER ENCLOSURE DESIGN AND CONSTRUCTION" published by FANE.

X-over Frequency (Hz)	Power to Bass (%)	Power to Mid +High (%)
250	40	60
350	50	50
500	60	40
1,200	65	35
3,000	85	15
5,000	90	10

Note that according to this table, the equal power point is 350Hz (which I calculated, since it was left out of the original). This is slightly different from my own measurements, but the error is of no consequence, regardless of who is right. As can be seen, the power requirement falls quite rapidly after 1200Hz, and although not shown, it also falls off with reducing frequency.

Since the last statement will possibly cause some discomfort or indeed confusion (after all, everyone knows that a subwoofer needs more power than the main speakers), I should explain myself. Most of the time in this article, I refer to power as average power, and indeed the average power falls with frequency below about 100Hz or so. The peak power is a different matter, and depends to a very great degree on the type of music.

The table assumes equal efficiencies for the bass and mid+high drivers. Should they be different, then a correction factor must be added in. For example, if the bass driver were to be 3dB less efficient than the mid+high drivers, then the bass power must be doubled (and of course vice versa). If the difference is less than 3dB, you may safely double the power anyway, or calculate the actual power needed - this I shall leave as an exercise for the reader.

This is a very important point, and cannot be over-emphasised. Some subs (particularly those using the "Extended Low Frequency" ELF™ technique) will need a huge amount of additional power at the bottom end because of the way they are driven. It is not easy to give a simple formula (so I'm not going to :-) to calculate the power needed, because there are so many variables.

Do not be tempted to reduce bass power below about the 40% level regardless of crossover frequency, because although the average power might be quite low, it is usually of relatively high peak amplitude. The wide dynamics of the bass content require an amplifier capable of far more power than might be imagined if clipping is to be avoided. Clipping is something that one should avoid at all costs, because apart from sounding horrid, it also places loudspeakers at risk. Having said that, some peak clipping in a subwoofer will be inaudible, provided the remainder of the signal is clean.

In general, I suggest that the bass amplifier should have at least the same power as that used for the mid+high frequencies, but if any equalisation is used (such as the Project 71 Linkwitz Transform circuit), this may need to be increased dramatically.

1.4 - Actual vs Effective Power

If we assume that our 100 Watt amplifiers will be handling exactly the same peak amplitudes with typical program input, then we have a total of 200 Watts for the combined program material. So, where does the magic come into this? This amp combination will sound (and measure) as if it were 400 Watts - twice as much "effective" power as there is real power. For this to make sense, we need to back track a little.

Imagine a sine wave signal of 100Hz at an amplitude of 28V RMS. For an 8 ohm load, this equates to about 100W (98 actually). The same amplitude at 1000Hz will be exactly the same power. Now add the two signals together, in the same way that signals add together in music. We are interested only in the peak amplitude, the RMS value indicates that the power is only 3dB higher, but it is only when an oscilloscope is used that the true picture emerges.

We will now see a low-frequency waveform, with a higher frequency waveform superimposed - the high frequency signal will be riding up and down the path of the low frequency signal. If we were to perform a calculation (or simply measure the combined signal with an oscilloscope), we will see that the peak amplitude has doubled. The effective RMS value (most multimeters will get this wrong unless they are true RMS types) is 40 Volts, and this would imply 200W. Although this is the real RMS voltage, it totally underestimates the amplifier power needed to reproduce it cleanly. An oscilloscope shows 80V peak for the same waveform, so the amplifier must be

capable of passing an 80V peak signal - a 400W amplifier.

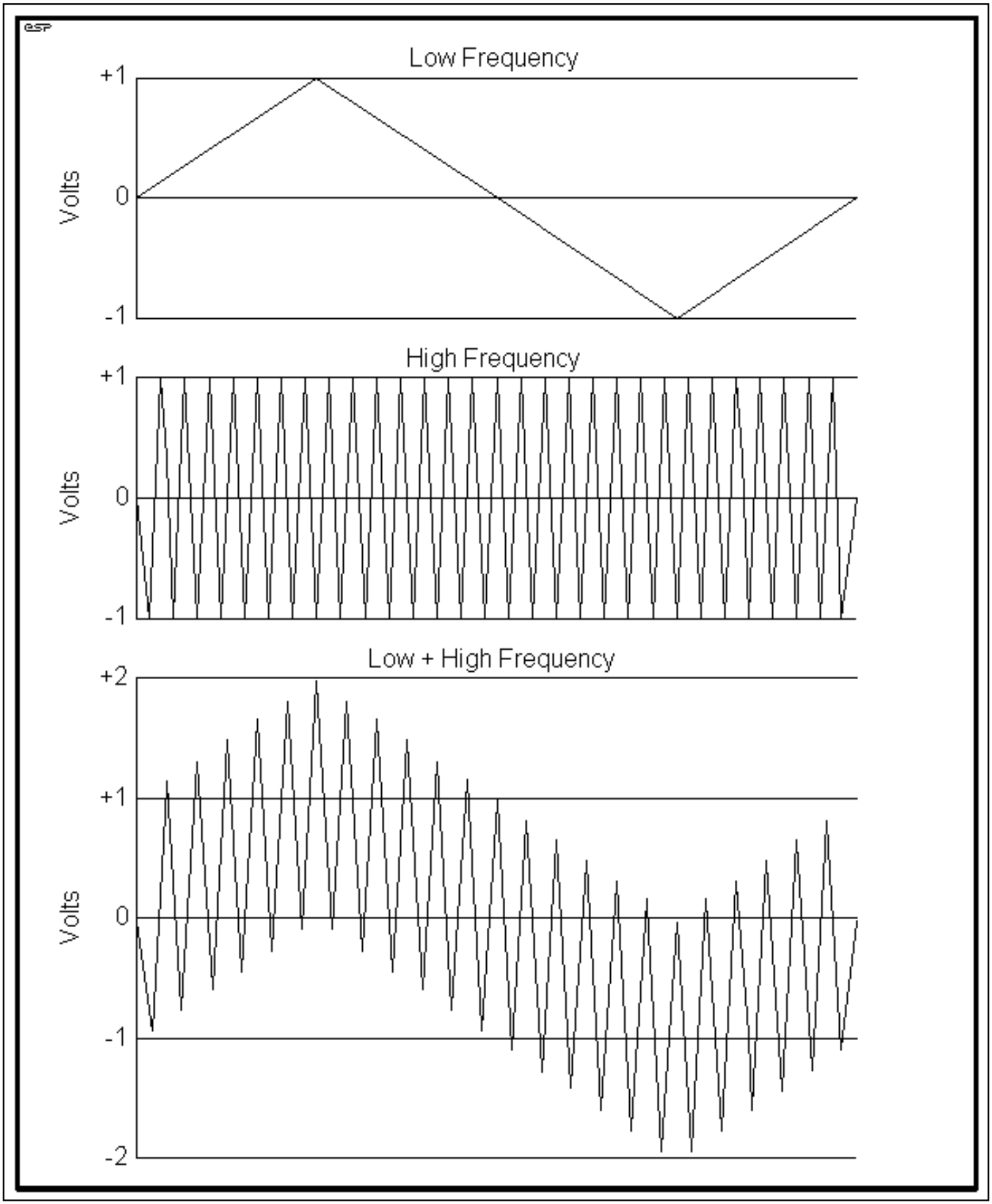


Figure 2 - Addition of Waveforms

To illustrate this point, Figure 2 shows two signals, each of 1 Unit peak amplitude. As can be seen, when the two are combined, the amplitude is much greater. The maximum peak amplitude is now 2 Units - double the peak voltage and four times the peak power of each signal individually. Power increases as the square of voltage, so twice the (peak) voltage is four times the power. Real ('RMS') power increases by 3dB or double the power, but this is a misleading figue and cannot be used. An oscilloscope is essential.

Note: Peak-to-peak amplitude is actually double the values quoted above, but since amplifiers are generally symmetrical (capable of equal positive and negative voltage swings) it is more convenient to simply refer to the peak amplitude only.

This is not to say that the actual music will be symmetrical. It isn't, but it is completely unpredictable in nature. As a result, it is possible (for example) to set up an amplifier asymmetrically and adjust the phase to suit with a switching circuit, since it will change. AM radio actually does this (or they used to) - a circuit is used to switch the phase so that slight over modulation causes more transmitter power, but never reduces it below the acceptable minimum. I shall not be going into details, since I believe few audiophiles would find this acceptable - I know I wouldn't.

All signal sources have the same characteristics as shown above in Figure 2, even a solo voice or musical instrument. In these cases, the fundamental frequency forms the low frequency component, while the harmonics 'ride the wave' as it were. Not surprisingly, the 'equal power' frequency will change (often dramatically) from the 250 to 350Hz range quoted above, but the basic principle does not alter.

Completely beside the point (but interesting anyway) is that in many musical instruments, the harmonics are actually at a greater amplitude than the fundamental. (File this away under 'Useless Information'.)

Note: It must be explained here that the 3dB effective power increase is the absolute maximum that can be obtained. In most cases it will be less - I have examined sections of music where the power gain was less than 1dB, and it can be reasonably safely assumed that the real gain will lie somewhere between 1-2dB in most cases. The real figure depends a lot on the type of music, the actual crossover frequency, and the peak to average ratio of the two separated signals. Just this topic alone is sufficient for a complete article in its own right.

1.5 - Separating The Signals

A passive crossover will separate the two signals shown above and feed each to the appropriate loudspeaker in the system. The amplifier must be capable of handling the entire composite waveform, so for our previous example of 100 Watts for each signal individually, must be capable

of 400 Watts to reproduce the waveform without distortion.

If we now we separate these signals again - prior to the power amp - and using an electronic crossover, we have an entirely different situation. (Note: It is assumed for the sake of this article that the crossover frequency is near the halfway point between the two discrete frequencies of Figure 2.) Each signal is now supplied to its own 100W amplifier (there will be but a hint of the other frequency still visible on an oscilloscope, since the filters are not "perfect") and thence to the loudspeakers. The amplifiers are not clipping, both signals are reproduced at their original power, and the effective result is that we are emulating a 400 Watt amp with two 100 Watt units.

By way of comparison, the waveforms in Figures 3A, 3B and 3C show what happens if the composite waveform is fed into a single 100 Watt amplifier, and we try to obtain the same power output as before. Once the amplifier's output voltage attempts to exceed the internal power supply voltage, the amplifier clips the tops and bottoms of the waveform - resulting in harsh distortion and placing tweeters at extreme risk due to the additional high frequency energy which is created by the sharp transitions of the clipped waveform, and even more so by the compression of the signal (see Speaker Damage). This also adds a considerable amount of intermodulation distortion to the signal, so the distortion is not just harmonic, but can also be discordant (not harmonically related). This is the worst kind of distortion, and sounds really gross.

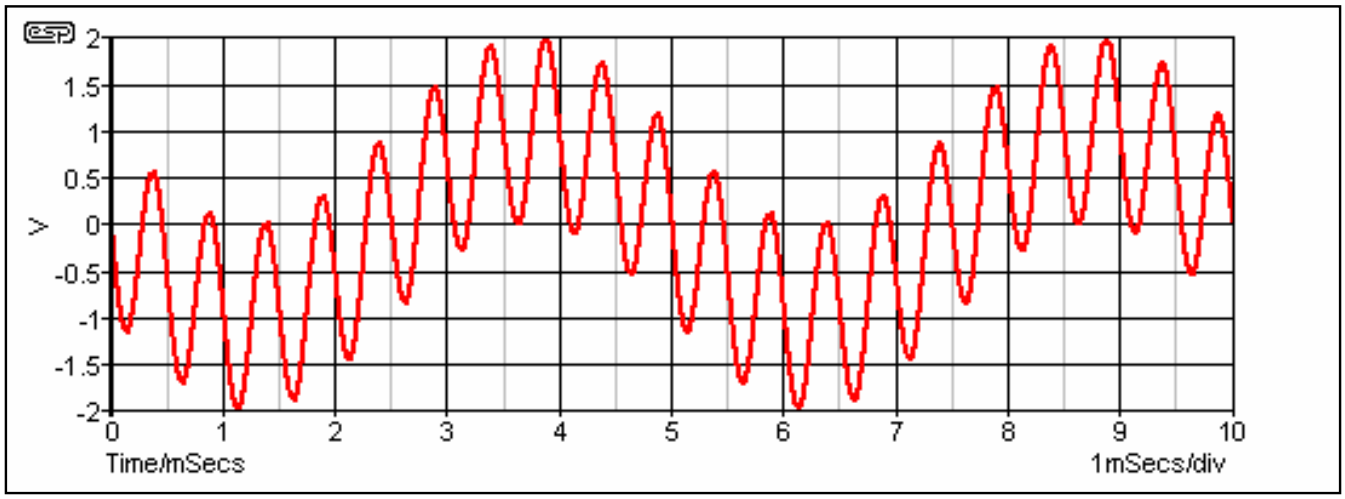


Figure 3A - Unclipped Waveform Expected from 400W Amplifier

Using the same principle outlined above, we add a 200Hz signal and a 2kHz signal, each having 1 unit (1V) amplitude. The result is a combined signal with a peak amplitude of 2 units. Again, if we equate 1 unit (1V) with a nominal 100W, then 2V is 400W.

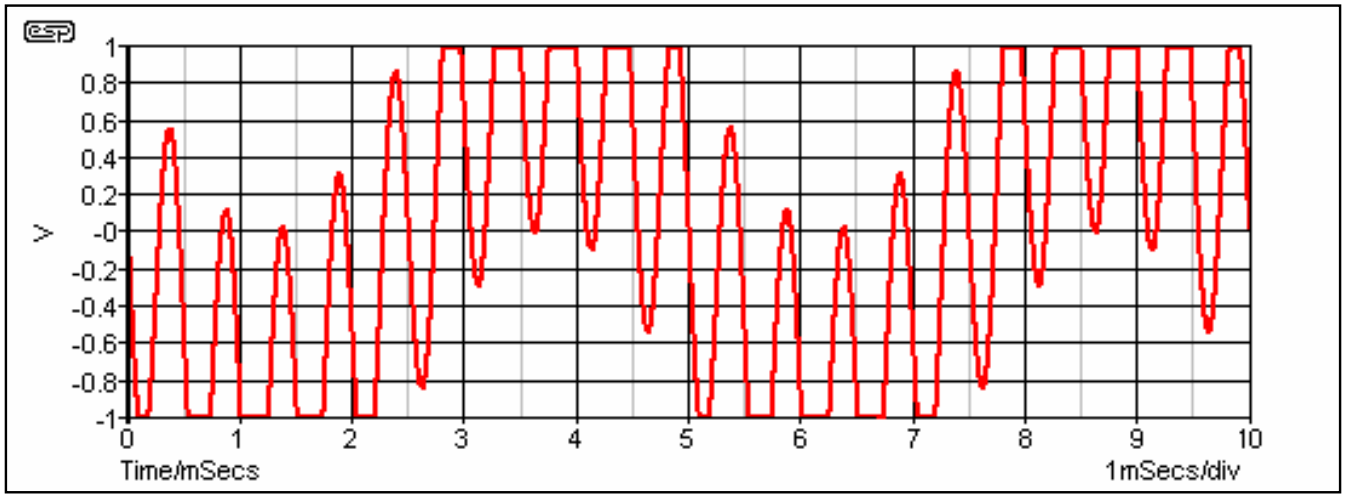


Figure 3B - Clipped Waveform From Underpowered Amplifier

The result of feeding a 2V input signal into an amplifier that is capable of reproducing 1V is shown in Figure 3B. The waveform in Fig 3B is *exactly* the same as that in Figure 3A, except the amplifier has limited the peak amplitude to ±1V, so causing the signal to be clipped. It is not immediately apparent, but both the low and high frequencies are distorted by the clipping action, and it is obvious that a significant part of the signal detail is no longer available as it has been 'clipped' off.

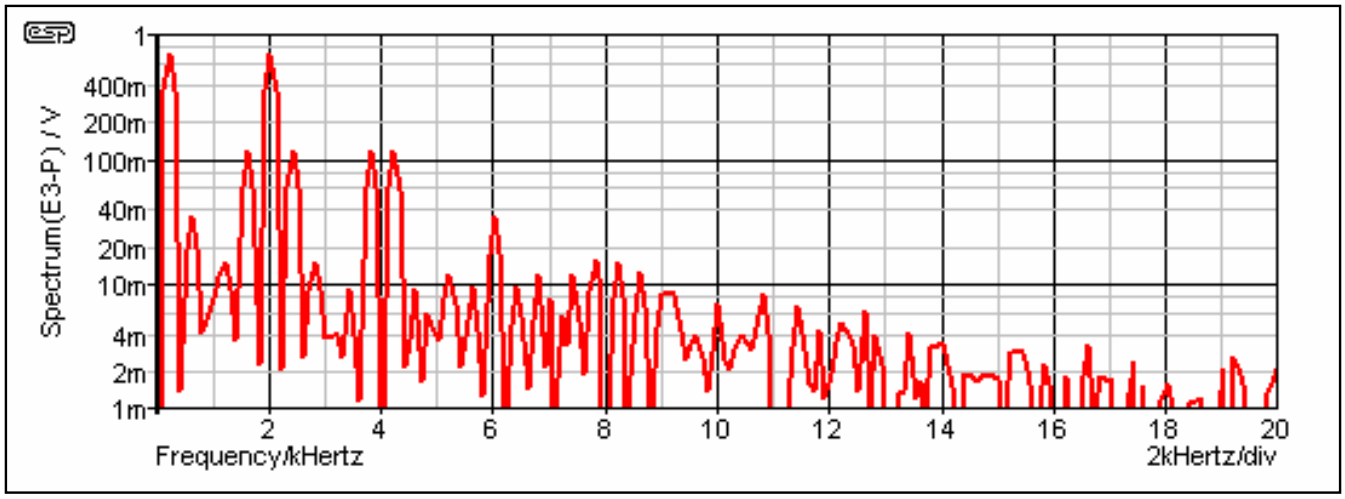


Figure 3C - Spectrum of Clipped Waveform

It is quite obvious that some of the signal has gone missing because of clipping. Not so obvious is that additional new frequencies are created, and this is shown in Figure 3C. This is a spectrum of the clipped waveform. The normal (unclipped) spectrum simply shows two peaks - one at 200Hz and another at 2kHz, with both being exactly 1V in amplitude.

Figure 3C shows that there is a multiplicity of 'new' frequencies. The original frequencies are each reduced to 714mV, and new frequencies are added. We have simple distortion, adding 600Hz to the signal, as well as sum and difference frequencies. These add 1.8kHz and 2.2kHz (at over 10% distortion level), as well as 3.8kHz and 4.2kHz. 6kHz is added (the third harmonic of 2kHz at the same level as the new 600Hz signal), plus 7.8kHz and 8.2kHz. All of these signals are above 10mV in amplitude (1% distortion referred to 1V), but there are a great many more new

frequencies below that level. The end result is a very harsh noise - it no longer qualifies as wanted sound (or music).

While it may seem that a bi-amped system gives you 'something for nothing' in the power department, this is not really the case. Four 100 Watt amps (2 x stereo 100W / channel) are going to be about the same price as (or perhaps more than) two 400 Watt amps (1 x stereo 400W / channel), but they will not be as highly stressed by high voltages, will probably run cooler, and each only has to handle a more limited frequency range. (For more good ideas on this concept, see Summary, below.). An electronic crossover is also needed, and this adds to the total cost of the system. Of course, the low frequency passive crossover isn't needed, so this offsets the overall cost somewhat. While you never get something for nothing, biamping probably comes as close as you'll get.

1.6 - High Frequency Energy Content

The basic principles described above also apply to the way high frequency signals are superimposed upon the low and middle signals. The main difference is in the energy (power) of the respective frequency bands. There is normally a relatively high amount of energy in the midrange band (see A Short Dissertation on Crossover Frequencies) as well as in the low frequency band. However, as the frequency increases beyond the upper fundamental frequencies of most musical instruments, the amount of energy falls off. Typically this will occur from about 800Hz and up (but will vary widely depending upon the type of program material), and the energy content will be seen to drop at a rate of about 3dB per octave (and more rapidly again above about 5kHz).

Since with a 3-way system the midrange to tweeter crossover frequency will be at perhaps 2500Hz or so, we can expect that the energy content of the high frequency band will be 9dB to 12dB down compared with the low and mid ranges. If we translate that back to our original 100 Watt amplifiers, this equates to somewhere between 7 and 12 Watts (peak) - giving an average power of around 1 Watt.

Because the high frequency energy content is such a low value (about 1/10th that of the midrange band), there is not a lot to be gained by using another electronic crossover network to separate this from the midrange signals. If the goal is to obtain the absolute maximum SPL (such as for sound reinforcement) it will be well worth the effort, but for hi-fi the law of diminishing returns indicates that it is not generally worthwhile. However, for optimum clarity, there is no comparison. An electronic crossover is not affected by driver impedance, and is (comparatively) infinitely stable.

One area where it is certainly better to use the additional electronic crossover and more amps for the tweeters, is where the sensitivities of the midrange driver and tweeter are more than 2 or 3dB different. In this case, using a separate amp will allow the levels to be matched far more easily, and will eliminate the use of resistive pads in the crossover network. There is also the potential for

a useful reduction in intermodulation distortion, although with good quality modern amplifiers this is normally very low.

1.7 - Intermodulation Distortion

Intermodulation distortion in an amplifier is a form of distortion created when two different frequencies are being amplified simultaneously. The effects of intermod are most noticeable when one of the frequencies is much lower than the other, and the high frequency signal is actually modulated by the low frequency. This is quite different from the signals simply adding as they are supposed to. The effect (musically speaking) is that the sound is muddied, and the highs lose their transparency. Individual instruments become difficult to separate as their harmonics all start to blend into a 'wall of sound' (have another look at Figure 3B - this is intermodulation distortion at its worst).

As described above, intermodulation distortion is not harmonically related, so its effect is worse than simple harmonic distortion. Transient Intermodulation Distortion (TIM or TID) is (supposedly) created when fast transients exceed the amplifier's ability to change its output voltage fast enough. Although uncommon in modern amplifiers, TIM is still theoretically possible, although it is very rare to find any programme material that will cause *any* reasonably competent amplifier any stress.

These effects can be hard to quantify, but by using two (or more) amplifiers, any problems will be greatly reduced.

By separating the low and mid+high frequencies from each other prior to the power amplifiers, we reduce (to a large degree) one of the major sources of intermodulation. This is a great benefit to the music lover, since the sound instantly becomes more open and cleaner.

+ So far we have identified two major plusses - effectively more power, meaning that transients are less likely to cause amplifier overload (clipping), and reduced intermodulation distortion. But wait, there's more .... (No, sorry, we're completely out of steak knives!)

1.8 - Passive Crossovers

For those who are unfamiliar with the setup of a three-way passive crossover, please refer to Figure 5, which shows (and the text explains) the connections. The diagram shown is for a "bi-wired" system, but includes the conventional connections.

When an amplifier reproduces the entire musical range, coils (inductors) and capacitors are used in the speaker cabinet to separate the high and low frequencies so that each may be supplied to the appropriate loudspeaker driver. A loudspeaker can be a difficult load for any amplifier, but

when additional inductance and capacitance enter the equation, this only makes matters worse. Add to this the fact that all passive crossovers introduce some degree of loss (in some cases as much as 3dB - which means that they are "stealing" half the available power), and one can see that getting rid of them cannot be such a bad thing.

Look at the impedance graphs for almost any speaker system, and it will be seen that there is almost always a dip in impedance (sometimes severe) at the crossover frequency. This is caused by the interactions of the loudspeakers and their inductor/capacitor networks, and in some cases can cause amplifiers considerable grief - especially at high power levels. Although few amps will fail, one can expect a reduction in effective output power as the protection circuits limit the maximum power available due to the loading of the crossover network.

These vague thoughts are brought into stark reality when one learns that the inductors and capacitors needed for the low frequency crossover are quite large values, which leads many speaker designers to compromise in the interests of economy. The inductors may have an iron or ferrite core - which improves its inductance, but ruins its linearity. So now the crossover behaves differently depending upon the amplitude of the signal. High value high quality capacitors are expensive, so again, bi-polar electrolytics are often used. It is often stated that these sound awful, although this is a somewhat contentious issue, but without any doubt their characteristics change with temperature and age. They also have rather mediocre accuracy against their claimed value (+20/-50% is typical), so a 10uF crossover cap may be 12uF, or as low as 5uF. This does not make for an accurate crossover network, and most reputable speaker manufacturers will not make this sort of compromise, at least not for their top-of-the-line models.

In addition, at high powers, the impedance of the voice coil rises because of the temperature rise in the voice coil. This is not stable, and varies widely with the music. So with loud passages, the voice coil temperature might rise significantly, which will severely impact the performance of the crossover - relying as it does on the load impedance being a constant. A loud bass solo followed by a relatively quiet but complex passage might create an interesting shift in the crossover frequency and phase response as the voice coil cools, which is unlikely to enhance the listening experience.

1.9 - Electronic Crossovers

In contrast, the electronic crossover uses active filters in the low-level signal path. These suffer none of the power "eating" problems of the passive variety, and are far more easily tuned to exactly match each other - both within the same unit (between low and mid+high), and from one unit to the next. Indeed, it is so easy to tune an electronic crossover that they could be (and often are) set up individually for the exact loudspeaker drivers installed in a given cabinet.

In the sound reinforcement industry, crossover frequencies may (or should) be changed to suit the type of music, or even to suit the acoustics of a particular venue.

There are no issues with the crossover frequency shifting, since it is stable and not at all reliant on the voice coil impedance. This will still change with the power level, but the effects are unlikely (certainly less likely) to be audible.

+ Now there are another two plusses to add to the list - elimination of the low frequency passive crossover, resplendent with its inherent losses, potentially poor linearity and crossover point inaccuracy (either as manufactured or with time, or both), and the reduction of the difficulty of the load presented to the power amplifier. Both of these result in more effective available power, ensuring that transients are preserved and overall linearity is improved markedly.

1.10 - Speaker Sensitivity

This rated a scanty mention above, but is a highly contentious issue and can cause (does cause?) many a fine speaker system to suffer from a relatively low overall sensitivity. When a speaker manufacturer chooses drivers for an enclosure, they should be the very best available for the intended final product. In many cases, although other characteristics may be ideal, the chosen drivers will have different sensitivities. This is generally solved by "padding", using resistive dividers to reduce the sensitivity of the more sensitive driver to match that of the least efficient. So if our hypothetical drivers (as described above) were to have the following efficiencies:

- Low 90dB / W @ 1m
- Mid 93dB / W @ 1m
- High 92dB / W @ 1m

It is immediately apparent that the midrange loudspeaker requires only half the power of the low frequency unit for the same output SPL. (It is 3dB more efficient, and 3dB equates to half (or double) the power.) This will never do for a quality unit, so it must be padded back by 3dB if the midrange is not to be prominent. Likewise, the efficiency of the tweeter is also too high, so this must be padded by 2dB to bring it into line with the others.

This represents a whole bunch of power being thrown away, simply dissipated as heat in resistors in the crossover network. But wait! Amplifier damping factor is a much quoted and highly sought after commodity. It mainly affects the low frequency drivers, but midrange loudspeakers are just as likely to have their own little resonances, too. Admittedly, these are much easier to control than the low frequency nasties, but it does seem to be such a shame to use all that expensive cable to ensure the best possible response and damping factor, and there it is - gone - filched by a couple of grotty little resistors.

Oh, and speaking of resistance. Remember the inductor for the low frequency crossover? Well the resistance of that is probably between 3 and 20 times greater than the resistance of your expensive cable. Don't assume that bi-wiring helps this either, because it doesn't (more on that a

little later).

With our bi-amplified model, we can simply adjust the relative gain of the amplifiers (and their power too, if maximum SPL is the goal) to bring everything back into balance. No power is lost as heat in redundant passive components, and we can ensure that the damping factor of both low and midrange drivers is not compromised by the crossover components. The low frequency loudspeaker in particular is driven directly by the power amplifier with only the speaker cable in between.

There are also some loudspeaker drivers that, for various reasons, will sound better if driven from a finite impedance. This may be to correct the Thiele-Small parameters, or (as has been suggested by one reader who referred me to a web site - in Russian!) to reduce driver intermodulation distortion. This is an area that I shall investigate when time permits, but I have been using this technique for many years, but for different reasons.

For more information on matching the amplifier power to the speakers, see Correcting Crossover Filter Amplitude Response, below.

\+ Add two more plusses. No padding is required to align the driver sensitivities, so we are not simply wasting power, and the damping factor is greatly improved for both the low and midrange loudspeakers (or can be individually set to the impedance that makes the speakers the happiest).

1.11 - Phase Response

This one is nearly as big a "killer" as the power gain - and from a musical point of view it may well be seen as even more important (I happen to think it is). The phase response of any crossover is quite predictable, as long as the source and load impedances are well defined and stable. In a passive crossover, this is rarely the case, and the results can be quite nasty. There is a phase transition around the crossover frequency, and with even-order crossovers (i.e. 12 and 24 dB/octave), there is actually a phase reversal between the low frequency and the mid+high frequency outputs. This can be seen if one examines the wiring of a speaker using an even-order crossover network, and it will be observed that the midrange driver is wired out-of-phase with the woofer. The same thing happens with the mid to high crossover, except that the tweeter is now back in phase with the low frequency driver.

It must be noted that the phase reversal is required only to ensure that the drivers are in phase at the crossover frequency. A couple of octaves each side, and with the inductive and capacitive load presented by a loudspeaker, the signals are out of phase to a greater or lesser degree. This can only be considered a completely unacceptable situation when you look at it, but there it is - alive and well in a great many speaker systems from the very beginnings of high fidelity right up to

the present day.

See Project 09 for a Linkwitz-Riley aligned 24dB/octave crossover that *is* phase coherent, and has the added benefit that there is exactly equal power at all times from the two drivers. A conventional crossover by way of comparison has a 3dB peak at the crossover frequency when the two outputs are summed.

I have made some amendments here, as I was previously caught out by "prior knowledge syndrome", a phenomenon that occurs when one has heard (and read) the "facts" so many times that one fails to check them properly. As it turns out, the so-called facts were wrong, and thus, so was I. With a resistive load, the phase response is as shown below, and the signals are 180 degrees out of phase at all times - not just at the crossover frequency. The problems arise because of the reactive nature of loudspeakers, which will mangle the phase response of almost any passive crossover. Many designs use complex networks in series and/or parallel with the drivers to attempt to get a net resistive load. This is not too difficult with the mid to high crossover, but is virtually impossible with the bass to mid+high crossover because of the very large resonances.

Figure 3 shows how the phase reversal in a 12dB/octave crossover comes about. At the crossover frequency, each waveform is subjected to a phase shift of 90 degrees. Since one is positive (called "leading" phase) and the other negative (lagging), the net result is that the two waveforms are 180 degrees apart - exactly out of phase. Notice that at frequencies significantly lower than the low-pass filter's cut-off frequency (defined as the -3dB frequency), there is little phase shift at all. The converse applies to the high-pass filter, so at significantly higher frequencies there is again little phase shift.

This gives rise to the phenomenon described above, where the driver phase reversal is needed to prevent massive cancellations at the crossover frequency. The cancellations will occur at other frequencies too, but are not audible because the level difference is so great.

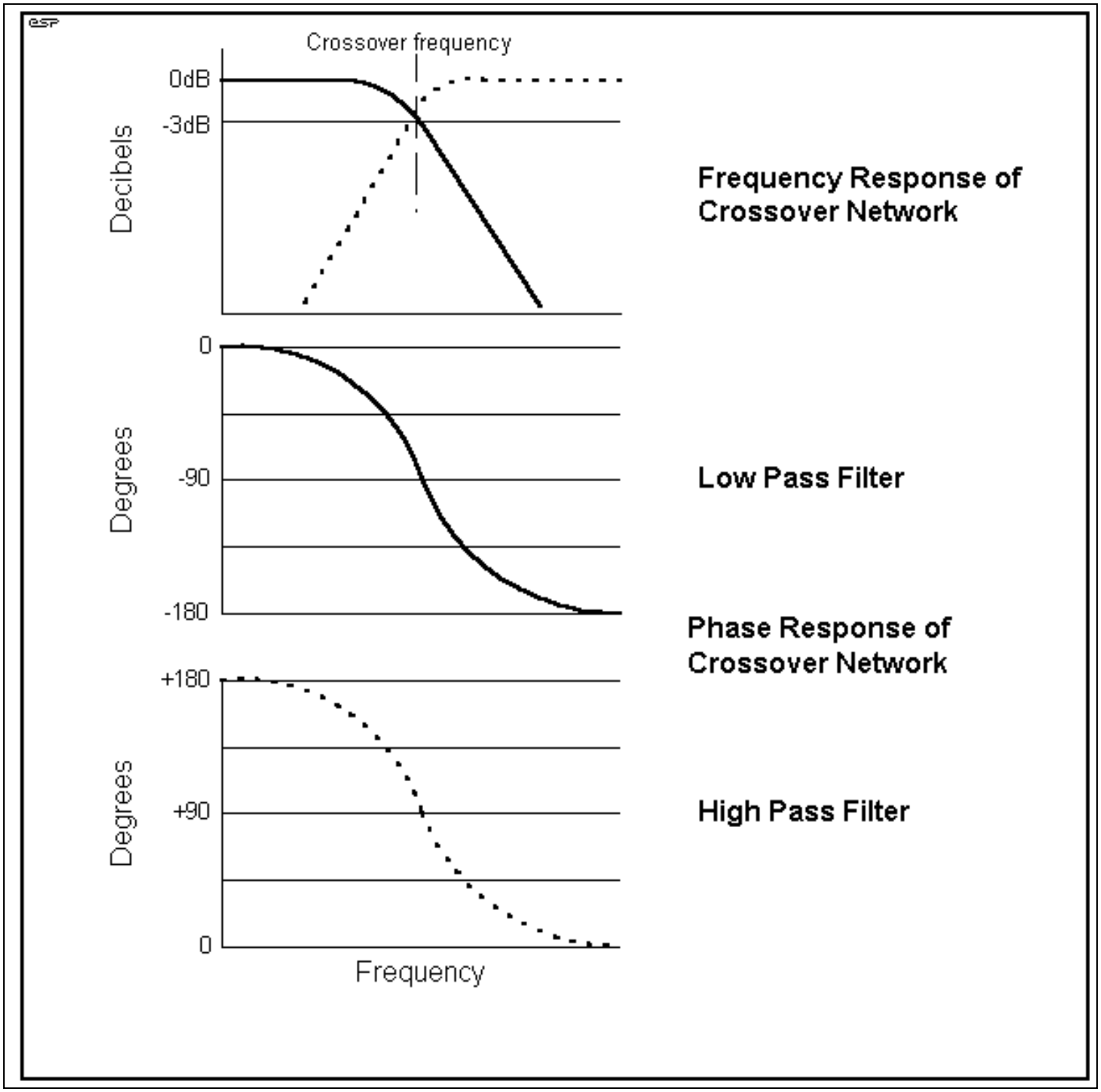


Figure 4 - Frequency And Phase Response of 12dB/Octave Crossover

It has been demonstrated by many workers in the field of acoustics that absolute phase is inaudible. Indeed, if this were not the case, then moving one's head 300mm closer or farther away from a sound source would give rise to a massive change in the perceived sound. As we all know, this is not the case. It has also been shown that some waveforms sound different if the phase is reversed, but the definitive word here is 'different' - there is no right or wrong involved. This topic is dealt with in greater detail elsewhere on the ESP site.

However, we are not talking about absolute phase but relative phase - the situation where the

phase of a signal is radiated from two different sources - *each with a different phase relationship from the other!* For this reason, many speaker manufacturers attempt to "time-align" the drivers so that the radiated signals are in the same physical plane - the idea being to combat additional phase distortion created by the loudspeaker drivers themselves. As the above shows, this is something of a lost battle - the crossover has already done plenty of damage to the phase response.

The only crossover which is relatively immune from the rapid phase transition around the crossover point as described above is a first order (6dB/octave) network, which is regrettably generally unsuitable for most loudspeakers because too much power is applied to the drivers outside their operating range. This can add considerable intermodulation distortion (this time loudspeaker induced), and is rarely an option in any system, especially between the low and mid +high frequencies. It may be an option with careful driver selection, but mostly this will only be applicable to the midrange to high frequency crossover (see below).

Although easily and cheaply built as an electronic filter, a passive third-order crossover is complex and expensive, and is more sensitive to variations in load impedance than the second-order filter.

No calculation is needed to demonstrate that if a speaker is pushed hard, its impedance will change - and this is completely aside from the reactive load presented. Most voice coils are wound using copper, and like all metals, copper has a positive coefficient of resistance. When you look at the specifications for most quality drivers, they will boast that they use a high temperature voice coil former - a good idea, since the voice coil can easily reach 150 degrees (Celsius) or more. This temperature change must cause a change in resistance, and any change will have an adverse effect on the alignment of the crossover, since impedance will change too.

Copper has a thermal coefficient of resistance such that its resistance increases by 0.39% per degree Celsius. Given a typical 6.6 ohm (DC) voice coil for an 8 ohm nominal speaker, at 150°C, the resistance will rise to over 10 ohms - naturally the impedance must be greater than this figure, so the loading on the crossover network is radically different from the design figure of 8 ohms.

At this point, the characteristics of the crossover are so far outside the design boundaries that any further calculation is futile.

1.12 - Phase-Coherent Electronic Crossover

It is possible to design a phase-coherent crossover (electronic, naturally), which exhibits none of the problem characteristics of the passive types. In an ideal world, the residual output (i.e. high frequencies below or low frequencies above the crossover point) would be in phase with the main output at any frequency in the spectrum. This ensures that there will be no cancellations or reinforcement of the signal as the outputs of the loudspeaker drivers re-combine in front of the cabinet. Regrettably, this is not the case with any passive crossover, and not all electronic crossovers exhibit this desirable characteristic either.

The author has designed and built such a unit however, and it's phase performance is exemplary. All output signals remain absolutely in phase for two octaves or more above and below the crossover frequencies. With a correctly time-aligned enclosure, phase and transient response are quite extraordinary.

It should be noted that the original phase-coherent crossover built by the author is now nearly 20 years old, and resides in a loudspeaker test amplifier (mono, tri-amped, with sweepable crossover frequencies, variable impedance - the lot) and is still in regular use.

See Project 09 for a Linkwitz-Riley aligned 24dB/octave crossover that *is* phase coherent. The question of phase coherency seems to have come of age, as it were. I have seen several designs advertised that are phase coherent, and more speaker designers are striving to achieve this goal. This is as it should be, and I am most pleased to see this happening at last. Although the filter whose graphs are shown in Fig 4 is phase coherent, it requires a polarity reversal to ensure correct phase response.

Advertisers, reviewers, manufacturers and the listening public do not seem to have seen the benefits of a phase-coherent system, and most available models of electronic crossovers seem to be electronic versions of ordinary passive crossovers. Although these units still provide many of the advantages listed in this article, phase-coherence is not necessarily included unless the manufacturer specifically states that this is a feature of the design.

\+ Phase Response - A big gain for the bi-amped system, since it can be driven from a phase-coherent crossover eliminating the rapid phase variations around the crossover frequency, and no phase reversals between drivers. Transients are cleaner and the sound is more open than can ever be achieved using passive crossovers.

1.13 - Bi-Wiring Facts And Myths (and More on Passive Crossovers)

Many speaker systems now cater for bi-wiring - running a separate speaker lead from the amplifier to the low and mid+high crossovers via separate terminals on the back of the enclosure. The benefits of this technique are said to be improved imaging due to the reduced interactions of the loudspeakers and their respective crossover networks, since the amplifier acts as an essentially zero impedance source for each section (the speaker cable now has no influence on crossover performance).

Some people equate bi-wiring as a cheaper method of achieving the same gains as one would with bi-amping. This is quite obviously not the case - there are gains to be had, but they are comparatively minor. This is not to say that the "minor" gains are not worth the effort, because as you will see this is not true at all.

For those who may not be sure of how bi-wiring really works, Figure 5 shows the setup. The broken line indicates where the connection would normally be made internally (i.e. inside the speaker cabinet). When bi-wiring is used, this connection is removed (usually with links on the connection panel), and a separate cable is run back to the amplifier.

This diagram also illustrates the composition of a 3-way crossover network. Low frequencies are fed to the woofer via a low-pass filter. The remaining signal is then fed through a high-pass filter to remove the bass energy. This is the mid+high component. To ensure that the midrange loudspeaker does not receive high frequencies as well (which it would otherwise proceed to mangle), a low-pass filter is used to filter out the high frequency component. Finally, to ensure that the tweeter is protected from the midrange signals, another high-pass filter is used.

The cutoff (i.e. crossover) frequencies for the two filter "groups" will typically be in the range of 300Hz to 800Hz for the low/mid+high section, and 2kHz to 6kHz for the mid/high section. Crossover frequency selection is discussed in a following section. (A Short Dissertation On Crossover Frequency Selection)

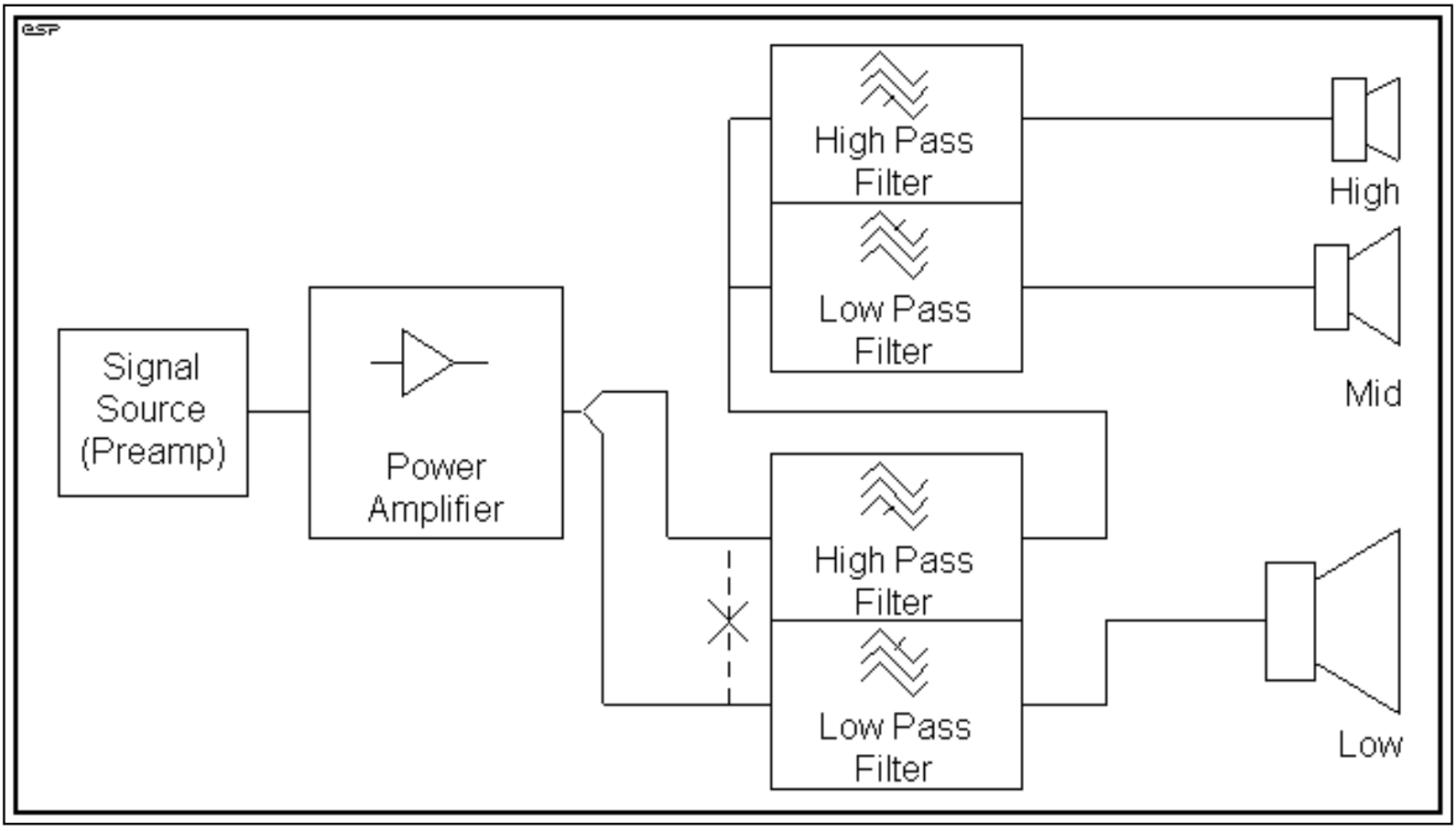


Figure 5 - Bi-Wiring Connections

As can be seen, the low frequency energy is now separated from the mid+high frequency energy in the cabling. The amplifier must still handle the full frequency range, but each section of the crossover has its own cable feed, which prevents some of the interactions between the crossover sections.

The overall effect is often (or so it has been said) a vast improvement, and largely because of the fundamental imperfections of passive crossover networks. These networks (regardless of their cost or complexity) have a few basic weaknesses which determine their overall performance. Basically, these are:

- Input impedance - a properly designed inductor / capacitor (L/C) filter relies on the input impedance (i.e. the impedance presented by the amplifier and speaker leads) to be zero ohms if the design crossover frequency is to be achieved. Any variation from this ideal will cause a small but measurable shift in the crossover's operating characteristics. When two such filters (low pass for the woofer and high pass for mid+high) are combined and share a common resistance (speaker leads) back to the source (the amplifier), interaction is inevitable, with the resulting final frequency and phase variations being unpredictable in real life.

- Load impedance - for any passive crossover to achieve its design frequency and phase response, the load must be purely resistive or wide variations in response are unavoidable. To combat this, many manufacturers design often highly complex impedance correction networks, placed across the loudspeaker in an attempt to make it "appear" resistive to the preceding filter. This task is by no means simple - a loudspeaker driver is a coil of wire (having resistance and inductance), connected to a moving mass (the diaphragm, cone, dome or whatever), all enclosed in a housing with its own resonances and reflections and coupled to the outside world via the air. This represents a highly complex electro-mechanical circuit, and presenting a pure resistance to the crossover can be approached, but never achieved (at least not over the desirable minimum of two octaves).

  In addition, these impedance correction circuits will invariably introduce additional losses, since they will generally consist of a combination of inductors, capacitors and resistors to achieve the desired result. Indeed, these networks can become more complex than the crossover itself in extreme cases, adding significantly to the final cost of the speaker system.

- Internal Impedance - as previously discussed, the coils (inductors) used in a crossover network must have resistance, since they are simply (hopefully!) an air-cored coil of copper wire. Being air-cored is the ideal, with no ferrous materials used at all. Using an iron (or any other magnetic material) core increases the inductance of the coil, so for the same value of inductance it will require fewer turns. Although this reduces internal resistance, it is at the expense of linearity and distortion, both of which are adversely affected by the use of magnetic materials.

  So, a good quality crossover inductor will have many turns of wire, which naturally has resistance. As discussed above, this should only be reduced by using a heavier gauge of wire - but compared to the cross-sectional area of your speaker cables, it is still going to be very thin! Again, power loss and variations from the ideal crossover frequency are the direct result. Naturally, speaker manufacturers are going to make the necessary frequency

corrections for the introduced resistance (in reality this is very small), but can do nothing to eliminate the power losses due to resistance. The crossover resistance also affects the damping factor presented to the loudspeaker by the amplifier - adversely of course.

- Variation of the Mechanical Components - The materials used for the manufacture of loudspeaker drivers are "real-world", which is to say they are imperfect. The roll surround of a woofer will not present the same impedance at all temperatures, and will change with age - as will all the other components. Of the fundamental ingredients which determine the overall characteristics of a loudspeaker, only the air is predictable over time (with the effects of greenhouse gasses, even this is debatable). But as we know, air varies its density and mass depending on altitude, barometric pressure, temperature and humidity, and even these variations will have some small effect on the overall characteristics of the driver. Since the response of the crossover network is dependent on all these mechanical factors as well as the electrical characteristics, it is small wonder that speaker systems seem to change their sound over time.

  Of the components which change over time, the loudspeaker itself is the greatest offender. Roll surrounds will stiffen (or start to decompose and "relax"), the cone will lose some of its stiffness and the spider (the section of the speaker which supports the voice coil) will also change. Damping materials in the speaker cabinet may settle or become more compacted as well, again changing the load presented to the crossover filter.

- Thermal Effects - So far, this discussion has taken into account only briefly the fact that the loudspeaker voice coil and inductance coil(s) in the crossover are generally of copper, and that its resistance will change with temperature. Particularly at high power levels, the self heating of these components cannot be neglected, so as the temperature increases, so too do the losses. There is also the small matter of the variation of load impedance presented to the crossover network as the loudspeaker voice coil heats up. Remember that the passive network relies on the load resistance (pure resistance is essential) remaining exactly the same at all times. Although these effects are measurable, it has neither been proven nor disproved that they are audible. I know of no speaker manufacturer who has included thermistors (thermal resistors) in a crossover network to maintain the exact design resistance at all voice coil temperatures. Indeed, the use of these devices would probably degrade the sound, since they are affected by self-heating (by the current flowing through them) which will vary with the signal.

  The use of oxygen free copper 'super' cable for speakers has many audiophiles convinced of the huge improvement (an argument I will not be drawn upon in this article, except for the note [1]below) - yet the variations in frequency response caused by the inductance, resistance and capacitance of the cable are barely perceptible (if at all) with the most sophisticated test equipment (other than ears). If these barely measurable effects are so apparent to listeners, why would the response variations of a crossover caused by temperature not be detectable? I believe that these changes are detectable, but since all (passive) speakers suffer from the same problem, this has simply been accepted as normal. These changes are also gradual, which allows time for our ears to adjust and compensate -

this is now delving into the area of psycho-acoustics, which I may examine in a future article - but not here and not now.

Again (this is starting to get tedious), an electronic crossover and separate power amplifiers will virtually eliminate the thermal effects other that a slight reduction of power as the loudspeaker voice coil heats. Even this can be corrected but the effort is not worth the gain, and due to additional circuit complexity will probably subtract more than it adds!

Many people have said that bi-wiring improved the sound quality, and although I have not used it myself (bi-amping being so far superior), I will reserve judgement until further notice. While there are some measurable differences, if sensibly sized cables are used the difference is unlikely to be audible unless the loudspeaker's crossover network has serious anomalies.

For more detailed information on the design of passive crossovers, and the many pitfalls involved, read the article Passive Crossover Design. While a mid to high passive crossover can be made reasonably economically and if well designed can sound very good (even excellent), there is often much to be gained by using a fully active system, where each driver has its own amplifier.

Note 1: It must be said that speaker cables in reality contribute little in the way of problems in hi-fi equipment. Much has been made of 'super' cables and the like, but in reality although measurable at audio frequencies, there is no proof that these effects are audible to the majority of listeners. The levels of performance variation caused by the cables are in fractions of a dB, so provided a sufficiently sized cable is used, it matters not whether it cost $2.50 / metre or ten times that amount - except to one's bank account, of course.

Note that the above may not necessarily be the case ... most speaker cables are benign, but some can easily cause a marginal amplifier to oscillate. In particular, be wary of those marketed as having a characteristic impedance of 8 ohms or others offering very low inductance. For more information on this topic, see Cable Impedance. There are solutions for these pointless (IMHO) constructions, but it's best to avoid them altogether. Any cable that is capable of making an amplifier unstable cannot offer an improvement, but can easily cause potentially serious problems.

To summarise this section, the complexity of a well designed passive crossover will be such that it will add significantly to the price of the speaker system, while still resolutely presenting the same old problems - power loss, interaction, imperfect impedance matching and rapid phase shifts around the crossover frequency. Variations of loudspeaker driver performance caused by ambient conditions (temperature and humidity for example) will affect all speaker systems (including bi-amped), but will only cause crossover frequency shifts with a passive network.

+ So now we can add a couple more plusses for bi-amplifying one's system - complete freedom from any interaction between the loudspeaker driver (and its environment) and the crossover network, and a potentially large cost saving for the now redundant complex passive crossover network.

  Add to this the benefit that many listeners already experience from bi-wiring - since a bi-amped system must be naturally be bi-wired since this is a fundamental part of its operation!

1.14 - Midrange to High Frequency Crossovers

So far I have only described the low to mid+high frequency crossover, because this is traditionally the most complex and power-hungry part of the network. There are potential gains to be had by tri-amping a system, but they become hard to justify when the added cost and complexity are considered.

The midrange to high frequency crossover still suffers from the same ailments as the low to mid +high, but they tend to be overshadowed by response aberrations created by the edges of the enclosure, grille cloth frames and other discontinuities. For this very reason, many manufacturers now use felt pads around the tweeter to reduce these effects. To some extent, the loudspeakers themselves will generally not be quite as difficult at higher frequencies, because the effects of cabinet resonances, cone mass and other mechanical factors are not as severe as at the low frequency end of the spectrum. This is admittedly a simplification, but the effects are more subtle unless the designer happens to be pushing loudspeakers beyond their limits.

Ferro-fluids are often introduced into the magnetic gap of mid and high frequency drivers, improving magnetic coupling and reducing resonances by way of the additional damping. This technique cannot be used on low frequency drivers, because the excursion of the voice coil is too great and the fluid would simply be flung from the gap as the cone moved.

Since the loudspeakers are actually more controlled at high frequencies, the complexity of the crossover is reduced. Some impedance correction is nearly always needed for the midrange driver (as its impedance will tend to rise with increasing frequency), but this is not especially arduous.

Likewise, padding may well be needed to make the driver sensitivities effectively equal, but especially in a bi-amped system this should only ever be done to reduce the sensitivity of a tweeter to match the midrange - never the other way around.

If one were to obtain drivers whose frequency response and power handling allowed it, a first-order 6dB/octave crossover network is ideal - good transient response (the best of all filter types, in fact), freedom from phase aberrations, no polarity reversal of drivers - the list goes on.

1.15 - Summary

During this article, I have given great account of the benefits of bi-amping, but nary a word about any negatives in the equation. There are some, of course, but they have actually been described already ...

- Two power amplifiers are required for each channel of a stereo system
- An electronic crossover is needed

An ideal solution would be to incorporate the power amplifiers and electronic crossover within the speaker cabinets themselves. This has been done by many manufacturers (including the author, many years ago), and is becoming quite common for high-end studio monitor speakers, especially some of the new "near-field" systems.

Even using conventional amplifiers and a separate electronic crossover network, real advantages are to be had. Imagine the best of all worlds - a really good transistor amp for the low end, providing tight and well controlled bass, and a valve (tube) amplifier for the midrange and high frequencies. With its soft overload characteristics and the renowned openness of a good valve amp, with none of the standard valve amp failings - woolly bass (generally due to the rather poor damping factor of valve amplifiers), and low frequency intermodulation distortion, caused primarily by the output transformer.

+ Another great benefit has revealed itself. Complete flexibility to choose amplifiers which are at their very best within a defined frequency range. Now the amp which all the reviewers said has "magnificent bass - but is disappointingly lacking at the top end", and the other one which is "glorious at the higher frequencies but suffers from lack of bass extension and tends towards woolliness" at low frequencies can find homes where they really excel!

1.16 - Adding up the Plusses

- Effectively (up to) twice the 'real' power of the amplifiers themselves
- Reduced intermodulation distortion
- Elimination of the low frequency passive crossover, its inherent losses, potentially poor linearity and crossover point inaccuracy
- Reduction of the difficulty of the load presented to the power amplifier
- No padding is required to align the driver sensitivities, so we are not simply wasting power
- The damping factor is greatly improved for both the low and midrange loudspeakers
- Complete freedom from any interaction between the loudspeaker driver (and its environment) and the crossover network
- Cost savings, since complex passive crossover networks an not needed
- Bi-wiring is included free!
- The flexibility to choose amplifiers which are at their best within a defined frequency range

- Ability to match amplifier power to the exact requirements of the drivers for maximum overall efficiency

I could go on (and on) here, but I shall resist the temptation. There is (IMHO) no reason to not use biamping wherever possible, from small (i.e. computer) speakers through to top of the line hi-fi. The benefits far outweigh the disadvantages in all cases.

I have seen many claims that loudspeaker manufacturers often go to extraordinary lengths to design the best possible crossover network for their products. I do not doubt that for many high-end systems, this is certainly the case. It must also be considered how much extra this costs, and we can be assured that many systems have a less than ideal network, simply to keep costs reasonable. Several times, I have seen reviews where expensive speakers use ferrite cored inductors for the low frequencies, and bipolar electrolytics are also common.

I do not consider these to be optimal or appropriate for a high quality system, and nor do many others. The truth is that cost considerations are nearly always made in any system, and much more so when the selling price becomes a consideration.

As I stated at the beginning, if you spend $25,000 or more for a pair of speakers, then we are into the "cost no object" area. Most people cannot afford such luxuries, and as a result they must settle for something they can afford. Only a very few systems will be as good as they can be, and you will pay dearly for it.

Biamping is not a simple tweak, and is not to be taken lightly. Make no mistake though, its application *will* improve almost any loudspeaker available, with very few exceptions.

There is one area where you may find that changing to a biamped system cases an apparent loss of bass response. This is rare, but in some cases the cabinet design may have been optimised for the woofer, including the resistance of the series inductor in the passive crossover. This is especially true for very well designed ported boxes. The result is that a biamped system is slightly over-damped by comparison, resulting in a loss of bass response.

There are two solutions for this ... either allow the system to 'break in' (which actually means that you get used to the new sound after a while), or increase the output impedance of the power amp. See the ESP articles index for more information on how to go about this. In general, the impedance variation is small, but it is certainly worth doing if you wish to get the best possible response from the system.

Part 2

Main Index

Articles Index

Copyright Notice

This article (including all images and diagrams) conceived and written by Rod Elliott. Copyright © 1998 - 2006 all rights reserved. Reproduction, storage or republication by any means whatsoever whether electronic, mechanical, or any combination thereof is strictly prohibited either in whole or in part without the express written permission of the author, with the sole exception that readers may print a copy of the article for their own personal use.

Some of the terms used in the descriptions of various design configurations may be registered trade marks. These terms (where used) are not to be taken as a reference to any particular product, company or corporation - they are used only in their generic or common technical sense and infer no affiliation with any third party.

Update Information: Page last updated: 05 Mar 06 - changed Fig 3 for Figs 3A, 3B and 3C. Added intermodulation data and some text./ 02 May 05 - divided page into two parts, updated drawings and text./ 28 jul 01 - added some minor commentary, and links to other pages./ 03 Jun - moved to sound.westhost.com./ 16 Aug 2000 - added small explanation of low freq power needs, minor reformat of page, added most common question./ 11 Dec 00 - added small systems to the intro, and speaker damage box + misc text mods./ 09 Dec 2000 - added tabulated TOC, modified conclusion, a few minor additions (speaker sensitivity, ELF subs, etc.)./ 29 Nov 1999 - added entry to table, and some extra comment about power distribution./ 19 Nov 1999 - Included table of power distribution (provided by a reader)./ 28 Aug 1999 - Added new information to tweeter protection section (use of DC detector, etc)./ 19 Nov 2000 - added thermal details and corrected error in phase response of conventional xover.

Elliott Sound Products | **Benefits of Bi-Amping (Not Quite Magic, But Close) - Part 2**

Rod Elliott - ESP

Last Updated 02 May 2005

Articles Index

Main Index

Contents of Part 2

- **2.0 - Crossover Frequency Selection**
  - 2.1 - Ideal Crossover Frequencies
  - 2.2 - The Ideal Compromise
- **3.0 - Tri-Amping**
  - 3.1 - DC Protection
  - 3.2 - Choice of Capacitor
  - 3.3 - Amplifier
  - 3.4 - Class-A Amplifiers
- **4.0 - Correcting Crossover Filter Amplitude Response**
  - 4.1 - Measuring Amplifier Gain
  - 4.2 - Speakers With Different Sensitivities
  - 4.3 - The Effect of Amplitude Inaccuracies
  - 4.4 - What Does This *Really* Mean?
- **Glossary of Terms**

- Copyright
- Update Info

- **Part 1**

2.0 - Crossover Frequency Selection

It would be remiss of me to not mention a few salient points about the choice of crossover frequencies. This applies to all system types where fidelity is expected (or demanded) - high power music, sound reinforcement, or hi-fi.

It is not at all uncommon to see systems where the crossover frequency is set right in the middle of what I call the "intelligence band". This is the range of frequencies from 300Hz to 3600Hz, and is extremely important from a psycho-acoustic point of view.

It is no accident that this is the range of the telephone system (and has been for many years). If we are only to hear a limited range, then this band of frequencies is by far the most important. Just from this we can recognise a person's voice, which musical instrument is being played (even bass instruments!), and - more importantly - what is being said. It contains nearly all the "intelligence" of the sound, which is to say that if this band is "corrupted", intelligibility is greatly reduced.

So why do speaker manufacturers insist on placing their crossover frequencies within this band of frequencies? The public address (PA) systems used by many rock bands are a case in point - how often does one find that the vocals are completely unintelligible? Mind you, it may also be the case that the band's lyrics just don't make sense, but that's another story altogether.

Often this occurs because the system is so loud that the amplifiers are clipping badly, but even at lower levels it is quite common. Place a common-or-garden crossover filter right in the middle of the "intelligence band" and this is exactly what will (and does) happen. With phase aberrations and cancellations, this most important frequency range becomes muddied and indistinct causing loss of intelligibility - not only on voice, but instruments as well.

The effect is also noticeable with some hi-fi speaker systems, except that it usually less pronounced, and it is far less likely that the amplifier will be driven to clipping. Reviewers will often say of a speaker that the vocals seem veiled, or that there is noticeable colouration of either male or female vocals. These effects are often caused by the effects of phase shift around the crossover frequency, coupled with the fact that the crossover frequency falls right in the middle of the intelligence band.

Should a crossover be unavoidable in this region - due (for example) to available loudspeaker drivers - then the manufacturer must go to great lengths to ensure that "artefacts" created by the crossover are not audible. This often causes greater problems with amplifier loading at the crossover frequency, since impedance dips seem a common problem with many speakers. It will be found that these almost invariably occur at the crossover frequency.

2.1 - Ideal Crossover Frequencies

Since we have already discussed the "equal power" crossover frequency between low and mid+high frequencies, it should come as no surprise that the author prefers between 275Hz and 300Hz as the ideal frequency. This is outside the intelligence band (albeit only just), but as discussed, a phase-coherent crossover network and a bi-amped system will tend to be far more tolerant than conventional (passive) crossover networks.

One problem this technique does cause, is that the demands placed on the midrange driver are greater than will normally be the case. This is because the low frequency end of the midrange is now extended to around 300Hz rather than the more "conventional" frequency of 500 or 600Hz. Few (none that I know of) so-called "enclosed" (i.e. those with their own integral enclosure) midrange drivers are capable of reproducing 300Hz accurately - indeed, many are quite inadequate even at 600Hz!

Even ruling out this style of driver altogether still leaves relatively few speakers which are small enough to be considered a point source at 3kHz (one wavelength at this frequency is only 115mm - assuming "British Standard" air temperature, etc.), yet is capable of reproducing signals down to 300Hz accurately. Ideally one would want a driver whose radiating surface is no greater than 100mm diameter (this is already a significant compromise), having high compliance for low frequency reproduction, and a stiff cone structure to prevent cone break-up at the upper limits.

Bear in mind that a loudspeaker which is going to be used to reproduce frequencies down to 300Hz should ideally be capable of uncoloured reproduction for at least one octave (and preferably two octaves) above and below the crossover frequencies. This means that a suitable midrange driver must be capable of reproducing from 150Hz to 6kHz with good efficiency and without significant colouration. This is not an easy task for any loudspeaker.

Many otherwise fine midrange drivers do not provide a wide enough safety margin below their recommended minimum crossover frequency, which causes resonances and other

effects to colour the sound. Also affected will be phase response, which will start to suffer badly as the driver approaches resonance - this rather negates the advantages of using a phase-coherent crossover network!

2.2 - The Ideal Compromise

Yes, I know the heading is an oxymoron, but that is what we really have to find. We cannot go further into discussion at this point (at least not without naming names, and deciding on some suitable loudspeaker drivers), since the "ideal compromise" will be different for every loudspeaker combination available, with added problems incurred by the selected cabinet design and the maker's design goals (price - as always - being a major player in all these calculations).

Having examined some of the factors which affect the performance of a speaker system, it is apparent that there are few hard and fast rules which can be applied, since there are so many variables. What has been presented here is a guideline which - assuming that suitable drivers can be obtained - will have a standard of performance well above average. This web site has now been updated many times, as more information comes to hand, and as I get responses from readers who have similar (or wildly different - rare!) views from my own.

It is to be hoped that this information will at least provide some further discussion and feedback from readers who share my interest in "the ultimate loudspeaker" - however it is configured (even with passive crossovers, perhaps).

3.0 - Tri-Amping

I have had many enquiries about extending the bi-amp principle to tri-amping, and offer a few thoughts here. There are some points which must be made, largely to protect the tweeters in such a system, but also to ensure that the system as a whole is coherent, with no one component of the music receiving more or less attention than the others.

3.1 - DC Protection

With a bi-amped system, the tweeters are protected by the mid-high crossover. Once the loudspeaker is tri-amped, this protection is lost, since the capacitor which is used to determine the crossover frequency is no longer present.

With most 'solid-state' amps, this places the tweeter at great risk during the (generally short) switch-on and switch-off periods. As the supply voltage is applied (or removed), most amplifiers will create a DC transient (if such a thing is possible) as the circuitry starts to operate. This causes the all too common speaker thump.

This is mildly annoying when applied to the low frequency drivers, but is capable of destroying a tweeter if allowed to persist for more than a few milliseconds.

In the case of amplifier failure, the tweeter is almost certain to protect any speaker fuse by blowing first - not exactly the desired effect! The "Poly-Switches" now available might help, but I don't like the idea of a non-linear resistor in series with my speakers.

If the direct coupled approach is contemplated, I would suggest the following:

- Use a relay in series between the tweeter and the amp with a suitable muting circuit
- Include a fast acting DC sensor to disconnect the load if amplifier DC is detected

A suitable circuit is available - see Project 33 in my Project Pages, which can be easily be modified to protect tweeters, where its DC detection circuit can be made vary fast indeed.

3.2 - Choice of Capacitor

A humble capacitor will prevent DC from reaching the tweeter voice coil, but the selection is critical to ensure that the sound is not degraded.

Value - The capacitor will almost always have to be at least 20uF, which for an 8 Ohm tweeter, will create a 3dB high pass crossover at about 995Hz. Given that this additional crossover should be ideally 1.5 to 2 octaves from the "real" crossover frequency (even more if possible), the values likely to be needed in real life will tend to get quite large. The reason that the protection cap needs to be so large is that smaller values introduce phase shift, which is significant for all frequencies within 2 octaves of the crossover point.

An alternative (I hope your maths are good) is to use a modified high pass section in the electronic crossover, and then use the protection cap to provide the last pole of the filter. This will work (it will work very well), but the mathematical complexities will be such that I

expect few constructors to go this way.

A further disadvantage is that the electronic crossover cannot simply be swapped for a different type to allow comparisons, and with some filter types the approach will not work at all.

Type - When we contemplate high value caps (greater than 20uF) there is an immediate tendency to think about using a bipolar electrolytic. For this application, I do not recommend them, but sometimes you may have little choice. According to some, they are not recommend for any application, since they are (supposedly) sonically disgusting. I have not been able to measure distortion in a bipolar electro, but there are many who claim that they destroy the sound. I shall not continue this debate.

The ideal is to use polyester or polypropylene caps, since their stability is so vastly superior to bipolar electrolytics that there is no comparison. They also have a comparatively unlimited life, but bipolars gradually lose capacitance (and sometimes not so gradually), thus changing the crossover frequency (or disabling the tweeters completely when they eventually fail.

Good caps can cause some degree of financial hardship, but be assured, that is as nothing compared to the utter despair when smoke is seen escaping from your precious tweets.

If you are on a budget (decent caps at these values are expensive), one possibility is to use power-factor correction or induction motor start capacitors. These used to be oil-filled paper (some still are), and are much cheaper than "electronics shop" devices. I can vouch for the sound quality, as I use these to protect my tweeters - most are polyproplyene are of film and foil construction, although metallised film is probably used as well. The stability and power handling will certainly be superior to that of bipolar electrolytics, and the high frequency response can be corrected (if necessary) by bypassing with a high grade polyester. I would expect that a 1uF bypass would suffice in most cases.

These caps should normally be available from electrical supply outlets, since they are commonly used in electrical (i.e. mains house/ factory/ office) installations.

3.3 - Amplifier

The amplifiers for a triamped system will have a (possibly) profound effect on the final

sound quality. This is especially true of the tweeter amp, which will generally not require a lot of power (depending on crossover frequency). If we assume that the power drops off at 3dB/octave above 1kHz for "typical" music signals, we can do a quick calculation - this is not difficult (nor is it terribly accurate), but will give an idea of how much power will be needed for the tweeters. Note that this "formula" errs on the side of safety (i.e. the tweeter amp will have more power than is really needed), and this provides a good safety margin - a tweeter driver amplifier which is clipping is not likely to enhance the sound quality !

We might quickly re-examine the power of the low and mid amps first, assuming that we have selected the "equal power" low/mid frequency of about 300Hz. For a typical system for home use, 50 Watts for each will generally be enough (especially when you remember that biamping gives the approximate equivalent of double the actual power of the amps - i.e. 200 Watts).

So, for this example, given that we have arrived at using a 50 Watt amp for mid+high, we are now going to triamp, with a crossover frequency of (say) 3kHz. This is approximately 1.5 octaves (it's a little more, but it is not worth worrying about).

At 3dB/octave, and 1.5 octaves, this results in a power requirement for the tweeters of -3 x 1.5 = -4.5dB relative to the midrange amplifier. Reversing the dB (power) formula, it can be seen that the high frequency amp will need 0.35 of the midrange amp's power.

0.35 x 50 Watts = 17.7 Watts, so a 20 Watt amp will be quite sufficient.

3.4 - Class-A Amplifiers

For 20 Watts, we can look seriously at using a Class-A amplifier, something that most of us would die for, but is unrealistic for higher powers.

High power Class-A amps are seriously expensive to build or buy, and create a lot of heat. At the small power of 20 Watts however, they start to become much more attractive. They still create a lot of heat, but since this is proportional to their output power it becomes manageable at low powers.

A typical 20 Watt Class-A amp will dissipate about 100 Watts worst case, and although this is not insignificant, it can still be dealt with by conventional heatsinks and no fan cooling. This is not to say that the heatsinks will be small - they most certainly will not - but 50W per device (assuming transistors) is not too hard to get rid of.

At these powers, one might even consider a valve (vacuum tube) Class-A design, but I would not be inclined to this approach (personal opinion), however it may be that this could make musical magic. It you try it and love it, please make sure that you let me know - I shall include your comments on the Readers' Response page.

Schematic Diagram

Refer to the Project Pages for a design of a couple of transistor 15 to 20 Watt Class-A amplifiers designed for general use, but are ideal for driving tweeters. Includes the design and basic/generic construction details. As yet, I have not had time to test one of these circuits, so final specifications are not complete, but the DoZ is a fairly nice little amp.

Also, have a look at the (now old but still useful) design - 10 Watt Class-A Amplifier (By John Linsley Hood)

4.0 - Correcting Crossover Filter Amplitude Response

When an electronic crossover is used together with the respective amplifiers for each channel, there is always going to be a temptation to experiment with the levels of the filters or amplifiers to act as a sort of tone control. To extract the maximum benefit from bi- or tri-amping, this should not be done, since it will effectively do a few things (all undesirable)

- The crossover frequency will be modified
- The phase relationships between crossovers may no longer be valid
- The frequency response will have peaks or dips (or both) where it should be nice and flat

The optimum settings for the relative gains are dependent on only one thing - speaker sensitivity.

In order for this to make some degree of sense, we must return to our previous examples, and look at a few more diagrams. First, let's look at the ideal, where both speakers have a sensitivity of 90dB/m/W. In this case, the gains of the crossover sections (if gain controls are provided) should be exactly equal. Likewise, the sensitivity of the power amplifiers must also be equal.

In many cases, different amplifiers will be used, often with differing power ratings as well. This is where some measurements are needed, since both amps must have the same gain.

4.1 - Measuring Amplifier Gain

This is quite easy to do, but you do need a single frequency stable tone source - music is of no use, because it is too dynamic so levels are constantly changing. However you might consider the use of a test CD, which will have various frequencies at predictable levels. In many respects, this will be easier to use than any other method, since it requires only that the CD is inserted, rather than dragging oscillators or other signal sources about.

Unless the amplifier is a valve unit, it is not necessary to have a speaker load connected for these tests, or a suitably high-powered resistor can be used as a load if you want.

If a known level (say 100mV) is injected into the input of the amp you are going to use for bass (for example), you will measure an output voltage of about 2.5V at the amplifier output (this represents a fairly typical gain of 28dB). This must be identical for the amp being used for mid+high - assuming that the speakers have the same sensitivity. If the gains are not the same, you must install a volume control on the amplifier whose gain is the higher, and adjust until both amps produce exactly the same voltage at their outputs for 100mV input.

For the test frequency, use an oscillator at about 400Hz, or if you don't have an oscillator at all, you can use the attenuated output from a small power transformer. This will not be as good, but it will work. The frequency will be either 50Hz or 60Hz, depending on your local supply. (If you don't know how to use a transformer to do this, ask someone to help - you can easily damage the input stage of the amp (and the rest of it !) if the level is too high). Alternately, use a test CD as mentioned above.

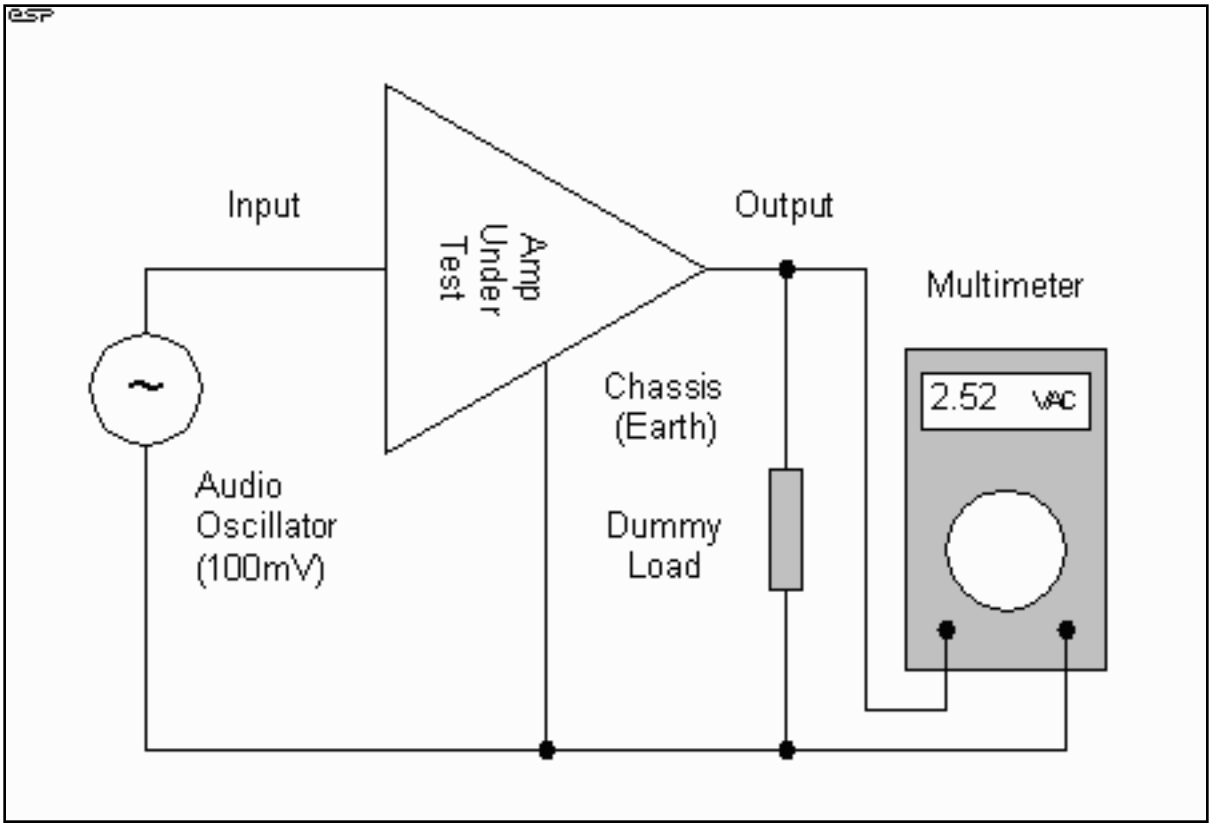


Figure 6 - Test Setup For Gain Measurement

The dummy load resistor should be equal to the speaker nominal impedance, and be rated at 5 to 10 Watts. Do not attempt to operate the amp at full power (especially if rated at more than 20W) into the load, or it will get very hot indeed. If you want to do this, then the resistor power rating should be at least double the expected amplifier output power. (Either that, or use lower power resistors and drop them in a bucket of water - it will not cause a short circuit, fear not).

The voltmeter used may be digital or analogue - as long as it can read the test frequency accurately.

4.2 - Speakers With Different Sensitivities

This is where things start to get a bit tricky. You will need to be able to calculate the required gain to suit the speaker sensitivities - not hard, but you might find that the scientific mode on the Windows calculator is useful (unless you already have a full scientific calc, of course).

Depending upon the crossover frequency, you might need to use a higher powered amp for the bass end, if its speaker has a lower sensitivity. For the purposes of the exercise, we will assume that the midrange (plus high frequency) speakers have a sensitivity of

90dB / W @1m as before. But the woofers have a sensitivity of 88dB / W @1m so we need to calculate the power and gain differences (assuming that the "equal power distribution" frequency of about 300Hz is being used - you want to use a different frequency? If you follow these procedures, you will become an expert at this stuff - guaranteed - because you will have to determine the relative power levels for the crossover frequency you are using - and I'm not going to help !

First, we will calculate the gain difference. Assume that the mid+high amp has a gain of 28dB, so the bass amp needs a gain of 30dB (the speaker is 2dB less efficient, so we just add the 2dB to the 28dB of the mid+high amp).

We will use the same 100mV input signal, so:

Gain = antilog (dB / 20) = antilog (30 / 20) = antilog (1.5) = 31.623

Since we started with 100mV (0.1V), the output voltage must be 3.16V from the output of the amp. That wasn't so hard. Now we need to determine the power output of the bass amp, if it is to exactly match the mid+high amp. Let's assume that we will use a 50W amplifier for the mid+high (with 28dB of gain).

P2 = antilog (dB / 10) * P1 (where P1 is the known power, and P2 is the unknown (higher) power)

P2 = antilog (2 / 10) * 50 = antilog (0.2) * 50 = 1.585 * 50 = 79.25W

Note that with the power calculation, the value of 10 is used, rather than 20 for voltage or current calculations.

We have now discovered that an 80W amplifier is needed with a gain of 31.6 (30dB), to exactly match the amp power and speaker efficiency of the mid+high combination.

Ah. So you have measured the amps, and have an output voltage, but cannot relate him to decibels. Fear not, another formula is at your disposal:

dB = 20 * log (V1 / V2) **(where V1 is the higher (i.e. output) voltage and V2 is the smaller)**

So if you measure an output of 2.32V at the output of the amp with an input of 100mV, its gain is

dB = 20 * log (2.32 / 0.1) = 20 * log (23.2) = 20 * 1.365 = 27.3dB

Note that in all calculations I have rounded the values to 3 decimal places, but when you do the calculations, retain all decimal places available for best accuracy. The difference is not great, but there is no need to introduce inaccuracies for no good reason.

4.3 - The Effect of Amplitude Inaccuracies

To see what happens when the gain is not correct, we need to look at the crossover curves again. Refer to Figure 4 again. This is the optimal frequency response of the crossover/amplifier/speaker combination, with the resulting output being virtually flat (there is a slight rise at the crossover frequency which can be corrected using Linkwitz Riley alignment, where the crossover point is 6dB down - see Project 09)

Have a look at what happens when the amplitude of one filter is different from what it should be. This is shown in Figure 7, and it is clear that the crossover frequency is shifted. What is not so clear is the final frequency response, and in the case of any crossover filter that is not phase-coherent, the adverse effects of the relative phase relationships. These are extremely difficult to quantify, but may be apparent in listening tests. The problem is that if you are unaware of the problems that can be created by modifying gain indiscriminately, it will be very hard indeed to determine why the system just doesn't sound right.

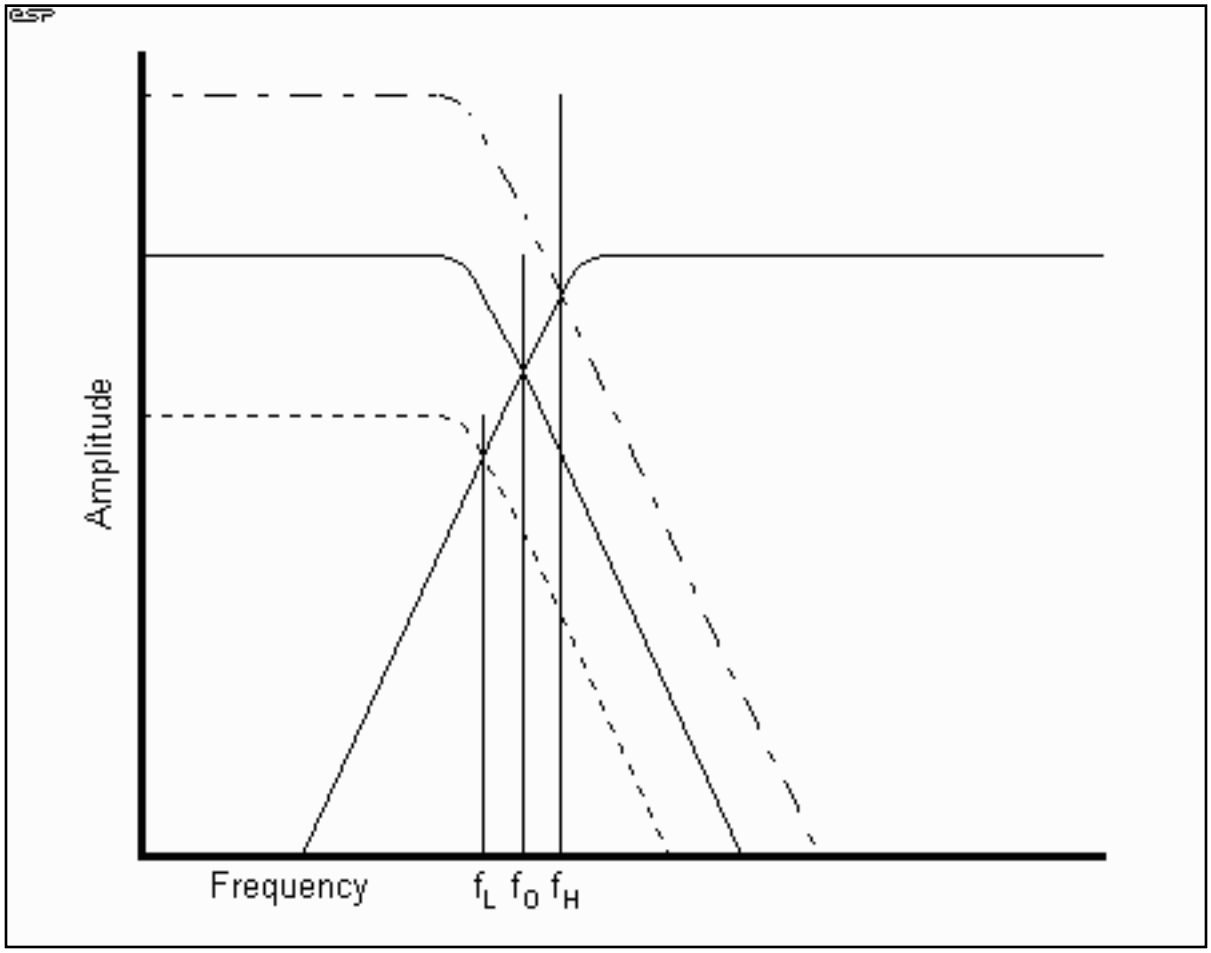


Figure 7 - The Effect of Changing Gain on Crossover Frequency

It is very obvious from Figure 7 that the crossover frequency has changed. At normal gain, the crossover is 295Hz, but if the gain is increased as shown, the crossover frequency shifts up to 500Hz. If the gain is reduced, the effective crossover frequency is now about 150Hz. Naturally, the same thing happens if we change the mid+high gain.

This is not just the output of the filter we are looking at, but rather the final output from the speakers - as shown above, it will often be necessary to change the gain of amplifiers to match the efficiency of the loudspeaker drivers used. This does not alter the crossover frequency as you might expect, but brings it back into proper alignment. In fact, if the gain is *not* changed, then you will get a result similar to that shown, by effectively amplifying one frequency band more than it should be for the correct tonal balance.

4.4 - What Does All This *Really* Mean?

The crossover / amp / speaker combination has to have the correct gain structure if a flat response is desired, and any variations can be quite audible. The audibility varies with

the type of music, and depends a lot on your hearing. In some cases, a slight unbalance can sound better than a perfectly flat system, and can be used to compensate for room influences or personal preference. Some passively crossed loudspeakers have a L-Pad level control for the tweeter, although these are a lot less common that they once were. If available, this does the same thing as changing the amp gain in an active system.

With a phase-coherent crossover, I have found that the ability to use the crossover gain controls as a "tone control" seems to work fine, and there are no real anomalies that I have heard (apart from the obvious prominence of the louder frequency band). This is something I often do with my workshop system. This unit runs from my tri-amped, phase-coherent 3-way variable impedance test amplifier (that's a mouthful). I am forever fiddling with the gain structure, amp output impedance and crossover frequencies, and although not ideal (although it has come close with some of the loudspeaker drivers I have there), it nearly always manages to sound much better than it has any right to.

With a conventional passive crossover network, the correct amplitude matching of the loudspeaker drivers is very important, but is usually fixed and cannot be altered. Even with an active system, correct level matching is not just to ensure a flat response, but to ensure that there are no additional phase problems created by the variations. There will nearly always be phase problems with passive crossovers even where the design is very complex, and these errors must create problems in either frequency response or overall accuracy (these don't necessarily coincide, although in an ideal situation they would).

Amplifier loading is another issue that cannot be ignored. The load presented should be essentially resistive (again in this mythical ideal situation), but in reality this is rare. Using electronic crossovers and separate power amplifiers alleviates the issue somewhat, especially for the mid+high section, but the variable load posed by a typical bass reflex type enclosure is always going to cause amplifiers (minor) problems. Reducing these problems by eliminating a crossover network (or part thereof) can go a long way towards improving the load seen by the amp, and potentially reducing amplifier intermodulation distortion (in particular). The reduction is largely brought about by making the amp work over a more limited frequency range than normal.

In short, there is no reason for a DIY enthusiast to avoid biamping, since the costs involved are very much lower if you make the crossover and power amplifiers. In addition, it is possible to obtain results that are startlingly good, without the considerable difficulty and expense of tweaking a passive crossover network. The latter can easily cost a hundred dollars or more, and if it's not quite right one can easily spend another hundred dollars trying to get it to sound like it should. By comparison, an active crossover costs between nothing and a few cents to change. Need I say any more?

Part 1

Main Index

Articles Index

Copyright Notice

This article (including all images and diagrams) conceived and written by Rod Elliott. Copyright © 1998 - 2005 all rights reserved. Reproduction, storage or republication by any means whatsoever whether electronic, mechanical, or any combination thereof is strictly prohibited either in whole or in part without the express written permission of the author, with the sole exception that readers may print a copy of the article for their own personal use.

Some of the terms used in the descriptions of various design configurations may be registered trade marks. These terms (where used) are not to be taken as a reference to any particular product, company or corporation - they are used only in their generic or common technical sense and infer no affiliation with any third party.

Update Information: Page last updated: 02 May 05 - divided page into two parts, updated drawings and text./ 28 jul 01 - added some minor commentary, and links to other pages./ 03 Jun - moved to sound.westhost.com./ 16 Aug 2000 - added small explanation of low freq power needs, minor reformat of page, added most common question./ 11 Dec 00 - added small systems to the intro, and speaker damage box + misc text mods./ 09 Dec 2000 - added tabulated TOC, modified conclusion, a few minor additions (speaker sensitivity, ELF subs, etc.)./ 29 Nov 1999 - added entry to table, and some extra comment about power distribution./ 19 Nov 1999 - Included table of power distribution (provided by a reader)./ 28 Aug 1999 - Added new information to tweeter protection section (use of DC detector, etc)./ 19 Nov 2000 - added thermal details and corrected error in phase response of conventional xover.

Elliott Sound Products **Counterfeit Semiconductors**

Last Updated 21 February 2006

Main Index

The devices described here *must not* be considered comprehensive, as I am sure that there are a great many additional counterfeit devices available that have either not been discovered yet, or whose purchasers are too embarrassed by being caught by the fraudsters, and are unwilling to admit that they were robbed.

I would like to thank those who have sent information, photos or links to other sites about fake devices. Your continued vigilance will help the fight against these practices. It is a fight that we cannot win, but at least by knowing what to look for, and where the problems have been found, there is at least some hope that we can make it harder for the criminals involved.

Counterfeits Index

The following is a list of faked components identified so far. As noted above, there will be many, many more - these are the ones that people have advised me they have found. You can be sure that *any* high priced component is a candidate for counterfeit fraud, just as the case with clothing, aircraft spares, DVD movies, etc., etc.

- MJ15003/4 Power Transistors
- Sanken 2SA1216 (and presumably the NPN type 2SC2922) - Other similar 'Sanken' devices are also affected
- 2SA1302 and 2SC3281 (Toshiba) Found *all over the world!* The ones seen are all marked Toshiba.
- OP-07 Opamps

- 2N2773 Power Transistors (probable)
- NTE37 and NTE37 (I am told these are equivalent to 2SC2581 and 2SA1106) - not confirmed, but highly suspect
- LM3915 LED VU Meter
- Toshiba 2SA1943 (and presumably 2SC5200) - this had to happen, and was only a matter of time.
- 2N3773 - Allegedly made by someone calling themselves 'MEV'
- MJL21193/4 ON Semi MJL21193/4 (branded Motorola) **Update - 21 Feb 2006**
- 2SA1386 and 2SC3519 Sanken (branded IEC)

... and just in case you thought that this was a conspiracy affecting hobbyists, I suggest a web search - the cost is in $millions, and is spread world-wide! Almost anything that can be faked is being faked. I'm sure everyone has received innumerable e-mails from spammers offering cheap watches, software, drugs, etc. Almost without exception, the goods are of poor quality, and many may not work at all (prescription drugs are a particularly worrying problem - many have zero active ingredients, placing people's health and lives at risk).

The Main Story ...

So, off you go to the local parts shop to buy some transistors (or indeed, other parts!). Having decided the devices that fit your needs (having selected any of a number of devices that are suitably rugged and powerful), you hand over a not insubstantial amount of cash and head home to build the masterpiece. Inexplicably, the expensive output transistors fail, but you know that their ratings are well within the design limits for the project you are building. This happens once, maybe twice, or perhaps more. You get discouraged, and shelve the project - having already spent quite a lot on all the parts needed.

Even worse, during testing, the transistors are (or seem to be) fine, only to fail later taking your expensive speakers with them. Now it is *really* serious.

So is the problem a bad design? Very probably not. Counterfeit power transistors are not only available (again!), they are rife in the industry. Don't bother re-reading that - you saw it correctly. Counterfeit power transistors !!

The first instance (that I know of) was in 1980, when MJ15003/4 transistors were sold under the brand name 'TIC'. These, and many similar counterfeits were in fact 2N3055

and MJ2955 aluminium cased devices, and the counterfeiters had removed the original markings and screen printed the fraudulent type numbers on the cases. Why? Because 2N3055 and MJ2955 devices are cheap, and genuine MJ15003 and MJ15004 transistors are not.

All over the world, people buy power transistors with amplifier kits or by themselves (for example to assemble any of the ESP designs), and they fail. One reader in New Zealand was caught with fake 2SA1302 & 2SC3281 devices, causing his P68 subwoofer amp to fail, and similar stories surface all the time.

Counterfeit transistors cost far more than their monetary value - loss of confidence, time, 'collateral damage', etc. are far worse. I hear these stories typically about once or twice a week, and the only difference is that a new target for the counterfeiters has been found.

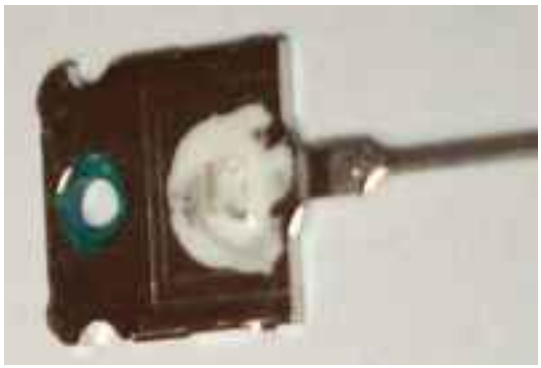
In most cases, I doubt that the supplier has deliberately supplied fake devices, but the fact remains that they have counterfeit stock and people have been affected. All suppliers should, nay **must** be advised immediately if you discover counterfeit devices. It is most unfortunate that so many suppliers have their heads buried so far in the sand (or see below) that they will not believe their customers. It may help to direct them to this page, but some are so far gone that they refuse to believe that counterfeiting even happens!

Supplier Reaction to Counterfeit Claims

For every e-mail I get on this topic, there will be hundreds of other people who don't

know about this page, and who think that the mistake must have been theirs - this is a sad situation indeed!

Fake MJ15003 - Typical



Fake 2SA1302 / 2SC3281

The photos above are indicative of what you will find inside a couple of typical fakes. The MJ15003 has two dies, each about 3 mm square, and simply wired in parallel. As can be seen, there is no heat distribution "coin" - they are simply bonded to the steel case. Motorola (or ON Semiconductor) has never made a power transistor in this way (to my knowledge, or that of anyone else I've spoken to about the fakes).

The situation with the 2SA/C devices is a little less obvious. The silicon die is just visible under a layer of silicone (the emitter and base leads came away with the epoxy, which is why you don't see them). The die is again 3mm square, and is bonded directly to the metal tab. In this case, the tab is of copper, so that part is acceptable. Unfortunately, I don't have an original device for comparison, so if anyone does (preferably a blown one - they are too expensive to sacrifice), I would appreciate a photo or even a description of the insides.

Vigilance

The only thing I can suggest is that you exercise extreme vigilance when purchasing semiconductors, and especially the premium devices. If they are normally expensive, then they are ripe for counterfeiting, since the potential gains for the criminals behind these rackets are very large indeed. Always beware if you see normally expensive devices being offered for sale much cheaper than normal. This is usually a good indicator that something is wrong - and in such cases, the supplier may know full well that he is selling sub-standard parts. This is a criminal act ('intent to defraud' or similar) in most countries, and should be reported to the police.

For many resellers, they are tempted by the low wholesale prices (and may not be

aware that fakes even exist). They see a way to maximise their own profits, . The backlash is that their customers will be bitterly disappointed, and will very likely take their custom elsewhere - a classic no-win situation.

Be more than careful with devices offered at auctions. Not all will be fakes, but you can almost guarantee that a fair proportion are counterfeit. I have had e-mails from a number of people who have purchased semiconductors at various on-line auctions, and the results were entirely predictable. The devices bought were fakes, and there is little or no recourse with an on-line seller who can happily disappear after unloading the goods.

Even some of the more established sellers will (inadvertently or otherwise) offer counterfeit components. Don't assume that semiconductors are the only parts that are affected - a great many polypropylene capacitors are nothing of the sort (they are polyester or mylar), and virtually anything that can be made to look like a more expensive part will be. Even though the individual gain may be small, if an unscrupulous dealer can make an extra ten cents on 10,000 parts, this represents a very worthwhile profit from their perspective.

Counterfeit components are not only those that are re-marked with a different manufacturer's logo and/or part number - no-one has been able to categorically deny that 'bad' batches of parts (where something went wrong in the manufacturing process) are destroyed. The most likely situation is that they are auctioned off as scrap, except the scrap metal merchants may well see a golden opportunity to make a lot more than the would by melting the parts down for their metal content.

Does this actually happen? I honestly don't know, but I would be much more surprised if it does not occur than if it can be shown that it is normal practice (in some countries at least).

To All Resellers - World Wide ...

If you discover that you have been supplied with fake components, come clean and admit your mistake. Otherwise, be willing to prove to your customers that you have never stocked counterfeit components. Please! This is not a big thing to ask, and will go a long way to showing that you value your customers and their custom. If not, you will be seen as the baddies, and rightfully so if you continue to defraud your customers!

All suppliers should be able (and willing) to provide the customer with an assurance that parts have been supplied by reputable wholesale outlets, preferably the manufacturer's

appointed distributor. This applies equally to independent distributors (brokers) who act for equipment manufacturers with 'distress' stock (i.e. stock that will not be used in production), and who supply devices from many different semiconductor suppliers. I suspect that these brokers are one of the main problems, because they will rarely (if ever) be in a position to supply the necessary paperwork to prove the device's authenticity.

This does not mean that no-one should ever use their services, but all end users and retail/ wholesale suppliers must be vigilant to make sure that no fakes are substituted for the real thing. No-one wants to be stuck with thousands of counterfeit components, so they tend to circulate world-wide until they all eventually blow up. The cost of counterfeits to everyone involved is high, and the sooner everyone is aware of the problem, the better.

When discovered, fakes should *always* be destroyed. This will help to prevent circulation, so that the overall impact is reduced. Unfortunately, no-one wants to do that because destruction represents a complete financial loss. Few suppliers are in a position (or will be willing) to suffer the loss - even if it was their own stupid fault for not checking the pedigree of the stock they purchase.

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000-2004. Reproduction or re-publication is allowed due to the importance of ensuring that everyone should be aware of these fraudulent practices, on condition that the name and URL of the original page and author (Rod Elliott) of the information herein is not removed or replaced.

Page created and Copyright (c) 14 June 2000 Rod Elliott./ Updated Apr 2002 - moved section to its own page./ 04 Feb 2006 - separated pages into sub-sections.

The Audio Pages

Elliott Sound Products **ESP Educational Projects**

Copyright © 2003 - Rod Elliott (ESP)
Page Created 30 Apr 2003

Articles Index

Main Index

Introduction

Technical colleges, schools and universities - ESP can design and supply educational kits, specialising in analogue electronics projects. Many of the featured projects have been used as student projects at educational institutions all over the world, but specialised projects designed to supplement specific parts of the curriculum can be developed, and either complete kits or just the blank PCB and full instructions can be supplied.

This is a new service from ESP, and I would be delighted to quote for any customised schematics, PCBs or complete kits, for almost any analogue or simple logic based project. As may be seen from the extensive range of projects and articles on The Audio Pages, there are a great many possibilities.

There is a big difference between knowing and understanding, and only by working with things at their most exposed level (discrete circuitry) does a student gain true understanding. A great many projects use ICs - sometimes highly specialised - the student may know how to use that IC (or similar ICs), but gets no understanding of the basic principles. With the passing years, the basics are being lost, with many engineers being unable to even bias a transistor for linear operation.

Examples

For more information and full details, please refer to the projects index - as can be seen, there are projects for construction of ever popular audio amplifiers and preamps, simple mixers, test equipment and lighting applications.

ESP is in a unique position to be able to offer anything from a complete service to just a design and/or a PCB to suit, along with comprehensive documentation. There are very few educational kit suppliers, and most are targetted at complete beginners - ESP can supply educational services for complete beginners, right through to advanced levels.

Some good examples of educational kits are ...

- Transistor and FET Circuits - biasing, different topologies - common emitter, common collector, common base, cascode, compound pairs, darlington pairs, direct coupled circuits, etc. Essential basic techniques for linear operation.
- Voltage Regulators - IC and discrete designs, where the student can learn the basic principles of series and shunt regulation, error amplifiers, series pass transistors, low drop-out circuits, etc. All of these devices are available in IC form, but the inner workings are hidden from view, thus preventing a full understanding of the principles of operation.
- Switching Regulators - again, using both IC and discrete designs to ensure that the student has a complete understanding of the principles of boost and buck regulators, switchmode power supplies, feedback principles, stability and efficiency.
- Operational Amplifiers - IC and discrete circuits, demonstrating voltage and current offsets, biasing, stability, feedback techniques (voltage and current) and output impedance. Filters, oscillators, comparators, level shifting circuits, precision rectifiers, current sources and sinks - the range is almost endless
- Analogue Servo Systems - simple servos that show the importance of servo gain, dead-band setting, etc. Stability and accuracy issues can be seen first hand, and the solutions investigated. Excellent introduction to electro-mechanical systems in general.
- Rectification - using basic rectifier topologies to gain an appreciation for the benefits and disadvantages of different topologies (e.g. half wave, full wave, bridge, voltage doublers, voltage multipliers, etc.).
- Mains Control - phase control using TRIACs and SCRs, zero voltage switching, linear control of AC systems, motor speed controls (servo and open loop systems), etc.

The above is a small sample of the possibilities, and there are obviously many more.

One of the difficulties of educational kits is that they are usually designed with a "one size fits all" approach, and this limits their usefulness. It is also important that the circuits demonstrated should do *exactly* what was intended - they must work flawlessly if properly assembled, and have a wide tolerance for normal component variations.

Further Information

Should you feel that well designed projects to reinforce the theoretical aspects of the curriculum would be helpful, please do not hesitate to contact me with details of your requirements. Having taught electronics, I know how important it is for students to have hands-on practical experience. Although discrete circuitry may well be considered "old hat" by the students of today, they will have a far, far greater understanding of basic principles if they are exposed to them in a way that makes the subject not only interesting, but *fun*. In my experience, people learn faster and better if the subject is fun - it greatly improves attention and (more importantly, retention), since people's minds don't wander very far when they are having a good time.

To contact ESP with any enquiry, please go to the Contact ESP page for details. I look forward to helping you to improve the education of our future engineers and technicians.

Articles Index

Main Index

Contact ESP

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 30 Apr 2003

The Audio Pages

Elliott Sound Products **The State of Manufacturing**

By Rod Elliott
(With additional material by Fred Newton and Mark Hammer,
as well as other material suggested by various readers)

Introduction

In one of the sorriest news items I have heard for some time, it was revealed that job losses in the manufacturing sector were the highest of any industry in Australia. Australia is not alone in this, as major companies and corporations rush to Asia to get the lowest prices they can for manufactured goods.

This push is partly because everyone wants their consumer goods to be cheap, and the labour content is the thing that (allegedly) pushes up the price. Cheap goods are not necessarily sold cheap, of course - profit maximisation is another reason to seek the lowest labour rates you can find, while selling the product at the same price.

By way of a direct example, a recent story in the newspaper ... A major Australian airline is getting uniforms made overseas and it saves them $1M / year - supposedly. The effect of this locally is that some 300 people will lose their jobs, so at even the most basic wage, that works out to around $6M in lost wages, plus flow-on effects. Sounds like a real bargain to me :-(

This same airline is now talking about outsourcing their aircraft *maintenance*! Suggested locations are New Zealand, China and Indonesia. I shall leave it to the reader to see the many flaws in this idea, but the world's safest airline is unlikely to remain that way if this lunacy is allowed to continue. It is incomprehensible that anyone could be so bloody stupid to think that this is a good idea.

Is Anyone Thinking?

It seems that none of the major corporations, and even less our governments, are looking to the future - what happens when a country has little or no manufacturing

capability left? This is already happening, and I'm sure that everyone who reads this will have knowledge of a factory (large or small) that has been shut down because the product(s) are being made elsewhere.

So, when we have closed all but a few small-time manufacturers, when the machinery has been sold off (or scrapped!), when everything is being made in the "Far East", where is our independence? What happens if there is a major crisis, and goods (for whatever reason) cannot be obtained? Worse, what will we do when (not if, when!) the cheap labour that we currently enjoy says "Sorry, but we are forced to increase our rates. All rates are now double." The chances are that they will double again soon, and we have absolutely no choice.

There is no competition, since we no longer have any factories capable of making the products, we no longer have the machinery needed to rebuild the factories, and we no longer even have the factories to make new machinery.

Greed

The 'Captains of Industry' (and that is meant as an insult), the stock holders and we, the consumers, are basically greedy - despite what was claimed in the '80s, greed is NOT good - it is selfish and self destructive. In the rush to be cheaper than everyone else, or to have higher profits than everyone else, or maximise profitability and 'shareholder value', most of the Western countries have blithely thrown out the baby with the bathwater.

This is serious. It will get worse. The electronics industry in Australia is decimated, ruined by incompetence and stupidity. Most other manufacturing is in the same position - the clothing and footwear manufacturing industries have been beaten to death, steelworks and shipbuilding yards have closed, and more disappear every day. Both governments and industry alike are to blame for this - and most Western countries are either the same, or heading down the same path of destruction.

Australia (like almost every other country) is still blessed with a few small 'boutique' manufacturers making whatever they can, and usually doing it well - despite government. Is it any different in the US? Europe? It may not be as bad in some other places, but it is certainly unlikely to get better. Very few governments are smart enough to be able to see the inevitable, and the major corporations are just as bad, if not worse. Would these boutique manufacturers be able to fill the void if the cost of all imported goods suddenly doubled? No, of course they couldn't, because they don't have the

machinery, the space or the capital, and even worse, they would be unable to find people to use the machines even if they did have them.

A Definition of Manufacture

It is time for all countries to look at manufacturing differently - manufacture is creation, it is a noble activity, taking raw materials and making something that wasn't there before. The whole concept of 'blue collar' work is wrong - it is not something to be looked down upon and disdained, it should be glorified and revered (Ok, that may be stretching the limits a little, but you get the idea :-)

We all need things to be made. Very few people could make their own car or washing machine, even if they had the tools. Only a relatively small number of people can make their own clothes - even with a sewing machine. The skills that we are losing are irreplaceable - it's not the tools or the materials that make manufacture what it is, it is skill. The ability to use a machine to make something is not 'manual labour' or 'production work' - it is a skill. The fewer factories we have, the fewer skilled people are employed, and the lower the demand for training - this is a self-fulfilling prophesy, and eventually there will be virtually no technical colleges teaching manufacturing skills, because there is no available employment in the industry, so there will be even less demand for the training.

It is the lack of skilled workers more than anything else that will be our undoing. There will be no-one left who has the knowledge of a production environment, the ability to use the machinery, or the logistics of a manufacturing plant.

Cascade Effect

The rot spreads much further afield though - if no-one is making anything, then the engineering suppliers (as well as metals and plastics suppliers) disappear as well. These are the people who would normally stock the raw materials, tools, drill bits, and other strange looking tools that most people have never even seen or heard of, that make manufacture possible (even on a small boutique scale). As these essential suppliers disappear, then it becomes more difficult for anyone to even contemplate starting a small manufacturing business. Large businesses grow from small ones, and these days they don't stand a chance.

The process continues all down the line (as always), and until we wake up to ourselves

and start encouraging manufacture, the situation will become critical - or more critical than it already is.

Large hardware (and electronics supply) stores used to stock things that people would need every so often - in other words, the stock would sit and gather dust until someone needed it. The bean counters don't like that - stock has to keep moving or it's dead. The remaining stuff that is useful on a daily basis is retained because it makes money - it keeps moving, but where does that leave anyone who needs a 19mm twist drill? Out in the cold, is where (and yes, this is a recent attempted purchase on my part - fortunately, a 3/4" drill bit is 19.05mm so I could use that - 3 hardware stores later I found one!).

Needless to say, if one is in need of lathe tools, or a local aluminium anodiser, then you are seriously out of luck. This is stuff that should be available, and if there was a vestige of manufacturing left (other than kitchens and bathrooms - they are everywhere, and don't really count - sorry) then these items/ services would be available.

Years ago, I could get aluminium cut and bent to the shape I needed, then anodised ready for installation of the components that made it into an amplifier. I still can, but I will have to travel a great deal further, have much less choice, and will pay considerably more (allowing for inflation) than I did before, simply because the demand no longer exists - because the manufacturers who used these supplier's services have gone!

This is not about me, or anything I may be doing. This is a general malaise that has struck down a vast number of manufacturers, both small and large.

The DIY Fraternity

Since my pages are devoted to hobbyists and DIY enthusiasts, it is worth mentioning that the DIY fraternity is capable of making a big difference - why buy something that was made somewhere else, uses proprietary parts that you can't get, is designed so that you can't even get it apart without the inside knowledge of the maker, when you can build it yourself.

Add to that the satisfaction of having made it, and the fact that you have acquired new skills and knowledge (which sadly will not help you get a job). The net result is a win - for you because you have something that you made, for the suppliers you deal with because you and your ilk help keep them in business, and for the country, because that is money that has stayed put, and not gone somewhere else across the sea.

This applies to every field of DIY - from wood turning, cabinet making, electronics, knitting - you name it. It is all local content, saves you money, and gives great personal satisfaction. If enough people do it, the difference to the economy and the community can be enormous. The only losers are the large corporations and their off-shore manufacturing plants.

Locally Made - Not Likely

It is time that governments realised what their policies are doing to local industry, and encouraged people to make things, by supporting the local product (not that there is any), and by offering incentives and even inducements ... "We need 5,000 WhatNots - locally manufactured products will be preferred, even if the price is higher."

The taxpayer obviously pays for this, but look at the advantages to the taxpayer - the more industry there is, the more money flows through the economy, and everyone benefits. There is always a flow-on effect, and it is high time that the direction of the flow is reversed.

An example is a small local manufacturer who made parts for commuter trains (this is a true story, but the actual items are not named). Despite assurances from government that local suppliers were preferred, he lost the contract to China. They saved less than AU$1 on each item, which was probably worth an average of about AU$200, and that only got the items landed in Australia - the customer still had to collect them at their own expense (the local supplier's cost included shipment to the works where the trains were built). The local manufacturer is now out of business, with the loss of half a dozen jobs and several lifetimes of accumulated knowledge. This is insignificant in the greater scheme of things, until you understand that it is happening every day - every day a small manufacturer (or several) will go out of business. Lost jobs, lost opportunities, and I'm lost for words!

Meanwhile, governments like to tell us about the 'level playing field' they have created by removing tariffs, and opening our markets to competition. Yes, the markets are open, but the playing field is not at all level. Subsidised transport (etc., etc.) is just one area where other countries can undercut the local product. This so-called 'level playing field has seen the demise of more farmers and manufacturers than ever before - somehow, I don't consider that to be level in any possible interpretation of the word.

Interdependence

We all depend on each other, and your business could not survive without other support businesses - but we are being stripped of one of our most important assets - independence, not from each other, but from the rest of the world. We need trade, but as support, not wholesale replacement of our own capabilities. In times of crisis, we all need to be able to do things on our own, since our major supplier may be the country in trouble (or in extreme crisis, even our enemy).

The oil situation is a perfect example - not everyone has oil to extract, of course, but the outcome is just as predictable with any commodity, whether manufactured or not. If war were to be declared against the country where your car/ fridge/ hi-fi/ dishwasher (etc.) was made, how would you get parts? Would you be considered a traitor for owning something made by 'the enemy'. If no-one in your country made the item in question, what would anyone be able to do without local manufacture? Go without? That is not something we are used to in our society, and it would cause much complaining - and very likely, those who would complain the loudest would be the ones who allowed it to happen in the first place!

Standards

The insistence on 'standards' has ruined a great many small manufacturers too. Not because their products are inferior, but because they have not achieved the required ISO stamp of mediocrity. The standards to a large extent regulate the way you do business, but not how well (or otherwise) your product performs, its build quality or longevity.

This is not to say that standards should not exist - they do help to keep the large corporations (marginally) honest. The small businessman needs integrity to stay in business, not standards. The way the business is operated (provided it is legal, of course) is of no concern - if the employees are happy, the goods are of high quality and the customers are happy, then what possible additional benefit can ISO9000 compliance produce? It certainly won't make the products any better, but it is likely to make them cost more.

Safety standards are another matter, and are in place in just about every country in the world. Daily newspapers and consumer magazines have legions of product recalls - very often because the product was sold without having been tested against the safety standards, or is defective. ISO9000 should prevent that from happening, but it doesn't.

One must also consider the integrity of the overseas supplier. Fake watches, clothing, transistors, aircraft parts and many, many more abound. Is a supplier who may be party to making fake products going to *really* care that the UL, CE or Australian Standard tests have all been properly conducted? Of course not. If they are found to breach the standards of any country, no-one can do a damn thing about it. The factory just changes its name and continues on as if nothing has happened. Countless safety breaches and non-compliance with applicable standards can be found on the Net, and they will get worse, not better.

Safety and the Environment

I have been advised recently (from someone who was working is Asia) that there is another side to outsourcing to developing nations. While I've also seen it myself, that particular penny hadn't dropped when this article was first written.

Most 'western' societies have very strict OH&S (Occupational Health and Safety) standards, and employers are expected to provide a safe working environment for all employees (including sub-contractors). Breaches cost companies dearly, both in bad publicity and financially - heavy fines are imposed for breaches, especially if someone is injured or killed in the workplace.

Likewise, companies that produce waste are expected to dispose of it in an approved manner, and this area is tightly regulated. Breaches can be very costly, since here too heavy fines are imposed for breaches, and the publicity can be particularly damaging. Anything that is even slightly toxic must be handled, stored and disposed of in a manner that is set down in the law. For example, spray painting booths must not allow any 'significant' solvent or particulate matter to escape, and there are very strict regulations that dictate exactly how the solvents and particles shall be filtered, and any waste disposed of.

Naturally, we also have minimum wage standards, compulsory superannuation contributions for employees (only so the government won't have to pay pensions in our latter years, but that's another story), etc. These are all burdens on local industry, but they are necessary to ensure a reasonable living standard, and to prevent screwing up the planet any more than is deemed necessary at the time.

Many of the 'cheap labour' countries do not have *any* of these schemes in place. OH&S is up to the individual - basically "you get hurt and can't work, you don't get paid". Superannuation? You must be kidding. Waste disposal? The recent massive toxic waste

spill(s) in China tell us what we need to know about that. In short, all the things we take for granted are missing. The workers are paid a pittance by our standards, but to them, it's very good money. Why would the employers make life any harder for themselves by employing best practices for OH&S, superannuation or environmentally friendly waste disposal?

Let's face facts here - most of the western corporations would quite happily dispense with these 'unwarranted exploits of their profitability' if they possibly could. Our governments have *imposed* these things upon them, very few are voluntary. Wages and conditions have been fought for over a long period by unions, and although it must be admitted that some union claims are outrageous, many of the benefits we now enjoy were the result of prolonged battles between unions and employers. No employer wants to pay any more than absolutely necessary, and every government or union demand is seen as a threat.

Since our scumbag government recently castrated our labour laws, we have seen a rash of unscrupulous employers 'offering' current employees their own jobs at a reduced wage - or risk the jobs going offshore. This is simply blackmail, and we have lost a significant number of manufacturing jobs recently to cheaper overseas alternatives. Unions in general have nipped such tactics in the bud in the past, but it is getting harder for employees to negotiate a fair deal - and this with the government's assistance!

This is not to say that all union claims (or their own blackmail tactics) ultimately benefit the workers - far from it. Nor are all government impositions well thought out - think of the European Union's 'ROHS' directive (restriction of hazardous substances), and the lead-free solder debacle!

Ultimately (and this is really the whole point), when a critical mass of jobs has gone overseas, the original countries (Oz, US, UK, Europe, etc.), will also lack the critical mass of purchasers. At around that time, the economy simply implodes.

I know the above is a bit of an over-simplification, but the net result is that economies *will* implode, and it will happen sooner rather than later if governments don't act. The signs are already present ... record numbers of bankruptcies, huge debt per head of population, more and more jobs disappearing overseas, and even migrant workers brought in to 'solve' the skills shortage. Why do we have a skills shortage? Might it be that people can see the writing on the wall for manufacturing anyway? Perhaps because the government has done *nothing whatsoever* to encourage apprentice employment? Maybe the consumer society in which we live considers 'manual labour' to be beneath

us? All of the above.

To an extent, we might say that the (major) retailers are true villains in this, but we are also to blame (we believe their advertising drivel, after all). Promote the consumer society, push the latest 'must have' models, squeeze the local manufacturers until they are no longer *capable* of bleeding, and sell, sell, sell (at all costs). Buyers must be convinced that the CD player or TV they bought last year is *so* passé that it must be replaced at once, lest they be seen as 'last week' types.

The time will come when the buyers become sufficiently scarce, and the competition for the lowest price sufficiently fierce - regardless of how insanely low it might be - then the retailers will implode too (it has happened to several here already). As soon as people feel scared for their job security (what job security??), they slow down their spending. All we need is the critical mass ...

Consumerism

This article is not about promotion of consumerism (rampant or otherwise), planned obsolescence and new models every three months. That is the 'model' pushed by large corporations, endorsed by governments and suffered by everyone else. Consumerism is the very thing that has created many of the manufacturing woes we see, and is putting repair people out of business as well. Why would you get something repaired for $150, when you can buy a new one for not much more - or perhaps considerably less?

What happens to all the old ones - whatever they may be? In some cases they will be recycled, but re-use is vastly kinder to our environment. Re-use means using the parts to make something else, or making the product work again so someone less fortunate than ourselves can have one. The economics of this don't make sense any more, because the goods we buy are made to be made - in other words, they are built with the intention that they will not be repaired. The manufacturing process is much more difficult if you have to consider the poor service person who will have to work on it later, so that too is history.

The ever increasing levels of complexity are inevitable, since we expect a new appliance to work at least as well as the old one (not necessarily the case, of course), and new techniques can improve efficiency and the environment is all the better for that. But even if efficiency is improved, if the expected life is reduced, then the net gain will very likely be negative - it will always take far more materials and energy to build a new one than to repair the old, so the alleged economy is false.

As an example, a 10 year old product is still working, and is maybe 70% efficient. Used (say) twice per week for perhaps one hour or so, it uses 1kWh* of electricity per usage. The 'new improved' model has an expected life of perhaps five years, but is now 80% efficient, so uses 0.88kWh per use to perform the same work. 0.12kWh per use is saved (0.24kWh per week), but by the time it has failed, a bit less than 63kWh has been saved compared to the old 'inefficient' model.

* 1KWh - 1000 watts per hour. A 1000W heater operated for 1 hour uses 1kWh of electricity

Since we do not know the exact figure for the energy usage to make the new product, it is difficult to make a fully informed decision, but a reasonable guess is probably not far off the mark. If we guessed that 100kWh of energy were used to make the product (not too unrealistic), there is a net loss of 47kWh for the new model versus the old - it would have to last almost 8 years just to break even. This is not an improvement. The efficiency is completely false, and also fails to take into consideration the energy that was used to make the old one in the first place.

To give you an idea, the iron and steel manufacturing sector is the largest energy consuming industry in the world. As an whole, it uses some 10 - 15% of the total industrial energy consumption. To exchange an old (steel) product for a new one, it will need to last for a very long time, or be extremely efficient before there is a net gain. At present, it takes 350kWh per tonne to smelt steel, which may not seem like a great deal (0.35kWh per kilogram). If an appliance is built using only 10kg of steel, there is 3.5kWh just for the initial production of the base metal. It hasn't even been formed, rolled, cut to the required size or shipped yet.

When all other processes are added in the production of a typical appliance, the energy costs mount up very quickly, since energy is used in every process of manufacture - from the reduction of raw materials into the base material (be it steel, aluminium, glass or plastic), then the rework to shape, mould, paint and pack the product. Then there is distribution, delivery to your home, and disposal of the old unit.

Perhaps manufacturers should be forced to provide not only an energy efficiency rating for new products, but should also indicate the total energy consumption required to build it. This may not directly influence the consumer's decision, but it should, since we all pay for the emissions of power generating plants, by way of greenhouse gases and depletion of the world's resources.

This is not a new topic, but it is more relevant now than ever before. New is 'good' - even

if we have to somehow convince ourselves that it really isn't as bad as we first thought. We are bombarded with advertisements for the latest and greatest product in every category - we MUST have this new improved model (or what exactly will happen? Oh, we may be thought unfashionable ... well deary me! - Hardly a heinous crime in my books

Am I cynical? Of course I am - and you should be too. There is a great deal to be cynical about - we all know that our governments lie to us when it suits them, and look at the large corporations that have collapsed because their web of lies and deceit finally unravelled. Are the remaining ones honest and caring, looking after their customers and employees as they should? I don't have to answer that, because you already know.

The European Union has introduced laws for recycling, but there is already some evidence that these are flawed. Certainly the idea is good, but the idea is in the execution and compliance with the laws. We will have to wait to see if any of this actually works.

The Future

Unless there is a turnaround by big business and governments, the situation will just get worse and worse. The attitude of "why should I make it here, when (somewhere else) can make it for less" has to go. Not that anyone would want to deny the Asian countries their income, but not at the expense of our own independence.

If other countries want to manufacture products, then that is how it should be, but not at the expense of loosing our own ability to make the same thing. It may well be more expensive, but if the quality is equivalent or higher, then a great many people will buy the local product as a gesture of 'patriotism'. Of course, there can still be problems with this. Current labelling laws in Australia are seriously defective (especially in the food sector), and allow importers and retailers to gloss over the truth.

Most countries have had, at some stage or another, a protective tariff on imported goods. This was seen as restrictive, and disadvantaged the consumer who was forced to pay more. All true. In the process, the local manufacturer was able to compete, the protective tariff entered government coffers, and other forms of taxation could have been reduced (I think I must be dreaming now :-)

The fact is that local manufacturing capability is vital, not just to the economy, but for national security. If all (or most) manufacturing is done by a few countries overseas,

what chance does anyone have of surviving a crisis? With no local skills, factories or machinery, where will we turn next when the current major manufacturing countries decide to increase their rates - knowing that we have no choice whatsoever?

Supply and demand is about competition - a monopoly is never a good thing, since there is no-one to compete with, so the monopoly can do whatever it pleases, at whatever rates it feels like charging. We are very close to that situation for many of the products we take for granted - we can no longer make them ourselves, so rely on others to make them for us, blindly (stupidly) assuming that everything will stay the same, and they will be cheap and available forever.

This is not the way real life is - it never was like that, and it certainly is not about to change.

We will pay for the short-sighted and stupid actions of companies and governments, that much is assured. When will it happen? My guess is about 10 years (somewhere around 2010-2015), unless action is taken now. Real, genuine action, not a few half-arsed politicians paying lip-service!

Thanks to a friend in the US, who does consultancy for manufacturers, I can add his perspective to this discussion. Fred Newton supplied this very revealing insight into the real economics involved when manufacturing is moved off shore or across the border.

My Thoughts on The State of Manufacturing (By Fred Newton)

In the US, the actual 'touch labour' portion or the percentage of 'Total Cost to build' of most consumer goods averages in the 17-20 % range. As an example let's take one of the popular brands of computers that is currently being marketed on a global basis. With an average delivered C.T.B. (cost to build at the factory level), price of US $350-$400. The portion of this that is actual touch labour is about $23.00. This is 5.75- 6.5 % of the shipped cost. The remaining cost is distributed between overhead, materials, and warranty, with materials on this specific product generally having the higher remaining percentage.

Now let's relocate this to a eastern rim country where the $23.00 decreases by $18.00 to an equivalent US$4.85 C.T.B. and the touch labour percentage of the new C.T.B. changes to 1.28 - 1.4% of the shipped cost. The new shipped cost is US$332 - $377.00.

Digging deeper into these numbers shows an average decrease in actual C.T.B. of 3.47% to 5.1%. Not a big number by itself, but there are other MAJOR costs that have not been discussed here so far. (Sorry for the economics lessons, but this dissertation is VERY important to all of us on a global basis!)

An average plant relocation from the US to an Eastern Rim country or to Mexico is in the $Hundred of Millions range. One recent plant relocation cost was well over US$200M. So if we begin to look closely at these large, no - *huge*, relocation costs we quickly begin to wonder where the profit is going to come from! At a marginal increase in profit from the slightly decreased labour cost it will require the relocating company to build and sell more than 11 million units to pay off the cost of the relocation, and start making a profit.

Now without going a lot deeper into this in numbers, it will take the company an average of 2-3 years of full production just to break even, and then only if the volume of products at the same quality level exists. If you add into the equation the required learning curve for the new location assemblers the 2-3 years changes into 3-4 very quickly. So where is the profit?

New Mass Market

In most countries around the world for a product to be sold there a certain percentage of that product is required to be manufactured there. With many of the more highly industrialised countries having had the current technology products available to them for the past twenty or more years, these markets are becoming saturated. The forecasted annual percentage of sales is MUCH lower in these replacement markets than for countries like China and Mexico where less than 1% of the population owns any of the modern products they are now manufacturing.

So ... to reduce all of this down to one line, the companies are relocating their products to the new MASS markets, and couldn't really give a damn about the livelihood of the country and its work force that allowed them to grow large enough to be Globally positioned as they are today!

This feeding frenzy is solely for the primary stakeholders within each of the organisations. Enron and the glut of centralised wealth that STILL exist within its senior staff members is a perfect example of lack of any sort of feeling for anyone other than themselves.

The real sad story here is that this is only the start of a vicious cycle! The economy

within these new countries will surely improve with the boost of the economies from these new jobs.

As the new (near poverty) workers get enough income to have purchasing power for more than the bare essentials, they will start demanding, (sound familiar?) a larger portion of the pie! There go the big new profit margins!

The corporations will then start looking for fresh masses to take advantage of. Guess where they will find them? Right back in the old locations where they were in the beginning! These economies will have been depleted from the relocation of their livelihood in the past, but this time it will most likely be the second generation workers of the original employees that supported the growth and development of these remaining companies originally. The first generation will have died off in the meantime.

I think the bottom line from my prospective is that the primary international Corporations are all short term gain oriented, with little or no long term growth and gain goals.

Now, I said at the beginning of this section that this was the US perspective, but it's not - it applies equally to all of the industrialised countries, all driven by greed, and the shareholders wanting more, more, more.

When are corporations going to realise that the shareholders are primarily gamblers ... they don't care about the company, its employees or its customers, only themselves. It is high time that management digs in its heels, and simply says "No!" to the ever increasing demands of the wealthy gamblers, who will buy and sell stock on the same day if it makes them a profit. Does anyone really think that this effects the way a company should do business? It is illogical in the extreme when a corporation sacks or retrenches a thousand workers, and the share price goes up. This is not business, it is just gambling, but gambling at the expense of the very people who made the corporation in the first place. A business without workers is a non-business, just like a worker with no job is a non worker.

There is supposed to be a mutual trust between the employer and employees, and in a few companies this still exists. For the vast majority, the worker is merely a pawn in a big game that no-one seems to really understand.

The smaller shareholders will eventually do well, if (and only if) they allow a well run company to do what it knows how to do best ... make its core business work, for the ultimate benefit of the employees, and following as a natural progression, the customers

and everyone else (such as the shareholders).

For the gamblers, they can feel free to visit their local bookmaker and gamble properly, rather than manipulate the odds in the stock market - let the people who actually create goods and products or supply services (i.e. The ones who actually make the economy work) get on with their jobs and lives.

As soon as any corporation stops looking to the future (not next week or next month, but a year - 5 years ahead), it immediately loses focus, and in the blind rush to try to keep up with the competition's share price, dishonest and illegal practices become common - until the whole thing falls in a pile - Enron, Worldcom, HIH Insurance are just three where the greed and corruption was exposed, and where the companies either folded or were disgraced. The loss of jobs, business confidence and the pain of the small investor are inestimable.

Why does a company try to insist on 30% growth per annum? 50%? To what end? Surely, the idea of business is to provide people with their needs - jobs, products, security. They don't care about these things any more, and look at the state of the world. Big business corrupts government if it suits it to do so, and we are able to see a pattern emerge, where the rich are very, very rich indeed, and the poor are exploited because they can't afford to fight back.

This is not an ethical way to run a business, a country or a planet - I think that it's high time that we all say that enough really is enough. Everyone has a right to a share of the pie, and making sure that the share is more evenly divided than at present is a good start.

No-one, absolutely no-one, is so smart, so good at what they do or so indispensable as to be actually worth 10 or 100 times (or even 1,000 times) what you or I would be paid for our job. Why should a worker, who knows his/her job and does it well, be paid $10/ hour, while someone else gets perhaps $100/ hr and another $1,000/ hr - in the same company? For some others, they wouldn't get out of bed for less than $10,000, let alone actually do any work. Are these people really that important, clever or hard-working?

No-one will deny that a CEO of a large corporation has a stressful and difficult job, just as no-one will deny that just about every job these days is stressful and difficult. There is not a person on this planet who is a thousand times smarter or better at his/her job than you or I, or who works a thousand times harder or a thousand times faster. There is obviously no reason that these people should be paid a thousand times more, nor 100 times more. Even at the other extreme, do you know anyone who is really (only) 10

times better, smarter or faster than you are?

No? I didn't think so.

Mark Hammer from Canada provided this additional insight, and has made some good observations that explain the malaise we see. His comments refer mainly to the actions of the stock market, and its effects on the small investor - these are the people who are often (financially) hurt badly by the "buy, buy, sell, sell!" antics of the gamblers, and the corporate attempts to satisfy what I believe is entirely the wrong sector of the market.

Economic "Rationalism" (by Mark Hammer)

The manoeuvrers that the corporate sector and their economist henchmen carry out are generally all in the name of good intentions. I work in government here, right in the same building as the ministry of finance, the treasury board and scads of economists, in the long shadow of parliament, and in the den of lawyers. They're all good people, but the majority have it wrong when it comes to the relationship between economy, economists and statesmanship. A thriving economy is certainly a worthy goal of a nation, but economic policy is not isomorphic with statesmanship.

So, these busy little beavers toil away on policies and negotiations designed to help 'the economy', often neglecting other things that matter. They believe they are doing the right thing, certainly cast no doubt in the minds of the corporate sector as to whether *they* are doing the right thing, and everyone sleeps soundly at night believing in what they believe in.

Western industrialised societies rest heavily these days on a number of things. One is certainly rampant consumerism and the mistaken equivalence between consumerism and 'progress' (''Hey, **my** phone has 10 brand new ways to annoy me, provoke rudeness on my part, require installation of new transmission lines, require teeny tiny fingertips, and generate embarrassing ring tones on the bus''). Another is certainly the disposable income of working adolescents (though this varies between North America and Europe). A third is the concomitant rise of retirement as a social institution and the stock market (and day trader) as the handmaiden and partner of retirement.

The stock market has always been there, and many people have tucked a little into the market for a few extra dollars. What has changed in the past 50 years is the concurrent increase in the duration/ length of the retirement period for many, the increasing duration

of what one might call the 'preparatory' years prior to establishing oneself (including college, university, and the usual adolescent tendency to try to escape the rat race entirely, until it becomes all too obvious that this is *not* going to work out in the long term). The net result is that a generally smaller portion of one's life is spent earning a decent wage.

The need to amass a personal net worth (current and future) which can provide for a retirement at the socially expected age leads many to either invest much more heavily and nervously than similar persons might have done in 1960 (when retiring at 65 or 67 was considered within the bounds of normal), or to be part of a work-related pension plan which is under pressure to deliver promised benefits. Indeed, retirement plans, pension funds, and personal nest eggs for the purposes of retirement make up a much larger chunk of the stock market than they might have many years ago. A lot of this money is related to retirement in opaque ways. For instance, when a bank issues guaranteed investment certificates, chances are that many of the owners of those GICs have purchased them expressly for retirement income purposes. The pressure on this money to perform is immense.

Correspondingly, the pressure on the organisations invested in to perform is similarly intense. I like to refer to this as 'nervous money'.

So what happens to 'nervous money'? Well, for one thing, to retain investors who are anxious about generating the kind of ROI (Return On Investment) that will keep them in the consumer style to which they have become accustomed - even after they don't work for a living any more - the profit margins have to be robust ... and the BS thick and savoury. My sense is that companies and boards of directors, elect to move, globalise, restructure, merge, buy out, down size, and do lots of economically and spiritually disruptive things in the name of retaining their investors.

Sometimes these moves are intended to increase profits, and sometimes they are simply intended to create the impression of the potential for profit (the stock market being mostly make-up, girdles, and falsies :-).

Whatever the case, if so many of these investors were not people who look at their money and see 25 years of their post-retirement life, it would be a different story. You will note that a good many of the folks who got stung by Enron were, in fact, stung for their retirement savings. I.e., they stuck with Enron and believed the BS because their need for retirement income urged them to. (This is not 'blame the victim' here, just trying to provide an aetiology of the current malaise).

The long and the short of it is that many of the things you mention are, to my mind, either direct consequences of nervous money or the spin-off to helpless victims (e.g., small non-public companies) who get caught in the wake of larger public organisations doing the kinds of things that nervous money does, like move out of country. Of course the countries where nervous money moves to are generally *not* replete with folks who are looking to invest their wages in retirement funds and pension plans because they actually don't *have* retirement as a social institution there - one would expect the financial planner business in many so called "third world" countries is rather slow - or certainly don't have early retirement as some kind of droit de seigneur.

I live in a town that has two major employers, civil service and telecommunications. Nortel, JDS Uniphase, and Alcatel are all situated here, as are Corel, Cognos and several other software places. Every day on the news, they list how local stocks are doing - something I don't remember hearing on the news as a youth. Do people actually know what the products of these organisations are?

Sometimes I'm not so sure. There are many days when you'd swear that their major business line was their stock. That companies can show huge leaps in stock value simply because the losses this quarter were not quite as bad as the same quarter last year (something which still baffles me) underscores the notion that folks care more about the stock than about what the company actually *does*.

Now I believe with all my heart that there ARE conscientious brokers who caution their clients to be patient, to accept the slow burn as opposed to being dazzled by risky high yield investments, to invest in companies that are more ethical, that give their commitment to local industry and don't simply chase money whichever border it happens to cross in its drive to evade taxes or fair wages. They're out there, but not everyone uses them.

What sorts of reasoning do the folks who invest multi-billion dollar teacher retirement funds apply? I have money in one of these, and I honestly couldn't tell you. I imagine the same goes for a lot of pension funds.

I suspect many folks who do public sector work 'for the good of humanity' would be quite surprised to know what their pension fund is doing and the kinds of corporate moves that may or may not have been provoked in an effort to hang onto that investment money. Has General Motors or John Deere closed plants locally and moved them south in order to show a higher profit margin or otherwise retain my TIAA-CREF contributions?

I certainly don't blame folks for wanting to retire, and I don't blame them for wanting to retire without being poor. That being said, however, if there are trickle-down effects of the kind of investment patterns that result from the way that social trends impact upon industry and the economy via the stock market and the nervousness of money, then I think people ought to take a moment to think through the choices they make, know more about the impact of their choices, and make some decisions. It's a *long* way from deciding to retire at 62, to the availability of decent potentiometers locally, to the living conditions of someone in a developing nation. But they ARE linked, even if by a great many intermediate stages, and one should consider one's values and the social repercussion of personal choices before making them. Yes, it's only one person, but one more person doing something is a lot different from one *less* person doing it.

Mark said that he often wonders is people know what the companies around them (or in which they have investments - knowingly or otherwise) actually make or do, and I am convinced that for the most part, the answer is "no". The same applies to the stock market, where investors (gamblers), know only that the company may have something to do with 'technology' - what they do with that information is based primarily on whether technology stocks are deemed 'good' or 'bad' on any given day.

The above is only one example, of course - similar 'decisions' are made on stocks of chemical companies, cigarette manufacturers, drug companies (making 'quit smoking' drugs :-), or anything else in the market. The problem is not with the small investor, the ethical investment broker, or the company itself. The greatest problems are created by the 'get rich quick' gamblers, who will artificially move a market based on rumour, whim or God only knows what, and cause a cascade effect that effects everyone else.

The company whose stocks are affected will then try to take 'remedial' action, which as often as not, is ill informed, and aimed at the wrong people. People lose their jobs, retirement funds, or in extreme cases (but sadly on the increase) their lives - the loss is more than they can handle, and they commit suicide.

The gamblers are of little consequence, and their actions are destructive to the company, its employees and its customers - should they lose their fortune, no one will care. That their actions can cause - directly or indirectly - the downfall of an entire corporation is deplorable, and it is high time that all companies (and governments) started looking to those who matter - those who are involved for the long term, and ignore the antics of the high flyers - let *them* crash, not the company or the economy.

Articles Index

Main Index

Copyright Notice. This material and all material contained in this web page, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott. Additional material is copyright (c) 2002, Fred Newton and Mark Hammer, and reproduction of their material is subject to the same conditions as shown above.

Copyright (c) 2002 Rod Elliott

The Audio Pages

Elliott Sound Products **Phase Angle Vs. Transistor Dissipation**

Copyright © 2005 - Rod Elliott (ESP)
Page Created 02 March 2005

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 The Loudspeaker Load
  - 2.1 Phase Angle Vs. Transistor Dissipation
  - 2.2 Test Design
- 3.0 Conclusion

1.0 - Introduction

This article will help those who have built an amplifier that just blew up for no apparent reason. It will also help those who are planning to build an amplifier, either from the ESP projects pages or elsewhere. Contained herein are answers to questions such as "why can't I use the P3A amplifier at +/-56V, or a single board P68 at +/-70V". These questions are common, and it is always good to know why something is so.

While this topic is also covered in the Safe Operating Area article, there is less detail in this present version, making it somewhat easier to read - especially for novices. In this case, I have concentrated on the specific effects of phase, and once this is understood, I suggest that you tackle the full version above.

While it is generally understood that a typical loudspeaker has an impedance (as opposed to resistance), the implications of this are not widely understood. Many of the designs

featured on The Audio Pages appear to be over-engineered, and the number of output devices seems excessive for the claimed output power.

This is entirely true! ... and for a very good reason ...

2.0 - The Loudspeaker Load

The problem with real loudspeakers is that they refuse to act like nice, well behaved resistors, and the impedance changes from being resistive, inductive and capacitive, depending on the frequency. Let's look at a typical speaker, whose equivalent circuit and impedance response are shown in Figures 1 and 2.

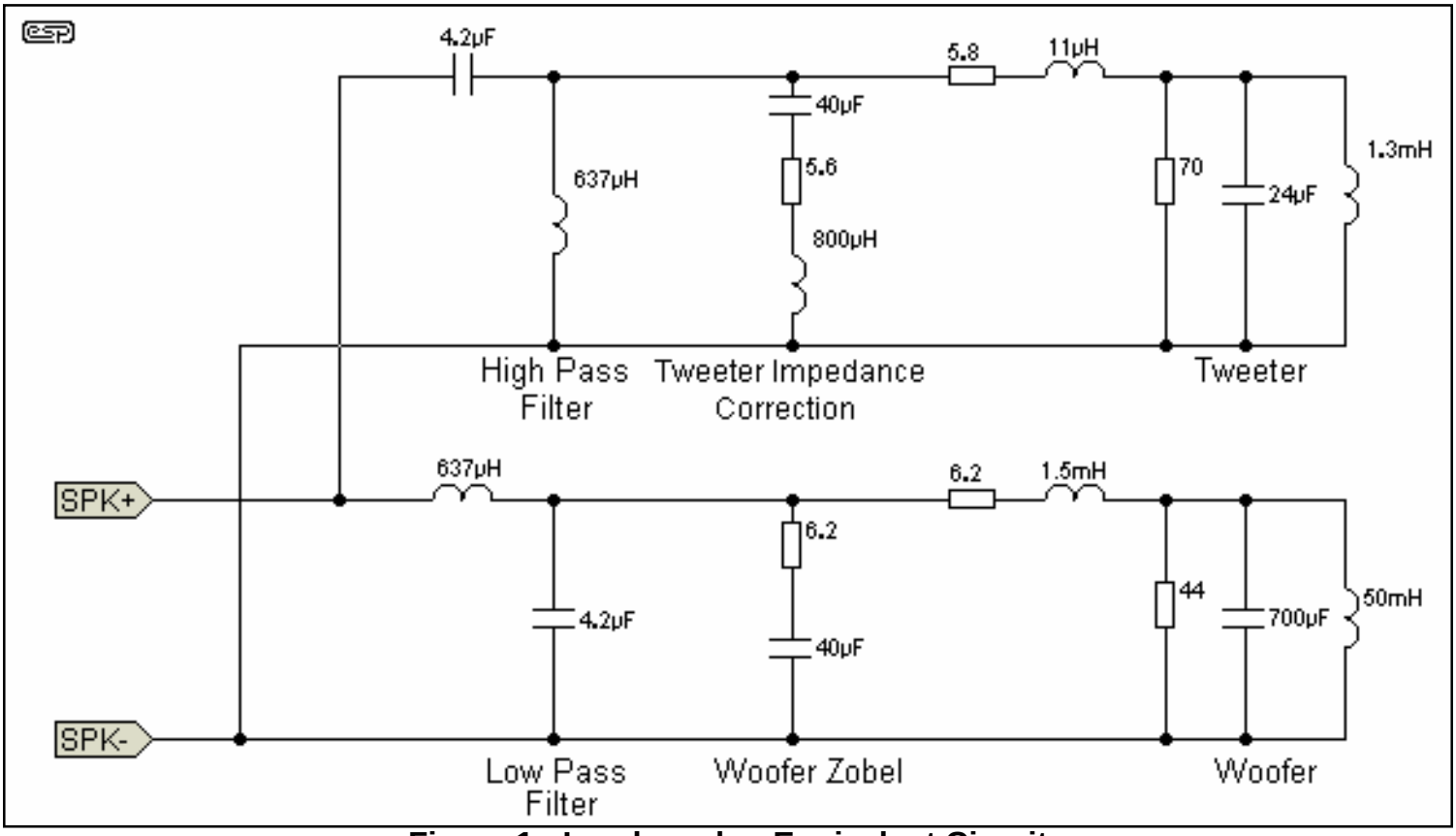


Figure 1 - Loudspeaker Equivalent Circuit

While it may look complex, it is simply a reasonable representation of a typical 2-way loudspeaker, having an impedance correction network to eliminate problems caused by the tweeter's resonance, and a basic Zobel across the woofer to damp the rising impedance cause by its voicecoil inductance. The crossover network is a conventional 12dB/octave parallel design.

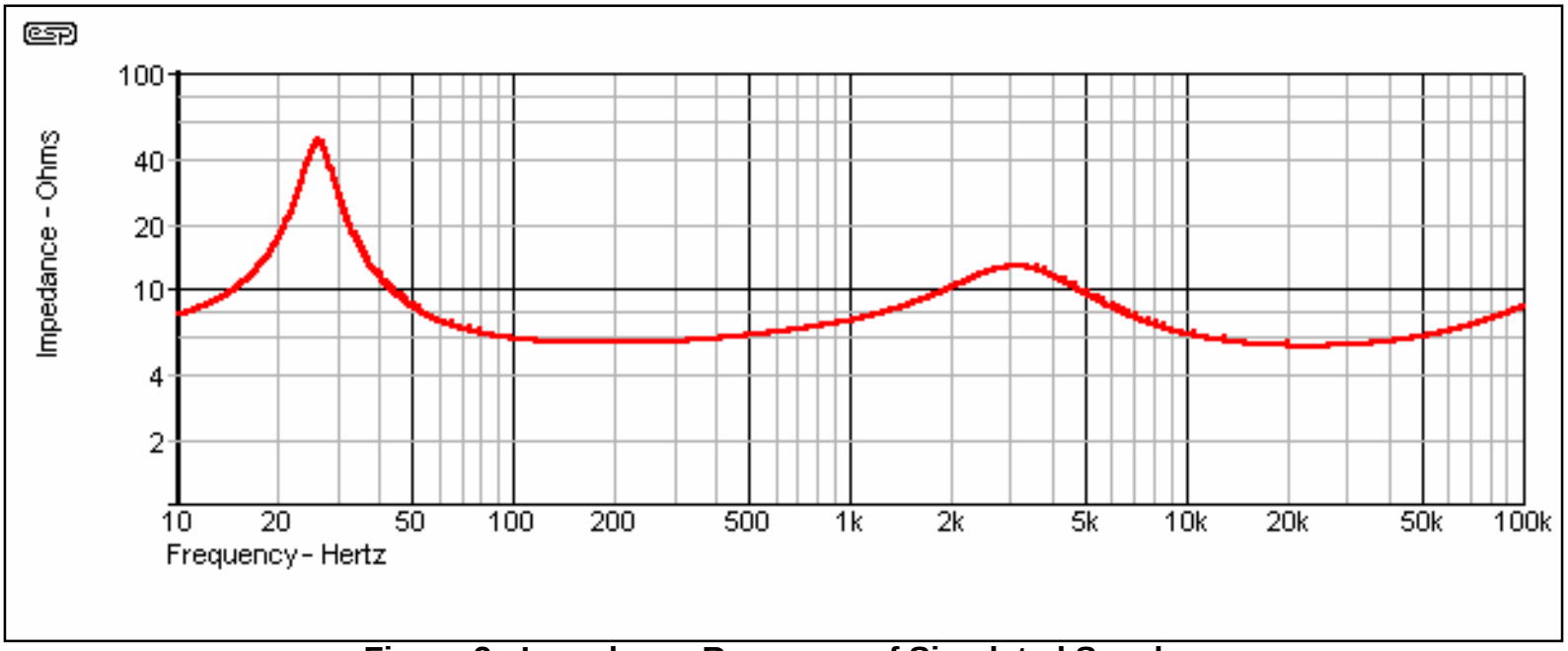


Figure 2 - Impedance Response of Simulated Speaker

Well below resonance, the system appears mildly inductive, with the inductive reactance component rising with frequency. At resonance, the load is purely resistive, and is at a relatively high value (typically from 20 to 50 ohms). Power at this frequency is very low - a 100W amplifier (8 ohms) will deliver less than 20W (19.6W to be exact) into an impedance of 40 ohms.

Above resonance, the load seen by the amplifier becomes capacitive, and impedance falls rapidly, eventually 'bottoming out' at some frequency (typically around 200Hz or so). This is the loudspeaker's 'nominal impedance' as quoted by the manufacturer. The impedance then starts to rise again as the voice coil inductance becomes significant. The crossover will almost always introduce further phase anomalies, and as you will see, any time the impedance changes, so too does the phase.

As long as the impedance is predominantly resistive, the amplifier has a relatively easy job, with voltage and current in phase, and amplifier dissipation is at the minimum possible for a given output power. The problems arise close to resonance - either above or below, where the load is highly reactive, or anywhere else where the impedance changes. Note that a vented enclosure has a second low frequency resonant peak, and the amplifier sees exactly the same reactive loading around the peak introduced by the vent tuning.

The worst case is that the amplifier sees either pure inductance or pure capacitance. Under those conditions, the voltage and current output from the amplifier are 180 degrees out of phase. Fortunately, this never happens with a loudspeaker.

2.1 - Phase Angle Vs. Transistor Dissipation

Although it may seem impossible (or at least highly unlikely), the voltage and current into anything other than a resistive load will not be in phase. Capacitive loads have a 'leading' phase, where the current waveform occurs first, followed by the voltage. Inductive loads have a 'lagging' phase shift, meaning that the current lags (is behind) the voltage.

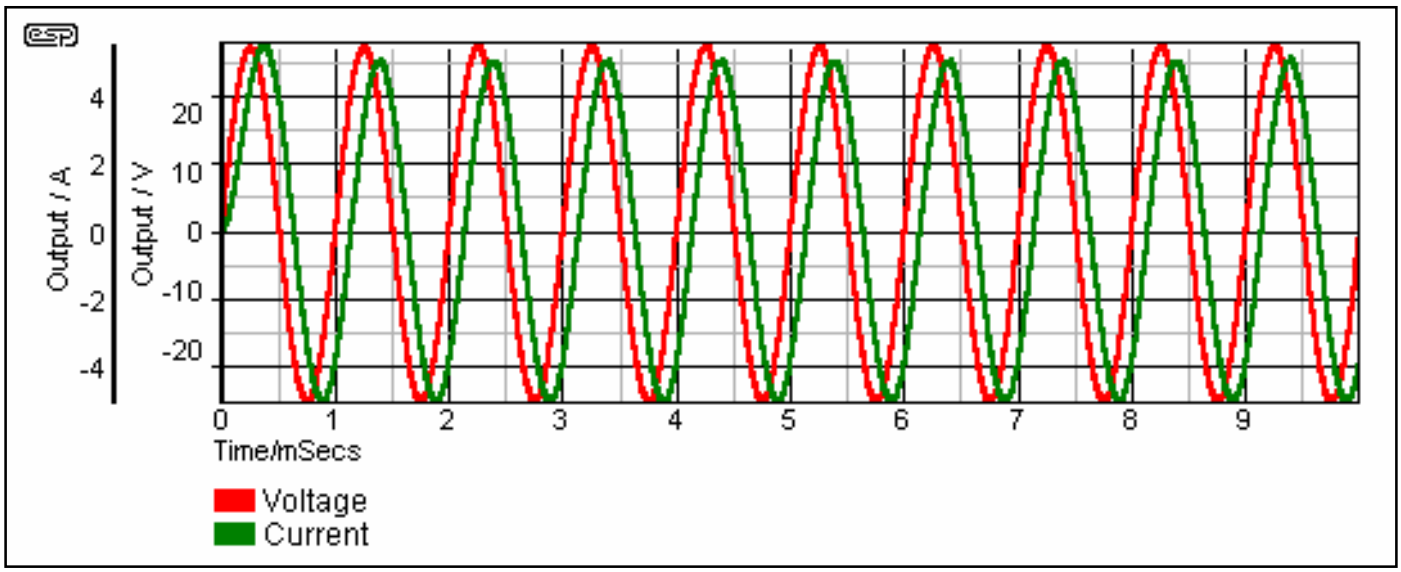


Figure 3 - Phase of Voltage Vs. Current With Reactive Load

As you can see, the voltage comes first, followed by the current (lagging power factor because of inductive load). In this case, the phase angle is a little over 45°, and was simulated by placing a 700uH inductor in series with the 4 Ohm load resistor. While not exactly the same as a loudspeaker load, it is good for simulations as it enables accurate determination of the peak transistor dissipation.

It must be understood that these conditions are steady state - the signal has to be present for a sustained period before the voltages and currents achieve their steady state values and relative phases. The actual time needed varies, based on the damping applied to the system, and many other factors, but in general, about 3 to 4 cycles will actually be enough. Normal music will easily be able to set up the necessary conditions for the voltage and current to be out of phase with any loudspeaker system. Let's have a look at the phase response of the speaker system shown in Fig 1.

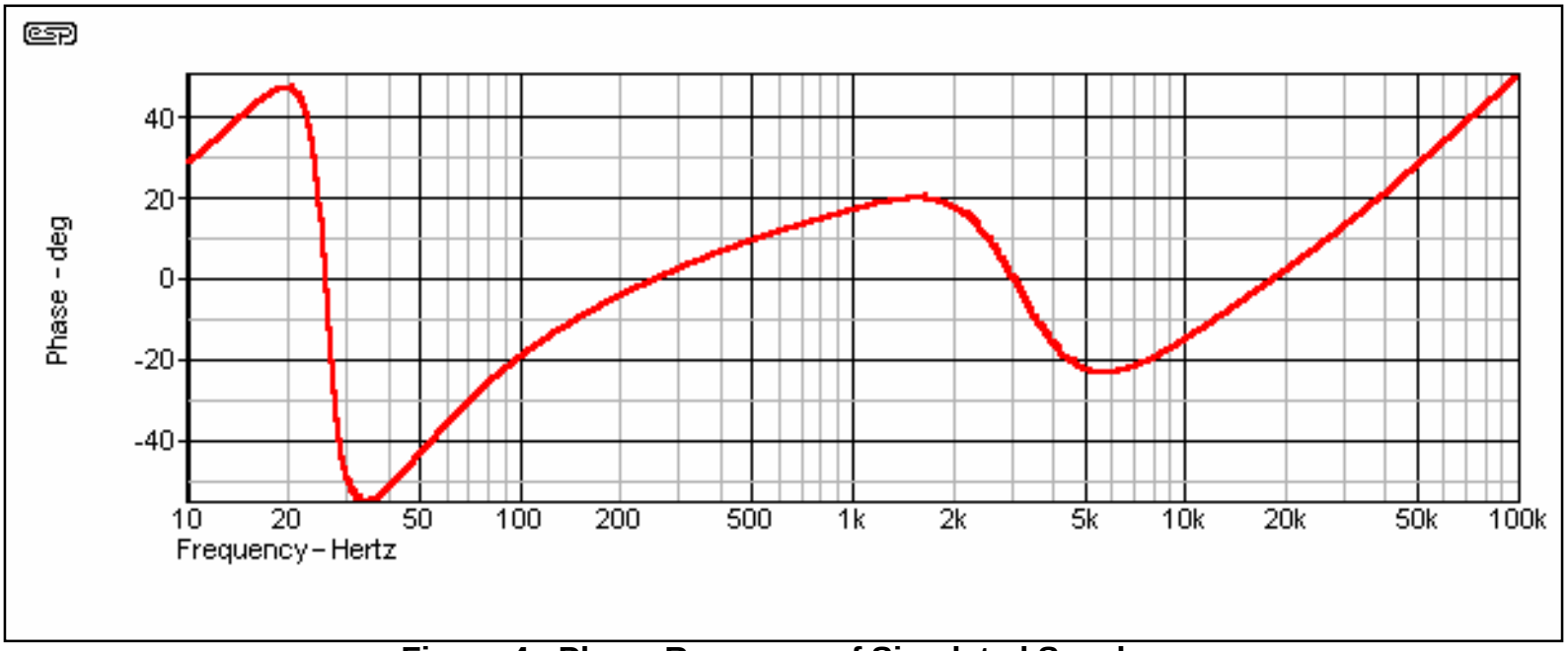


Figure 4 - Phase Response of Simulated Speaker

As you can see, the phase varies over approximately +/-50 degrees, and while this is fairly realistic, some speakers will exceed this. The majority (and especially mid-woofers as used in most 2-way systems) will have a phase response of +/-45 degrees or so, and some will be even less. It is worth noting just how little of the frequency range appears resistive - the phase angle over the majority of the frequency range is greater than 10° in one direction or the other.

Assume a 100W amplifier and a nominal 8 ohm load. Full power output is reached at a voltage of 40V peak (28.29V RMS). Maximum current is 40 / 8 = 5A peak or a little over 3.53A RMS. This is exactly the voltage and current at full power into a resistive load, and the peak transistor dissipation occurs at 1/2 the supply voltage.

At an instantaneous level 20V into 8 ohms, current is 2.5A, and transistor peak dissipation is 20 * 2.5 = 50W. If voltage and current are out of phase, the power delivered to the load is decreased, and the power dissipated by the transistor is increased.

Worst case (never achieved with any loudspeaker), is a 90 degree phase shift. This means that when the voltage across the transistors is at the minimum (turned fully on), the current is also at a minimum. That seems pretty good - zero dissipation can't be all bad. The problem is that the converse also applies, so when the voltage across the transistor is at the maximum, so is the current!

Assuming zero losses, that means that when the transistor has the full 40V supply across it, it is simultaneously supplying the peak current of 5A. Instantaneous dissipation is therefore 40 * 5 = 200W.

So, where it appeared that a 50W transistor was quite adequate, it is obvious that it will fail under these conditions. Add to this the fact that transistors have a SOA (Safe Operating Area) that limits the peak dissipation to the maximum rating or less (depending on voltage and current), and it is easily seen that more powerful transistors must be used.

Now, I stated earlier that the 'worst case' was never achieved in practice, and this is the only thing that saves us. In reality, the voice coil resistance is always in circuit, and this limits the maximum phase angle.

Looking back to Fig. 1, you can see that the voice coil resistance is not 8 ohms as you might expect, but 6.8 ohms. This resistance is the only factor that stands between your amplifier and a 100% reactive load, and the typical phase angle as a result is a little over 45 degrees. The speaker in Fig. 1 has a maximum phase angle of about 50° - not an unrealistic figure.

At 45 degrees, the transistor peak dissipation is doubled, compared to the case with a resistive load. This means that for our 100W 8 ohm amp, the transistor dissipation will be 100W instead of 50W - any increase of phase angle over 45 degrees increases the peak dissipated power vs. the power delivered to the load. It is fair to assume that the 'real life' worst case phase angle will be around 60 degrees, and will occur only near (above or below) resonance, or around the crossover frequency.

Table 1 shows the relationship between phase angle (the difference between voltage and current, measured in degrees), peak power dissipated in the amplifier and average power delivered to the load, normalised to 1W. The type of amplifier is unimportant - transistors, MOSFETs, valves or magic - all are affected equally.

Note that peak power is *not* the average power - these are entirely different things. Average power is used to determine the heatsink requirements, but peak power is the killer of bipolar junction transistors (BJTs) due to second breakdown. MOSFETs and valves will survive these momentary peaks without complaint - BJTs will not!

This table is for peak dissipation at the onset of clipping, and the situation is very different at worst case output level (¼ the maximum power, or ½ total voltage swing).

Phase Angle	**Power Factor**	**Power (Amp)**	**Power (Load)**
0	1.000	1	1
15	0.966	1.38	0.94
30	0.866	1.76	0.75
45	0.707	2	0.5

60	0.500	1.66	0.24
75	0.259	1.2	0.08
90	0.00	4	0

Table 1 - Power Factor, Normalised Amplifier/Load Dissipation Vs. Phase

Power Factor has been included, not so much because it is something you specifically need to know, but because some knowledge of it is likely to be useful. Power factor is determined by taking the cosine of the phase angle, and is essentially a 'figure of merit' for an AC load.

The 'magic' figure of 45° is worst case with typical systems, where the transistors must dissipate double the normal peak power, while the load only receives half the power it would normally get. Note that the amplifier peak dissipation *appears* to fall after 45 degrees - this is only because of the increased impedance presented because of the reactance. While these figures are reasonably accurate, it must be understood that the situation varies depending on output power and supply voltage - there are a great many variables, and it is not practical to try to cover them all.

The load power is interesting. Notice that at 90 degrees, there is zero power delivered to the load! There is voltage and current (referred to as VA with transformers), but no power, so no physical work is done - the speaker would be silent. This phenomenon is well known in the power industry, and is called power factor. The ideal case is a power factor of 1, where every volt and amp is converted into work - heat, light or rotation (for example).

Obviously, the worst case is a power factor of zero. Volts and amps are readily measured, yet no work is done, and the electricity meter remains still. The majority of real loads are somewhere in between, and the loudspeaker is no exception.

This is all the proof that is needed that a 90 degree phase shift never occurs with a loudspeaker load (i.e. the load is never pure inductance or capacitance). Although the sound output from a subwoofer at 20kHz is very low, this is due to many factors - none of which involves a power factor of zero. It is beyond the scope or intent of this article to discuss this further, so I won't.

The essential point here is that any power amplifier must have sufficient reserve power in the output (and to a lesser extent, the driver transistors as well) devices to handle the maximum possible dissipation. In real terms, this always means more output devices than you anticipated.

2.2 - Test Design

Let's assume that we want 100W into 4 ohms, based on the design of P3A. That means 20V RMS, or 28.2V peak. Assuming no losses, we shall use a power supply of +/-35V. Peak output transistor dissipation at 110W into a 4Ω resistive load is 70W, at an average of just under 23W (each transistor - 46W for the pair). This represents an easy load for the amp, and could easily lull one into a false sense of security.

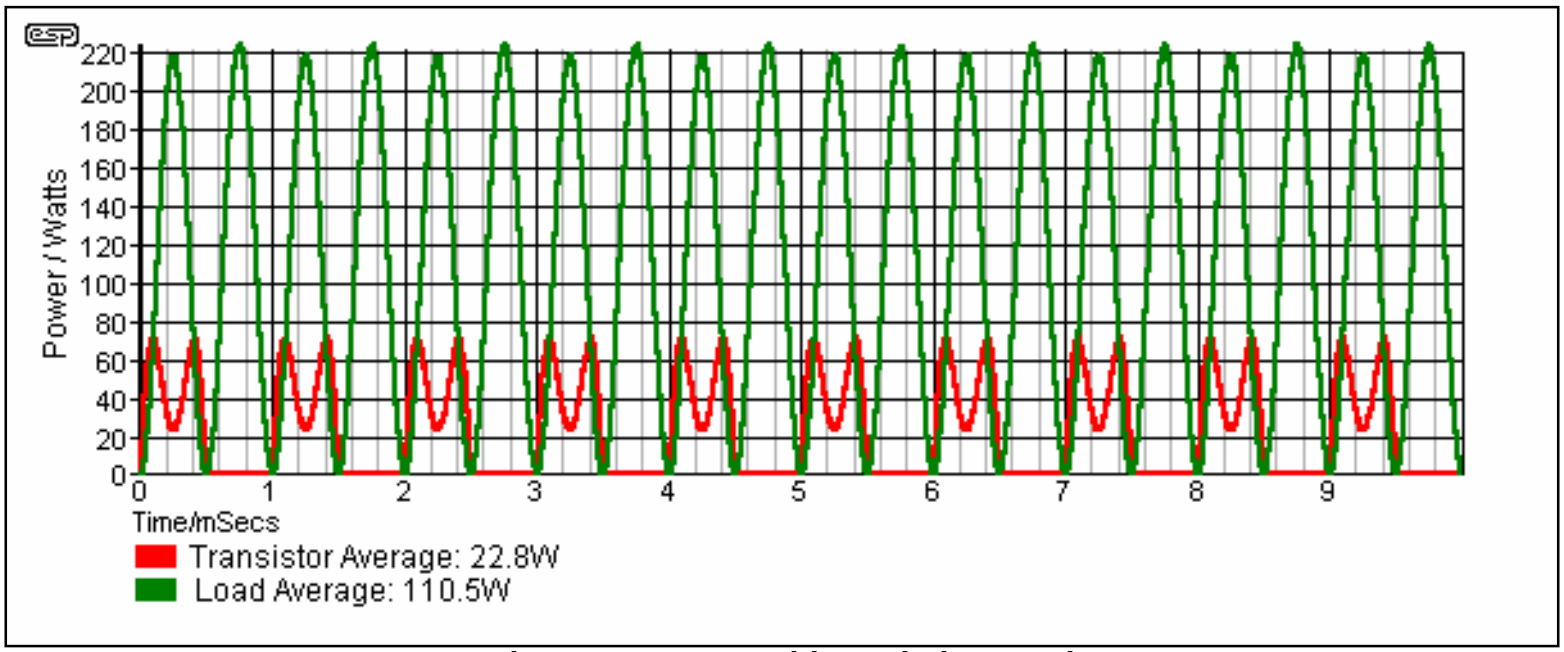


Figure 5 - Power With Resistive Load

Worst case power dissipation (resistive) is actually at around 0.707 of the maximum output voltage (½ power), or 14V RMS. Peak transistor dissipation is still 70W, but the average power increases to a little over 28W because of the way amplifiers work (this topic is covered in greater depth in the Amplifier Design article). This is not a large increase, but every Watt that needs to be disposed of means a greater load on the heatsink.

It is not unreasonable to design for a worst case phase angle of 45°, and as shown in the table above, power actually falls slightly above this. At 45° the peak dissipation is as shown in Figure 6 - this is where things can go pear-shaped in a hurry if you underestimate the operating conditions of a real-world amplifier. Transistor dissipation just before clipping is a bit over 120W (with an average of 28W). Any increase of the supply voltage will push the repetitive peak dissipation into the danger zone - especially when the power transistors are at an elevated temperature.

The above is based on a ±35V supply, but if you increase this to ±42V, at 4 Ohms reactive the peak dissipation will be over 200W ... the maximum rating for the transistors ... *at 25°C!*. Since it is highly unlikely (impossible, more like it) that the transistor die will be maintained at 25°C, the device must also be derated accordingly. Therefore, P3A cannot be used at 4 Ohms with a ±42V supply (something I have been claiming all along).

The bottom line is that to deliver 100W safely into 8 ohms, and allow 4 ohm operation, you need 400W of available transistor dissipation to ensure that the transistor SOA will not be exceeded at any time.

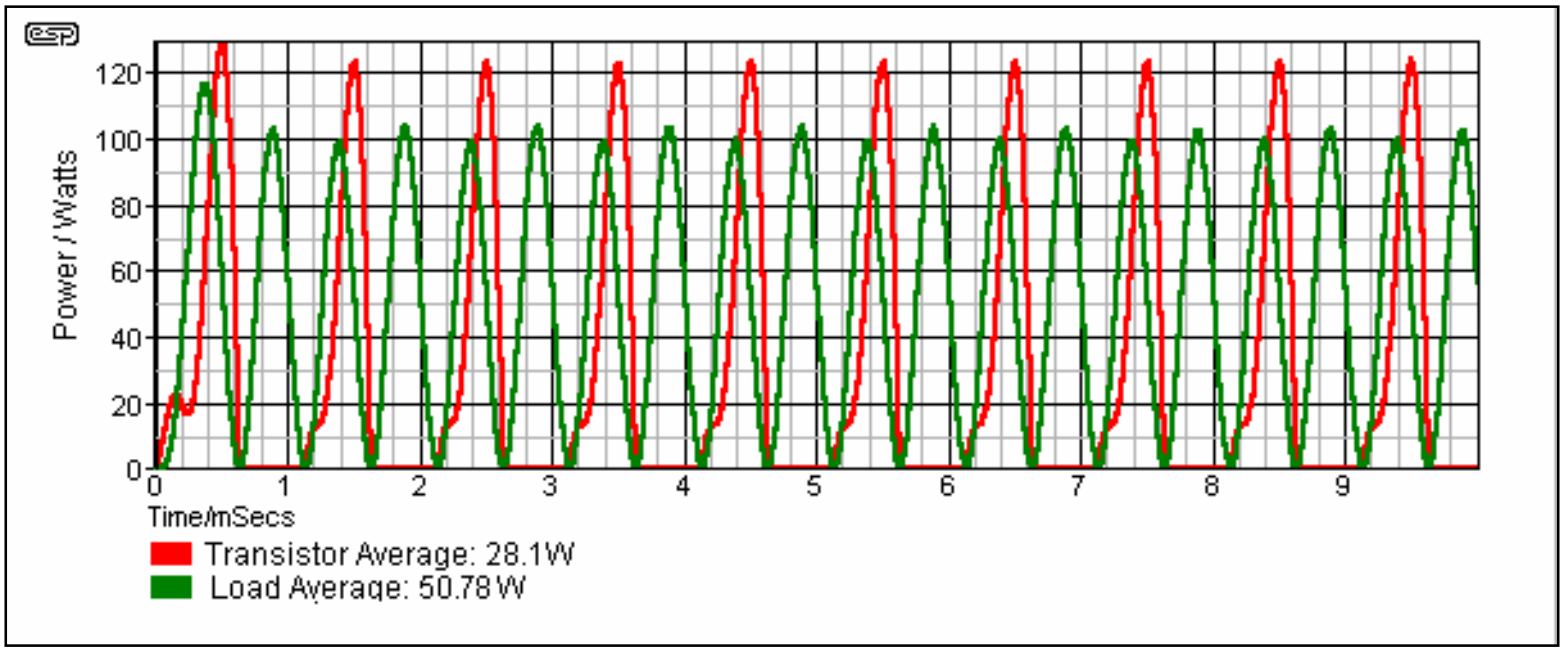


Figure 6 - Power With Reactive Load

Now you know why P3A (for example) is designed for operation at +/-35V, and the suggested upper limit (8 Ohms only) is +/-42V. When losses are taken into consideration, this is the absolute maximum recommended operating voltage. Under these conditions, using 200W output transistors, it is perfectly fine if the recommended supply is used, but is at risk if you go for the upper limit. The amp *might* be saved from destruction into 4 ohms (at the maximum voltage) by the fact that the supply voltage will collapse, and although this is common feature with many amplifiers, it is not recommended. Alternatively, you can always try your luck - it is amazing just what some people have managed to get away with, but this is not an approach that I am comfortable with.

It is quite obvious that the heatsink is of paramount importance, as is the transistor mounting. Maintaining the lowest possible thermal resistance keeps the transistors cool, and limits the amount of derating that must be applied. Other techniques that may be used include protection circuits, but these must also take the maximum operating temperature into account to be effective. It is widely believed that protection circuits contribute additional distortion and are 'audible', even when not activated. While this is possible with some designs, there is no doubt that aggressive protection most certainly *is* audible, as evidenced by many IC power amps - these must be kept below the protection threshold at all times.

3.0 - Conclusion

While this article has taken a somewhat simplistic approach to the issue, it is a reasonable description of reality. Real loudspeakers in enclosures will invariably make the amplifier's job harder than any resistive load, and even more so with a vented box. Few loudspeakers will present anything that looks even remotely like a resistive load, so amplifier dissipation will always be worse than simple analysis would indicate.

Of enclosures, the transmission line will usually present the easiest load to an amplifier, but unfortunately these are much larger than a conventional sealed or vented box. In all cases, a passive crossover network will also present additional phase shift. In some cases this can be extreme (usually due to poor design IMO).

At high frequencies, the amplitude is much lower than at low and mid frequencies, and even quite radical phase shifts do not cause undue amplifier stress.

The thing that saves some amplifiers is the power supply impedance, and careful design (hint - the cheapest alternative) ensures that there is enough power available for transients, but it will collapse sufficiently to allow for worst case conditions. This is not a good method to rely on.

Some commercial amplifiers use a tapped power transformer, and have settings for 8 and 4 ohms. The voltage is reduced for 4 ohm operation to make sure that the transistor SOA is not exceeded. Others take a more simplistic approach (many subwoofer amps fall into this category), where the transformer is simply too small for the job. If loaded heavily and driven hard, the supply voltage will collapse because of the under-rated transformer, and the amp will survive. Fortunately, music is dynamic, so the transformer will not have to suffer a sustained overload, and will usually live a long and happy life.

Remember, see the Safe Operating Area article for more detail on this topic.

All graphs shown were captured from SIMetrix circuit simulation software.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 02 Mar 2005

Elliott Sound Products **Sound Impairment Monitor - The Answer?**

Copyright (c) 2000 - Rod Elliott (ESP)

Under Construction - Hard Hat Area

Return to Articles Index

e-mail ESP

Back to Home Page

Introduction

There has been fierce debate for some time now about all sorts of problems that exist (or are supposed to exist) in amplifiers, speaker leads, interconnects, mains leads (power cords) and so on. Many of these are supposed to be so subtle that they cannot be measured by any known means, and the claims and counter claims can never be proven because of this.

The idea behind the Sound Impairment Monitor (SIM) is not new (I wish it were - I might be able to make my pages pay for themselves at last), but the application is unique. The thing that makes the SIM unique is that it has been specifically designed to be able to work with any amplifier that is reasonably sensibly behaved, and does not have massive phase shift at either frequency extreme within the audio spectrum. It is even possible to adapt the circuit to compensate for this, but the effort does not seem worthwhile (at least not for the time being). Having said that, some degree of compensation is essential to obtain satisfactory sensitivity to low level signals or small aberrations in the amplifier's linearity.

The important part of this is that if the sensitivity can be made high enough, then impairment may be seen at extremely low levels. I suggest that a well set up SIM should be able to show ...

- Distortion of waveform from any source. This could include alleged resonant (or microphonic) behaviour of amplifiers or other components
- The effect (or lack of effect) of so-called tuning, where an amplifier is claimed to contribute artefacts to the sound because of mechanical vibration
- The effect (or again, lack of effect) of different mains leads (power cords), interconnects or other

components within the SIM loop

The above is only a small sample, and I expect that many more will come to light if the principle is used in earnest. If (for example) a mains lead is claimed to improve bass response, then this must show up as a change in the amplifier's response. We do not generally percieve a dramatic (or even subtle) change in overall tonal balance without a corresponding electrical change (all other things being equal, of course). While this may possibly be elusive in static tests, it must be visible in a dynamic test under normal listening conditions, if the original signal can be subtracted from the amplified version to leave the smallest possible residual. The smallest change in performance will therefore be visible, since the input signal has not changed, but the output signal must. This will show up for phase or amplitude changes alike, since the subtraction process will be different from that when it was set up originally.

Let's face it - if the signal is exactly the same before and after a tweak, then the tweak in fact did nothing. If the signal is exactly (note: *exactly!*) the same, then we cannot possibly be hearing a difference. We might *think* there is a difference, but this in no way means that there *is* a difference - especially if the audio signal is unchanged. This last point is crucial to our understanding of the "mysterious" and "unexplained" differences between components.

In the same way, if any tweak causes the output of the SIM to change, then it did something. We don't know by looking at the signal if the "something" was good or bad - that's what we have ears for. What we do know is that what we are hearing is real, since we now have some physical evidence that we can analyse, disect and discuss until the reasons can be determined.

No amount of critical listening can reveal something that simply does not exist. Likewise, no amount of "superb specifications" will make an amplifier actually sound good in all cases. The goal is to isolate what small differences do exist, and to corelate these with what we hear. This is the goal of the SIM project, and with the use of a tool to allow us to see the effect of any change in amplifier topology, cables or damping that we may care to try, some plausible explanations will come to see the light of day.

In the meantime, we are all subject to our own prejudices and beliefs, "what we think we know *makes* a difference", and "what we think we know does *not* make a difference". For example, I believe that mains leads make no difference, while others believe the opposite. Now we can put it to the test, with an instrument that works in real time, with real music. Will differences in mains leads be revealed at last? I don't know, since I have not had the time I would like to experiment and perfect the external SIM.

System Description

The Sound Impairment Monitor concept is quite simple, and just subtracts the input signal from the output signal to indicate the difference between the two. An ideal amplifier will show a difference in amplitude only, and this is nulled using the Calibrate control. The resulting signal is the difference between the input and output of the amp, and will include any phase shift, frequency variation, or distortion. It is quite possible that the SIM could be implemented in the digital domain, using a DSP (Digital Signal Processor) to make the subtraction process more accurate.

The version I have developed is analogue, and requires the best opamp possible to get good results. The connection of the SIM is shown in Figure 1, and it includes a LED to indicate that something interesting (i.e. bad) is happening. An oscilloscope can be connected to allow us to see exactly what is causing the problem, and the SIM can be used with normal programme material. Indeed, this is the whole idea, since it avoids all the criticism of static (sine wave) tests. The SIM will be just as revealing (and in some cases more so) with so-called "static"

sinewave testing, since the exact distortion waveform will be quite visible. Square wave testing will also be very revealing. All amplifiers will show a difference between a fast rise-time square wave at the input versus output, but a modified (i.e. filtered to some agreed standard) square wave is still a very hard test on an amp, and may be helpful in identifying the problems that are supposedly audible, and isolating those that are not.

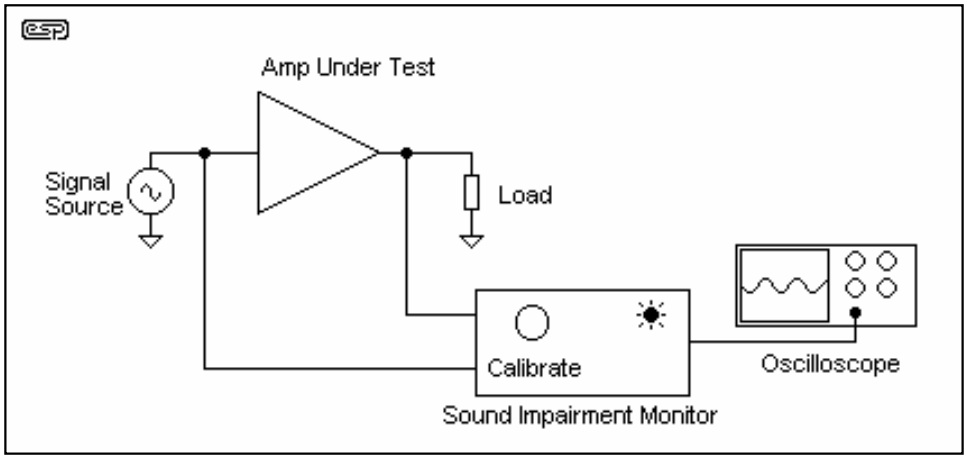


Figure 1 - Connecting The SIM To An Amplifier

When the SIM is connected, a normal signal (at a normal to low level) is applied. The calibration controls are set to ensure that the LED is not lit and the monitor amp should be almost completely quiet. Sensitivity (not shown) is increased, and calibration is readjusted until the highest possible sensitivity is obtained with the LED not illuminated. Some experimentation will be needed to ensure that small phase differences do not cause the SIM to indicate a problem that does not exist, but initial testing indicates that this may be an extraordinarily useful test for amplifiers.

With the monitor amplifier, we will be able to hear any difference between the original and amplified signal, and for the very few amps with distortion that is low enough, this should be mainly noise. There will always be some signal present, but it should not sound at all like music since we are interested only in distortion products. Note that distortion covers everything in this context - frequency and phase response, amplitude variations, and actual distortion.

Any distortion (regardless of origin) that arises as the level is increased or changes are made to leads or other components will cause the LED to light, and we can observe the difference between input and output on the oscilloscope. Figure 2 shows the waveform we will obtain if the amplifier clips and has some degree of overhang ("power rail sticking") - a sinewave is shown for clarity, but the same principle applies with music.

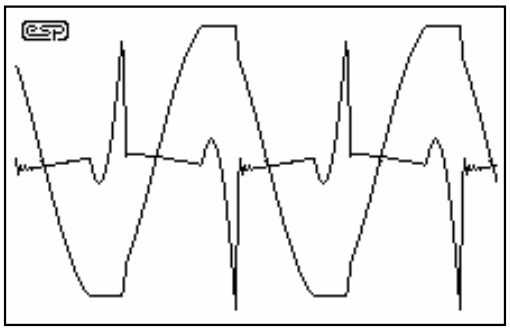


Figure 2 - An Amplifier With "Overhang" Clipping

Is this the perfect amplifier monitor? No. In some cases normal variations in input level will cause the LED to light, but the oscilloscope should show exactly what is happening. Any variation in the shape or sound (rather than the amplitude or volume) of the waveform indicates a problem, but it may not be possible with normal

programme material to see the exact mechanism that is the root cause.

It is potentially a much better test than the standard tests we have seen a thousand times (all looking pretty much the same as each other), but are still unable to reveal everything. Perhaps someone out there with DSP programming skills will be able to do a better job than a simple analogue circuit can achieve, but even as it stands, it is better than the current methods. By how much? This remains to be seen ....

Or heard. The output of the SIM really should be monitored with headphones or a loudspeaker, so not only do we see the difference on an oscilloscope, but we can listen to the difference signal. In a perfect amplifier we will hear silence - there will be no difference at all, but in the real world we will *always* be able to hear and see a residual signal.

Here is where listening is important! The residual signal may have a hard "gritty" edge with an amplifier and one set of speaker leads (for example). By substituting the leads, this may become a "smooth" signal, which although bearing no resemblance to the actual music, has no harsh overtones or other nastiness. We can safely assume that the difference is beneficial, that the second set of leads really does improve the sound.

We still don't know why, but armed with some evidence we have at least a fighting chance of finding out.

Initial Tests

I ran some tests on the SIM concept, with a normal signal and my 60W amp. The reaction to signal clipping was instant and highly visible. Even clipping that was of such magnitude and duration that I could not see it on the oscilloscope waveform of the amp's output showed up instantly, as will any other variation. Likewise, any signal that exceeds the amplifier's slew rate also shows up, but I had to deliberately slow the amp down by increasing the miller capacitance to see this.

If this amp had any form of protection, activation of the protection circuit would also show up as large spikes on the output waveform, so I am well satisfied that this is an extremely valid test. Unlike a clipping indicator, the SIM shows any deviation from the normal expected waveform, howsoever caused.

I found that phase shift at low frequencies (caused by the amp's input capacitor) made a good null very difficult, and after some experimentation I found that a high pass filter at the detector output tuned to about 700Hz eliminated the low frequency signal with no apparent loss of accuracy for signal impairment detection.

Next Steps - Involvement, Anyone?

I experimented with a square wave signal, and quite predictably the SIM went crazy, indicating every level change. The rise and fall time of my oscillator is very fast (it will give an excellent square wave at 100kHz), so this was entirely expected. The same will happen with almost any amp in the known universe, since few (if any) can reproduce a square wave perfectly, especially when the input signal is as fast as mine.

To be able to get an instant feel for an amp's ability to reproduce normal music, I suggest that for square wave testing with the SIM, a low pass filter be used, having a 6dB per octave rolloff from 1kHz. This will limit the harmonics to something passably sensible - it is still a severe test, and one that few amps will ace, because the SIM is so sensitive to any impairment to the signal - of amplitude or phase or any combination.

I am anxious to hear suggestions from anyone who might be interested in assisting with the development of a standard test method for amplifiers - one that will show the things we need to see, and reject those that are inaudible. The SIM is potentially the first real advance in amplifier testing methods for many years, and if properly

implemented could become a standard test that can be included in all reviews.

For example, if standardised, we could have an oscilloscope trace of the SIM filtered square wave performance for each amp tested and reviewed, and if left connected to the review amp (without the filtered square wave, of course), will enable the reviewer to ensure that no protection circuit is activated and there is no clipping during the listening test. I suspect that a few people might be surprised at the results, as I was when I was sure that the amp was not clipping with an applied music signal (according to the oscilloscope), yet the SIM showed quite plainly that it was.

One of the best things about the SIM is that it is relatively cheap to build, requires minimal calibration other than compensating for the gain (and in some cases phase) of the amp under test, and shows up imperfections that occur in music and in "real time" - not just steady tones. As mentioned above, it will also show that protection circuits have activated, and can display the instantaneous magnitude of the error signal on a LED bar graph, peak programme meter or monitor speaker.

If it appears that I am a little excited by my findings so far, this is only because I am! Can anyone imagine the jubilation if the SIM were to show that the perceived differences in various amplifiers showed up clearly, and in a way that could be easily analysed.

Not only amplifiers themselves, but speaker leads, interconnects, and even mains leads will all show a difference if it exists - we might be able to eliminate some of the fierce debate that currently exists if it can be shown that the effects are real (or equally that they are not). All of this with real music, connected to real speakers and in real time.

There is not a test methodology currently used that has this capability, so you bet I'm excited. I am even prepared to give this technology away to anyone who wants to try it. A project for the internal SIM is already published, and the external SIM will be published when completed.

External SIM

The idea of the external SIM is actually not new, having first been proposed by Peter Baxandall many years ago. The current design bears some similarity to the original, but there are important differences. The original Baxandall circuit is shown in Figure 3, but has a small problem in that the amplifier under test is expected to be inverting. This is very rare in modern designs, but in its time was relatively easy, since the amplifiers of the day were valve, and the polarity of the output transformer winding could simply be reversed. This is no longer possible.

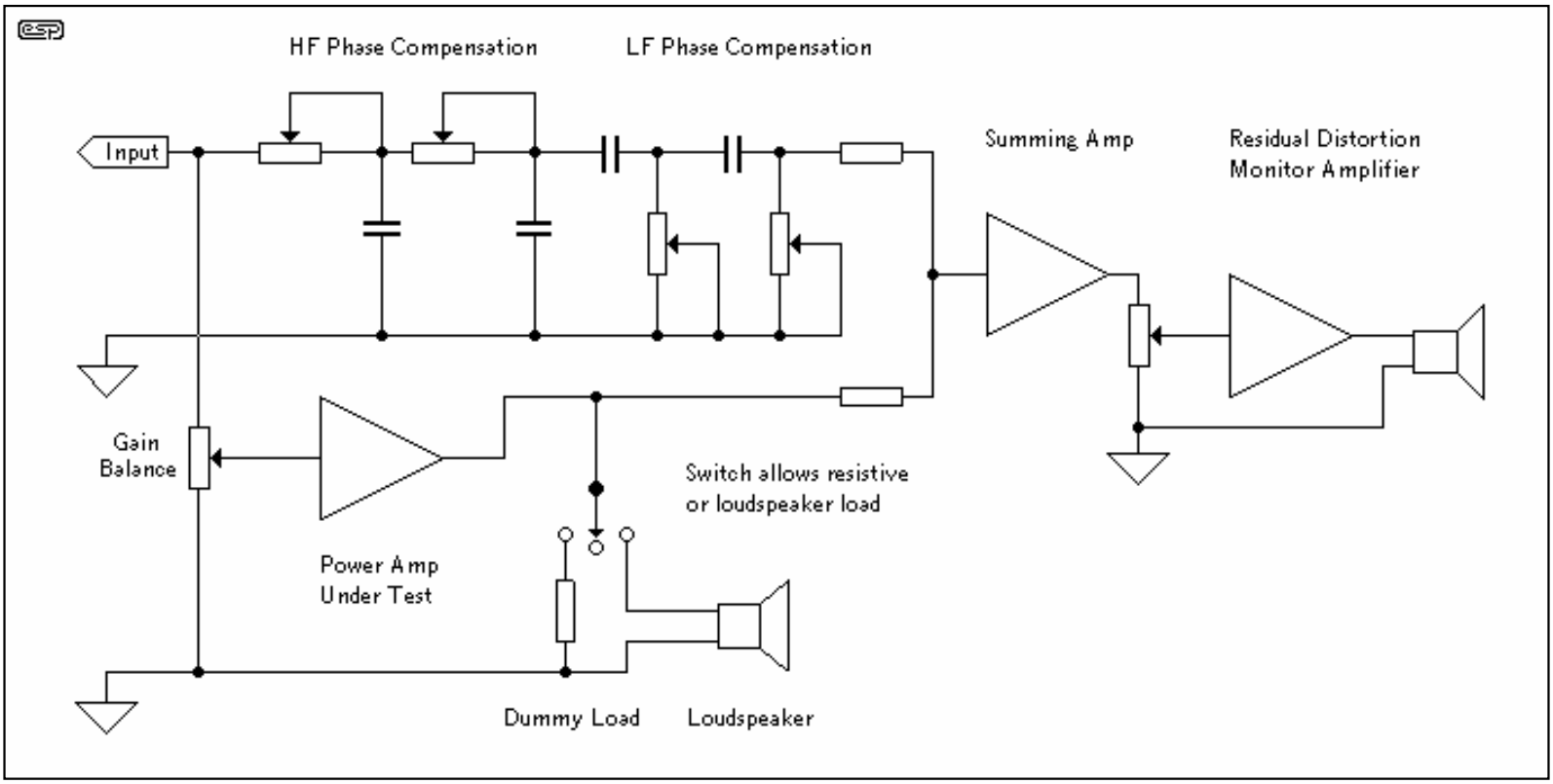


Figure 3 - The Original Baxandall Subtractive Test Set

Naturally enough, the term SIM was not applied to the original circuit, but name aside, I am puzzled as to why it never gained the popularity of "static" sine wave distortion testing (which is actually anything but static, a sine wave is a signal like any other, and stresses the amplifier in the same way as a music signal).

The original circuit proposed that gain and phase errors at high and low frequencies could be nulled by the compensation pots, and that the amplifier's gain would be equalised by adjusting the gain balance. Any residual signal would then be a direct representation of the distortion components in the amplifier. This residual may be monitored, and only the distortion components will be audible.

The Final New Design Criteria

Modern amplifiers will generally need minimal phase compensation, at least within the audio range. Because of this, the circuit can be simplified to some degree by reducing the phase nulling circuits to a single pole each. The external SIM is based on this principle, and with careful adjustment will allow very small differences to be heard (and seen on an oscilloscope).

To reduce the loading on the input source, the final SIM will use a high grade opamp to buffer and invert the signal before the phase compensation circuits. This has the added advantage that it will also allow a better match to the amplifier because of the very small phase differences it will introduce itself. To verify that the input buffer is not a contributor to the residual signal, I will be setting up a calibrate switch, allowing the amp under test to be bypassed entirely and replaced by a direct connection.

To minimise any loading on the input, a unity gain buffer will also be incorporated. This may change the sound to a small degree in itself, but will not affect the validity of any comparisons done on the test amplifier, as it will be in circuit all the time. To completely eliminate mains interference, the SIM will be battery powered, but will drive an external small power amp for monitoring. This connection will be balanced to eliminate earth loops and hum injection back into the SIM and the amp under test.

The 700Hz filter I mentioned above is an option I will have to think about very carefully. I do think it is a good

idea, but it should be able to be bypassed (or perhaps made variable).

As you can imagine, there is a lot of work to be done before the design is complete, and the basic criteria above are likely to change as I progress.

Internal SIM

The internal SIM is published as Project 57, and can be applied to any amplifier. It does require that the amp be modified, and it is very important that leads are kept short and all wiring is carefully done so as not to adversely affect the amplifier. There is little point in an instrument that will show you how badly the amplifier performs when the instrument is connected!

The internal SIM is only capable of showing differences that occur within the amplifier itself, although in some cases the speaker leads may affect the calibration. Again, it is possible to use a small headphone amp to allow the residual signal to be monitored. Now when you change the stock mains lead for a set of nice new SupaTune9000s, you will hear the difference in the amplifier itself, if indeed any difference exists.

Is the difference good or bad? This is for you to determine, but at least you will *know* there *is* a difference. This is the first principle of experimentation. There must be an observable change in an amplifiers output for any difference to be heard. This much is plain. What is not plain (using conventional test methods) is how this can be detcted. The SIM might well be the missing link, revealing all (well, some, at least) to those who really want to know.

Return to Articles Index

e-mail ESP

Back to Home Page

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 27 Feb 2000./ Updated 23 May

The Audio Pages

Elliott Sound Products **Semiconductor Safe Operating Area**

Copyright © 2003 - Rod Elliott (ESP)
(With thanks to ON Semiconductor for additional material)
Page Created 08 August 2003

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Amplifier Power Demands
  - 2.1 Amplifier Design Parameters
    - 2.1.1 Output Power
    - 2.1.2 Current
    - 2.1.3 Dissipation
  - 2.2 Loudspeaker Loads
- 3.0 Transistor Limitations
  - 3.1 Second Breakdown
  - 3.2 Device Parameters
- 4.0 Destruction!
- 5.0 Protection Schemes
- 6.0 Conclusion
- 7.0 References

1.0 Introduction

Safe Operating Area (SOA) for semiconductors is a little understood topic. Although the chart is generally provided in the data sheet, there is a great deal you need to know to be able to make proper use of it. Without a thorough understanding of the loudspeaker load, instantaneous voltage and currents, and what happens to transistors if the SOA is exceeded, it is easy to imagine that the supply voltage for an amplifier can be increased up to the maximum voltage allowed by the transistors used.

This is not the case at all, and this article discusses the problems faced in any amplifier design to create a reliable circuit that (ideally) can never place the power devices at risk. This is much easier said than

done, unfortunately.

2.0 Amplifier Power Demands

Most basic analysis of a power amplifier design is done (at least initially) using circuit simulation and basic theory. None of this is at all difficult, and is essentially a matter of current analysis through the amplifier into the load. For the sake of simplicity, a resistive load is generally used for all but the most rigorous analysis, and for low powered amplifiers, this is quite sufficient.

When you make (for example) a 20 Watt amplifier using discrete components, most of the power transistors available have so much reserve current and power available that few problems will be encountered. Even a 100W amp is not a problem if the impedance is known in advance, and reasonable care will give a reliable circuit.

The problem is that real life loads are neither predictable nor reasonable, with nominal* 8 ohm loads perhaps plunging to 3 ohms or less at some frequencies, and soaring to 50 ohms or more at loudspeaker driver resonance. Four ohm loads are no better, and 2 ohm loads are a nightmare for most amps.

One of the problems is that music is also unpredictable. Some music has a very low "crest factor" (the ratio, in dB, between the average and peak power), so relatively high power levels are present on a more or less constant basis. Other music has a high crest factor, with a peak to average ratio of up to 20dB (a power ratio of 100:1). Classical recordings are commonly thought to have a high crest factor, but this is not always the case, with some having as little as 6dB. Yes, this is uncommon, but it can (and does) happen).

"Modern" music (a term that has a different meaning to everyone ;-) is not immune from high crest factors, but they are less common than in unprocessed orchestral recordings (for example).

It is the combination of unpredictable loads, very different musical styles and power demands, different listening preferences and (we must never forget this one!) ... heat, that can spell doom for even the best designed amplifier, if it is used outside of the original design parameters.

* Nominal - in name only. As used in electronics, "nominal" refers to the expected voltage, power, impedance, etc., under "normal" (or sometimes ideal) conditions.

2.1 Amplifier Design Parameters

So, what are the factors that determine if an amp will be reliable or an owner's nightmare? There are quite a few, and as an example I will use the P3A 60/100W amp design from The Audio Pages projects section.

2.1.1 Output Power

The primary objective is to produce a design that has sufficient power to suit the needs of the owner - this

is a very difficult, because of the vast differences in loudspeaker efficiency, preferred listening level, and type of music. Nonetheless, 70W is not an unreasonable figure, and is well suited for many speaker systems. Smaller 2-way units are very popular because of their relatively high spousal acceptance factor, and they are convenient, reasonably priced and can give very good performance.

Such systems simply will not take the full continuous power of a 250W amplifier. The sales blurb may claim they are suited to amps of "20-200W", but this often assumes a "typical" crest factor of around 10dB (10:1), where the peak power will be 10 times the average, or in some cases even more.

This has everything to do with the amplifier design, as it sets a reasonable expectation of the power needed, and sets us towards an understanding of the load the amp is expected to drive. A quick analysis of any 2-way or 3-way speaker will show that the impedance is far from flat, it has peaks and dips at various frequencies, and will only show the nominal impedance at a few frequencies.

Again, for the purpose of explanation, I must choose a speaker system, and the one used for the remainder of this article is completely imaginary. It exists in simulation only, but has a reasonably close resemblance to many small/medium sized 2-way loudspeakers. The reason for a simulated speaker is simple ... the effects of the impedance variations are easily seen, and are actually very similar to a "real" speaker.

The primary requirement for obtaining power is voltage swing. This in turn is determined by the supply voltage, and the supply voltage and lowest impedance determines the maximum current.

2.1.2 Current

Using the (nominal) ±35V supply for P3A as the example, we must accept that even for an 8 ohm loudspeaker, the minimum impedance will be lower than claimed. Six ohms is a realistic figure (assuming a well engineered speaker), but it could be less. A 4 ohm speaker can be expected to have a minimum impedance of around 3 ohms.

Power transistors are assumed to be rated at 200W dissipation, and have adequate voltage and current capabilities (e.g. the recommended devices for the P3A amplifier).

For the 6 ohm case, peak current will be 5.8A, or 11.6A into 3 ohms. This is with a +/-35V supply, but the transformer voltage is always quoted at full (resistive) load, so typically with normal mains voltage, the supplies can be expected to be ±38V or so with no load. Large filter caps will hold this voltage for several milliseconds, so the maximum peak currents are probably closer to 6A and 12A for 8 and 4 ohm nominal loads (respectively).

2.1.3 Dissipation

Under ideal conditions, a transistor's power dissipation rating refers to the maximum average power that the device can handle, with the junction temperature at 25° C. At any temperature above 25°, the power is derated (reduced) linearly, until it reaches zero at around 150° C.

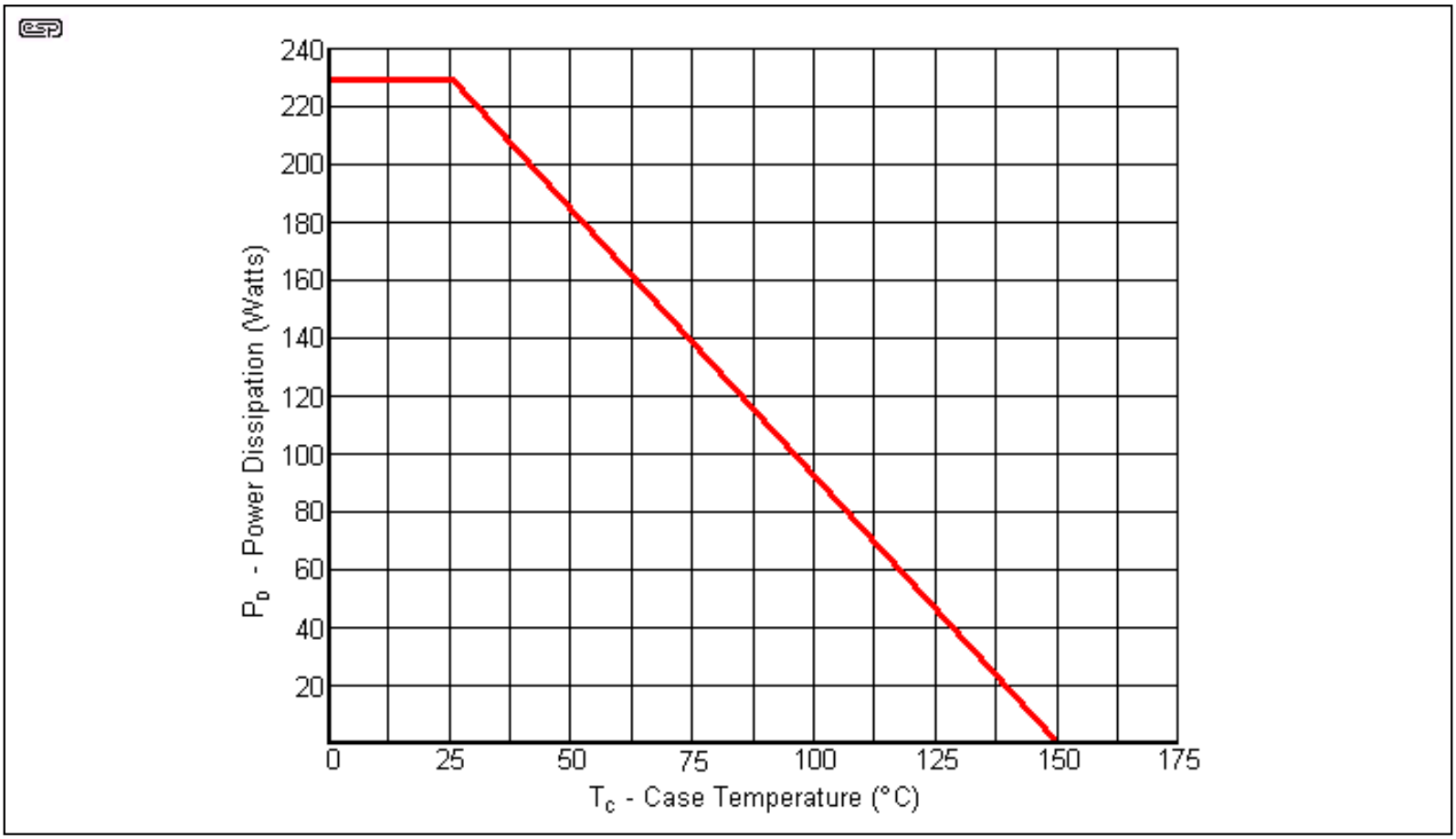


Figure 2.1 - Typical Power Derating Curve

Based on this, it is obvious that keeping the temperature down is critical, since elevated temperatures reduce the available dissipation, and reduce any safety margin that has been incorporated into the design. High temperatures also reduce device life, and the hotter a transistor runs, the expected life is reduced roughly exponentially.

To place the thermal issues into perspective, there is a calculation in the next section that shows average transistor dissipation to be about 40W (42.5W actually), when driving a 4 ohm reactive load at full power using ±35V supplies. For the sake of simplicity, we could assume about half that to be the continuous average with music at the highest level before clipping. 20W continuous does not sound like very much, but the thermal resistance from junction to air (per transistor) might be around 5.5° C/W ...

Rth(j-c) = 0.54°C/W (junction to case, assume 0.5°/W for simplicity)
Rth(c-h) = 1°C/W (case to heatsink - a very good figure, and difficult to achieve in practice)
Rth(h-a) = 4°C/W (heatsink to ambient, based on a 1°C/W heatsink and 4 transistors [two amps])

... so the temperature rise is 110°. Now, add the ambient temperature - say 25°C - but it could be a lot higher!. The die temperature is therefore 135°C, and from the above chart, that means that continuous allowable power dissipation is reduced to about 23W, so we are way too close to the thermal limit of the transistors - even a minor obstruction over the heatsink would be more than enough to tip the balance.

There is a lot more than just simple dissipation though. Figure 2.2 shows the power in a transistor driving a resistive load at the onset of clipping. As you can see, the power increases until the voltage reaches the halfway point between zero volts and full supply. After that, it goes down again - in a perfect amplifier,

dissipation will fall to almost zero at the clipping point. For Figure 2.2, the applied voltage was ±35V, with a 3 ohm resistive load.

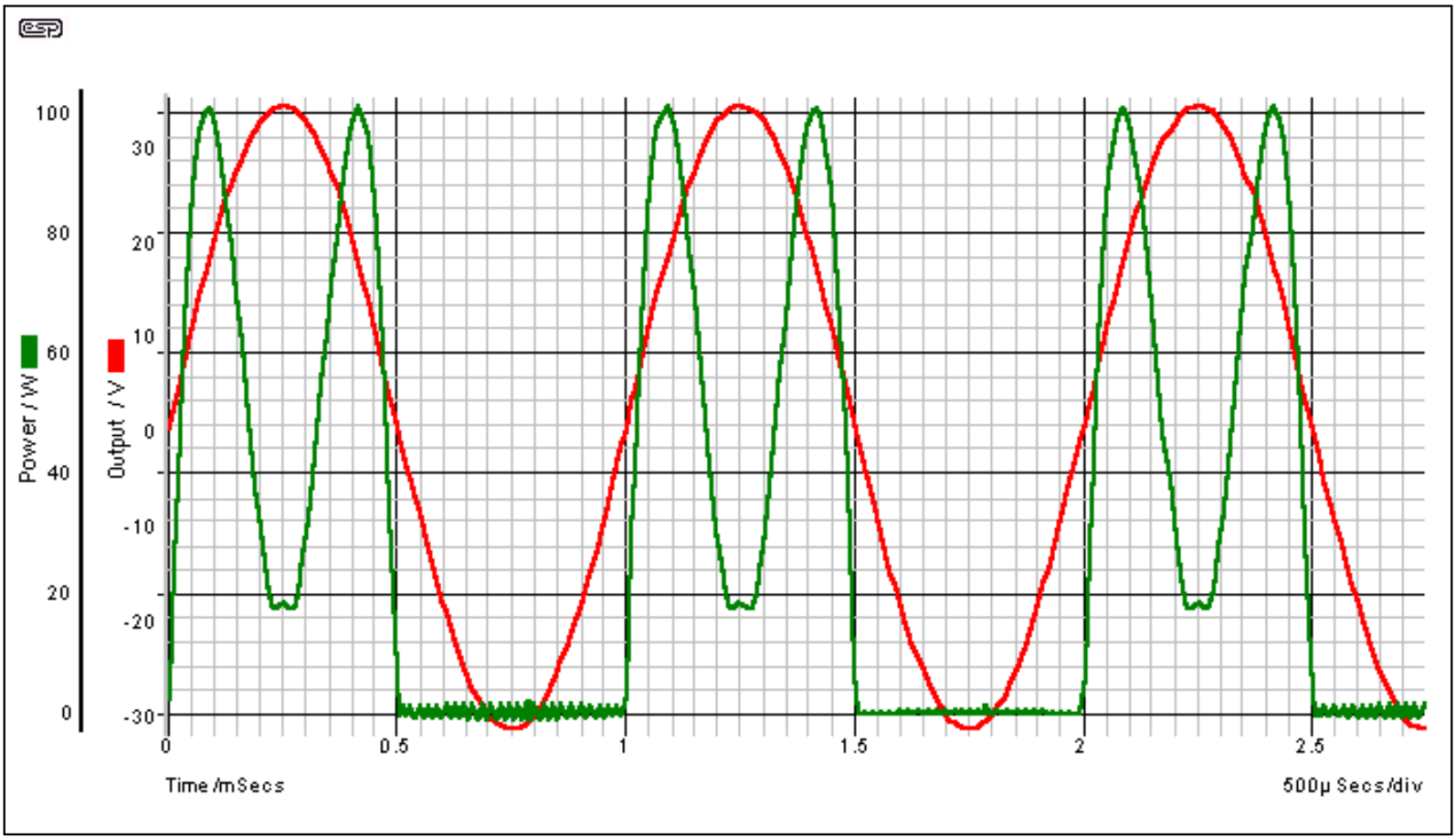


Figure 2.2 - Transistor Power Dissipation

The peak power is (using 6 ohms and 35V supplies) $(35/2)^2 / 6 = 51W$, or 102W for a 3 ohm load. Average dissipation is difficult to calculate because of the waveform, but my simulator tells me that it is 15 and 30 watts respectively. Note that for the simulations, zero quiescent current has been assumed - in real circuits, this just adds to the average dissipation.

2.2 Loudspeaker Loads

Where things rapidly get out of hand is with the loudspeaker load - it is not resistive (or even close to resistive) for 99% of all loudspeakers. The impedance and phase angle of a loudspeaker varies, and as phase angle changes from zero degrees (voltage vs. current), dissipation increases further.

For the 3 ohm case, a reactive load (at 45° phase angle) can be simulated by using a 477uH inductor in series with the 3 ohm load (for a frequency of 1kHz). With this combination, the peak transistor dissipation is 200W, with an average of almost 42.5W - note that the peak transistor power has doubled, and the average has increased by 1.414. Of particular interest is that the maximum power occurs at the voltage zero crossing point, when the maximum voltage is across the device. This is what causes transistors to fail, and the higher the voltage, the greater the risk.

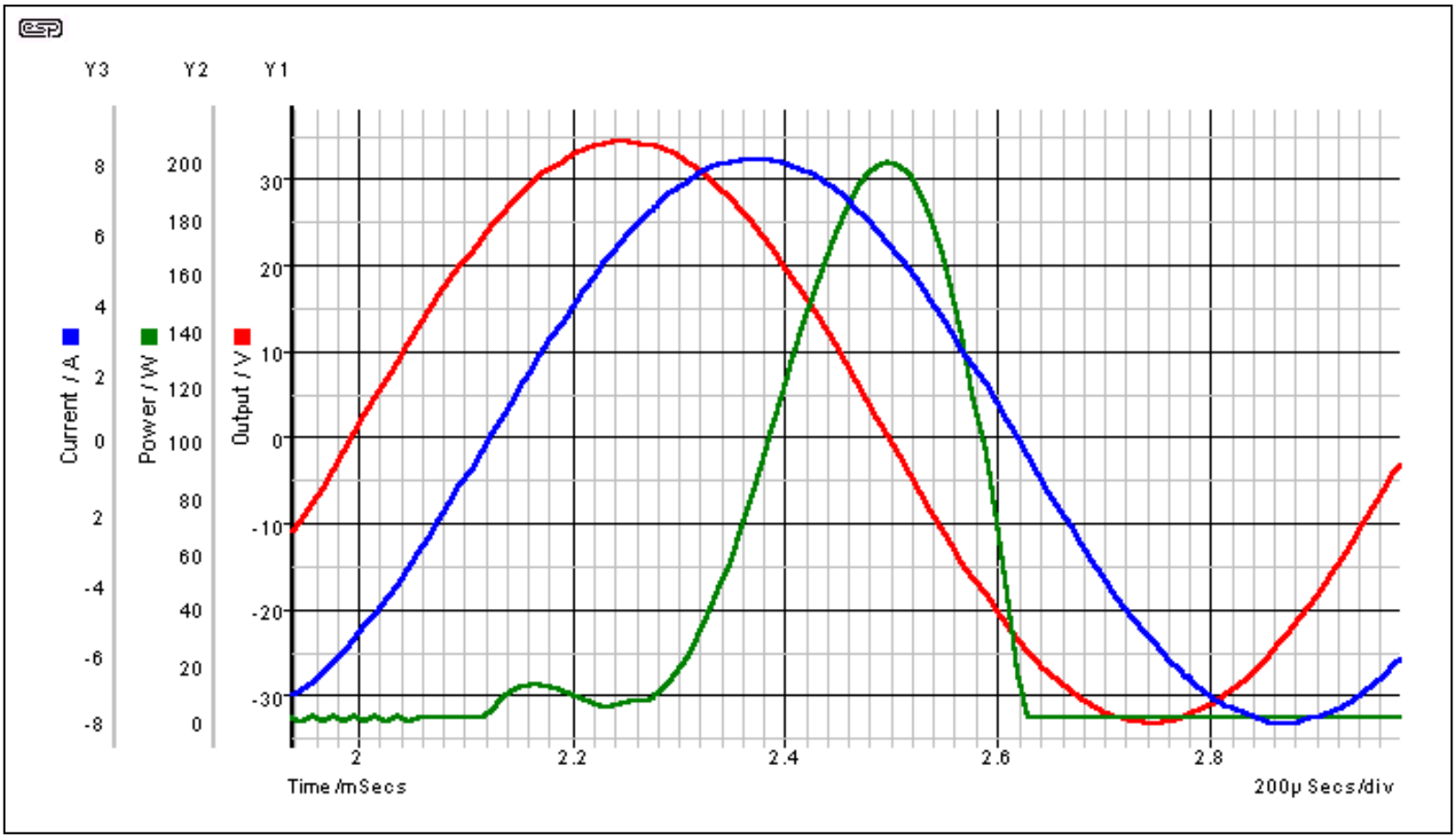


Figure 2.3 - Voltage, Current and Power Dissipation

While average power is well within the maximum ratings, the peak has reached the maximum device power, and we are now constrained by the SOA of the devices. Remember, this is with a supply voltage of +/-35V - higher voltages will create higher peak powers with real loudspeaker loads!

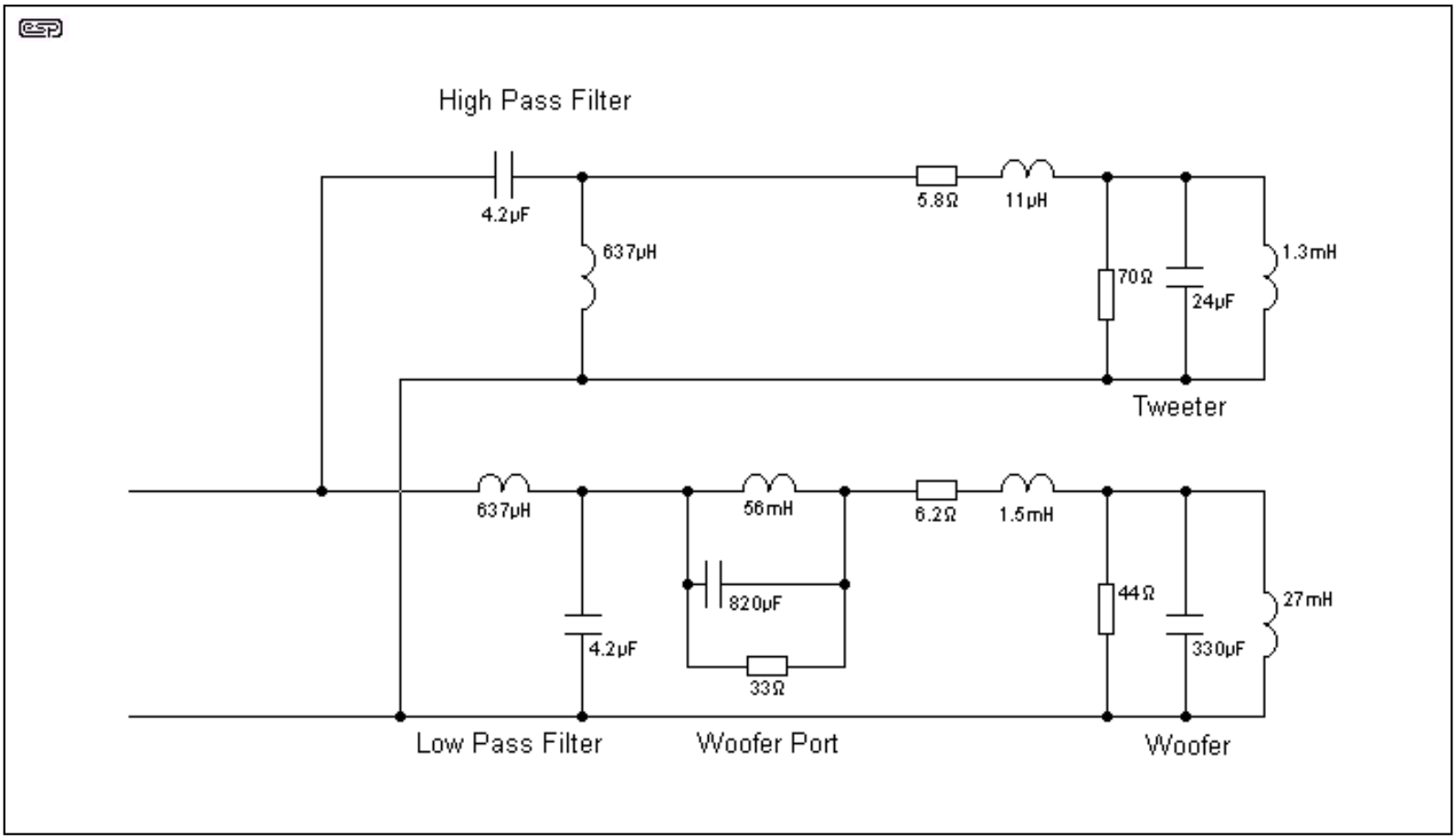


Figure 2.4 - Simulated Loudspeaker System

The combination of resistors, capacitors and inductors simulates the 2 drivers in the system, along with their crossover networks. For simplicity, no impedance correction networks have been included, and the loudspeaker is vented (note the double low frequency peaks shown in Fig. 2.5). This is a typical response, but remember that this is only the electrical response of the system - acoustically, it might be good, bad or indifferent (the electrical response gives clues, but cannot be used to predict the acoustical performance - I would expect it to be very ordinary however, based on the lack of impedance correction that is clearly visible looking at the phase angle and impedance curves).

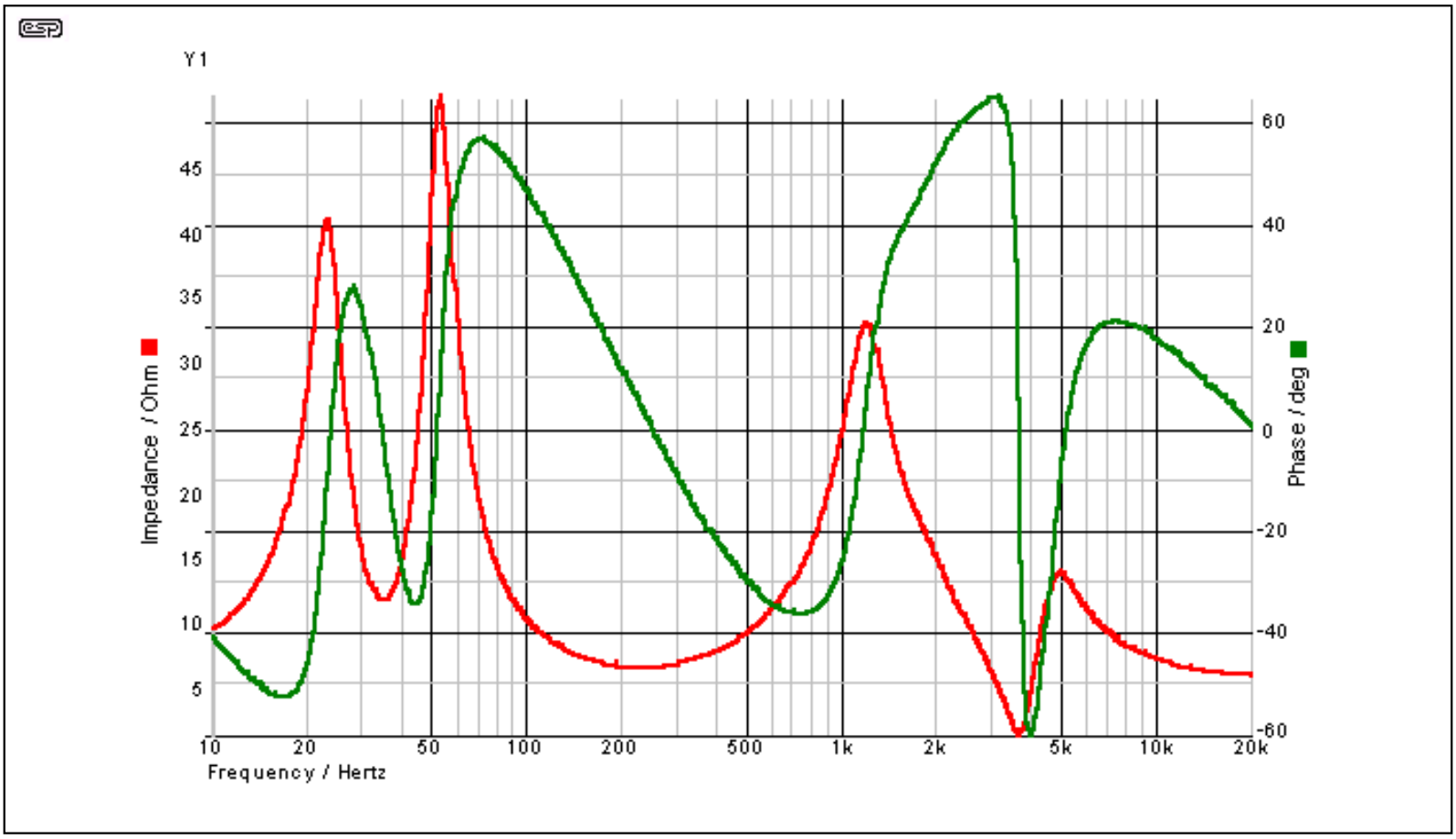


Figure 2.5 - Impedance and Phase Response Of Simulated Speaker

While the impedance is more or less as expected, the phase is another matter. At a phase angle of other than zero, the voltage and current are not simultaneous - the voltage may occur before the current (leading phase, capacitive load) or after the current (lagging phase, inductive load).

This is a major problem for amplifier designs, since at any phase other than zero, the power delivered to the load decreases, while the transistor dissipation increases. At 45°, peak transistor dissipation doubles, and power into the load is halved.

As the impedance rises with increasing frequency, the load appears as an inductor, and when it falls with increasing frequency, it is capacitive. Note how little time is spent at zero degrees phase shift! This means that at nearly all frequencies in the spectrum, the amplifier sees not a resistive load, but a reactive load. A significant number of frequencies are at or near 45° phase shift, so the amplifier is actually working nearly twice as hard as you thought it was.

3.0 Transistor Limitations

A bipolar junction transistor (BJT) has a negative temperature coefficient. As temperature rises, the junction voltage falls, and gain increases. Transistors are not perfect - there are always minute flaws in the fabrication, causing tiny variations in the characteristics of different parts of the transistor die.

Modern fabrication techniques have minimised these to a huge extent, but they still exist. Even the resistance of the conductive layers within the device becomes very significant at high currents, so perfect

current distribution cannot happen.

Now, there is a sequence of events than can (and does) occur within the transistor. If the instantaneous power dissipation is too high, parts of the transistor die will get hotter than others. This means that the junction voltage falls, and the gain increases - but only at the most sensitive part(s) of the die. If Vbe (base to emitter voltage) falls and gain increases - at one spot in the transistor - it will naturally take more of the current, which means it gets hotter, so it takes even more of the current (and so on). This can happen in a few milliseconds! That part of the transistor will quickly exceed the maximum permissible temperature, and the transistor will short-circuit internally.

All of this has happened in perhaps 10 milliseconds, and the case is not even warm. This phenomenon is called "second breakdown" (or secondary breakdown), and is the single greatest reason for transistor failure in a working circuit.

3.1 Second Breakdown

Data sheets usually have a full set of graphs and charts, showing the various device parameters as a function of voltage, current and frequency. In the design phase, all are important, but the most important of all are the two that are most often overlooked by hobbyists and experimenters - thermal derating and safe operating area.

From the data sheet for the MJL4381A, Fig. 3.1 shows the SOA curve for these devices. Non-repetitive peak currents of up to 30A are permissible for 10ms, but only for collector voltages up to 30V, and only with the junction temperature at 25 degrees. This is a peak power of 300W (the device rating is 230W), but it must be stressed that these conditions must not be allowed to continue beyond the time specified - 10ms is not very long!

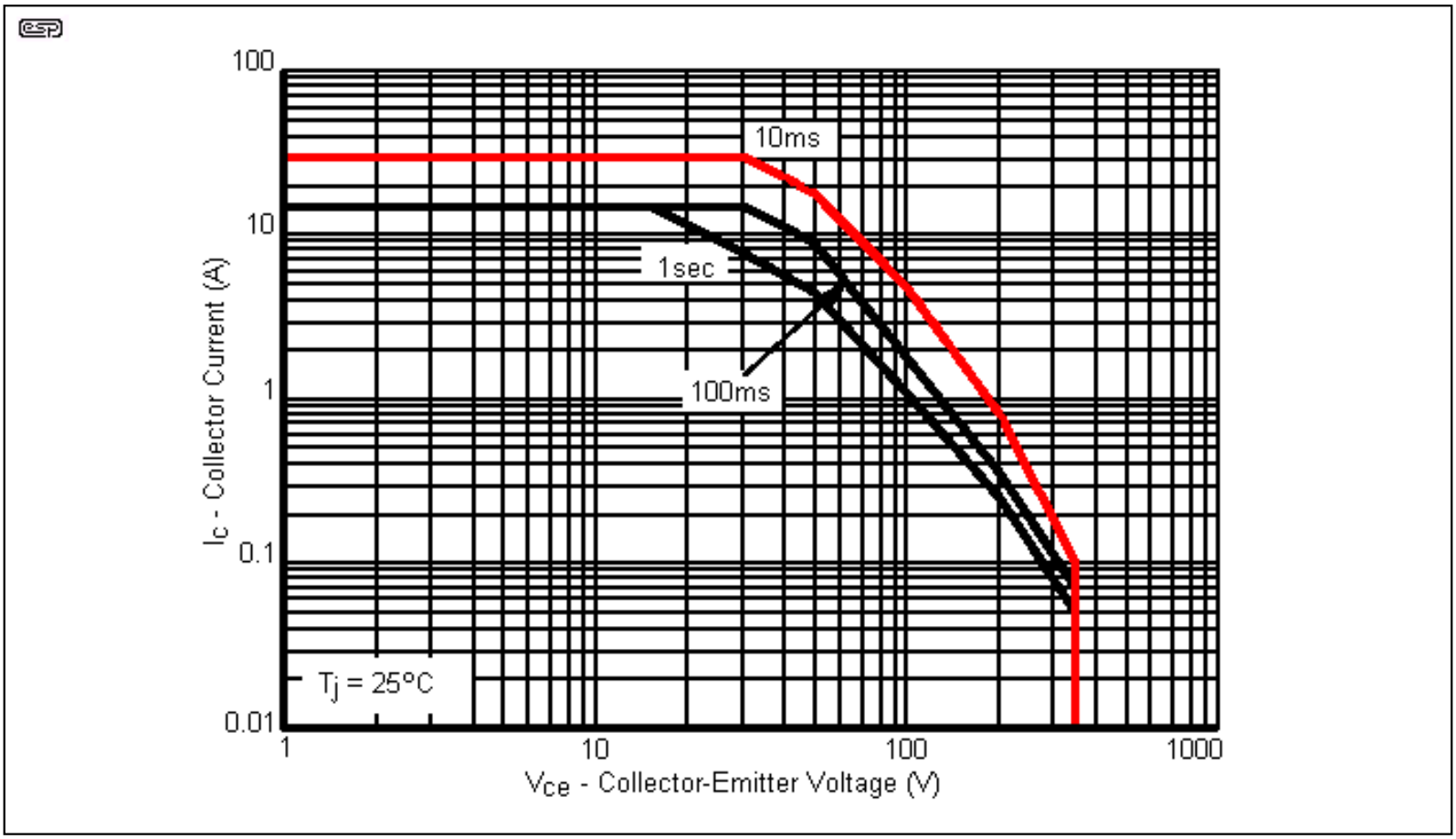


Figure 3.1 - SOA Curves for MJL4281A/4302A

If the time is extended, then the peak current is reduced for a given voltage, and for 1 second, the maximum rated current (15A) may only be drawn at collector-emitter voltages below 15V. This region is limited by the maximum rated current of the transistor, and will *never* allow continuous operation at maximum power (remember thermal derating? This is where it *must* be applied).

So far, all this looks pretty good if you look at it in conjunction with the demands outlined above, and it even *looks* as if it would be safe with 4 ohm loads at greater than rated ±35V. Appearances can be deceptive though! Remember that all peak currents and power dissipations referred to were for a junction temperature of 25 degrees - no transistor can maintain that temperature in real life, since there is thermal resistance between the die and case, and further thermal resistance between case and heatsink (see Heatsink Design for more information on thermal resistance and heatsinking of transistors).

The devices must be derated by 1.84° C/W above 25° (see Fig 2.1), which gives *zero* dissipation at 150° C. The thermal resistance from junction to ambient air (via the case, insulating washer and heatsink) can be expected to be around 1.5-2° C/W (for a *big* heatsink), so all dissipation limits quoted can be expected to be as little as 1/2 of those shown in the specifications.

That means that the 230W transistor is really capable of a dissipation of around 120W at typical (relatively high) operating temperatures. As a result, at ±35V with a 3 ohm resistive + 3 ohm reactive load (representing a typical 4 ohm speaker either side of resonance), the maximum limits will be exceeded!

Although this is completely true, in reality there are two things that will ensure that the amp remains functional (for many years) - the nature of music itself, and the collapse of the power supply under

sustained load. However, continuous operation at full power into a reactance that gives a 45° phase angle will cause the amp to fail, even with ±35V supply rails.

The variable nature of music, where the frequency and instantaneous amplitude are continually changing, means that potentially destructive signals do not last long enough to cause a problem, however increasing the supply voltage or reducing the load impedance further will almost certainly cause device failure.

Likwise, the bigger the heatsink, the better. The thermal resistances that cause the semiconductor die to operate at a much higher temperature than you may expect are the limiting parts of the equation - and they cannot be eliminated - at least not sensibly. It is generally considered uneconomical to provide a refrigeration system to keep the transistor temperature at low enough temperatures to avoid problems.

3.2 Device Parameters

Maximum Current: The emitter area determines the maximum current capability of the device, there are many design options for emitters, for audio transistors nearly all manufacturers use perforated emitter (also called mesh emitter) designs. The perforated emitter design also gives better gain linearity than a regular "interdigitised" emitter finger design (double or single comb). The other benefit for perforated emitter designs is silicon area utilisation, you can put a lot more emitter in a given piece of silicon with this design type for lower cost. One trade-off of the perforated design is switching, devices won't be as fast but we really don't need good switching capabilities for linear transistors and for audio this is a non issue. The bonding wire size depends on the current rating of course. Typical is around 15 mils (0.38mm) aluminium for audio high current devices.

Maximum Power Dissipation: Die size is the main parameter, ON Semi's Thermal Characterization Lab has done extensive studies and created some formulas for each package type so they can predict the thermal resistance (J-C) for any Die size in a given package. There are other factors besides die size that can affect the power dissipation, like solder line thickness, solder alloy, die thickness, etc. For good SOA performance good power dissipation is a must, thin die and very thin die attach solder are very important factors.

Second Breakdown: This is tough to determine and normally is determined by testing devices in a SOA tester by forcing power between collector and emitter and measuring the power dissipation time to secondary breakdown. Vertical structure of the device (collector and base thickness and resistivity) are important device design parameters, as well as die design geometry.

Current Gain: Emitter area determines the maximum current gain at high current levels of a device, too high peak hFE may result in lower BVCEO (Breakdown Voltage - collector to emitter, base open), higher hFE results in lower VCE(sat), VBE(on). As mentioned before, for good current gain linearity a perforated emitter design is best.

fT (Current Gain bandwidth Product): This is directly related to device gain and also to the device physical base width (wb). Most of the audio transistors in the industry have high fT (~30MHz), the trade-off is SOA performance with high voltage conditions. ON Semi Power Base Technology (which is unique in the market) has low/medium fT devices (8 to 12MHz) devices like the MJL21193/94 which have

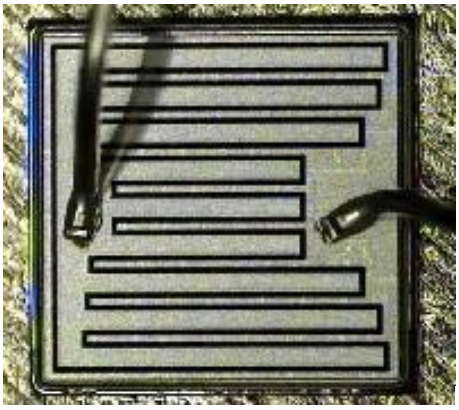
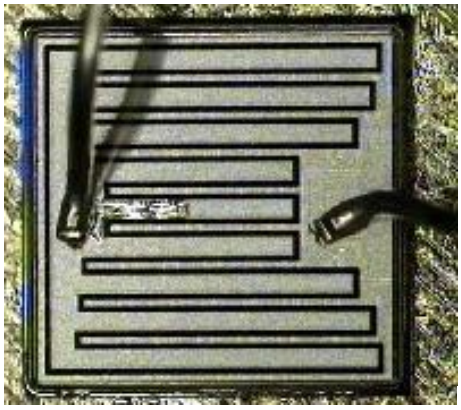
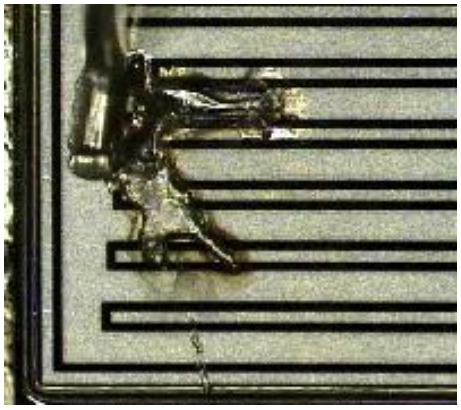
excellent SOA above 100V, these devices have wider Bases and also some unique "base spreading resistor" design which make them extremely rugged, used by most high end audio manufacturers.

FBSOA (Forward Biased Safe Operating Area): Die size, power dissipation, die geometry and base width are some important parameters.

(The above information kindly provided by ON Semiconductor)

4.0 Destruction!

The following photos show a typical (functional) die, and two shots demonstrating destruction. The shiny sections are melted silicon! This is also a good example of the "interdigitised finger" type of emitter and base construction referred to above.

Figures 4.1, 2, 3 - Transistor Dies (click for full size image)

The functional die (left) shows clearly what a typical transistor looks like. The emitter and base sections are clearly visible, with the emitter having the thicker "fingers" for best current carrying ability. This is not one of the new ON Semi transistors - the photo is representative only.

The damage in the failed die is quite obvious (centre), and there is a section of melted silicon where the transistor failed. As is the case almost 100% of the time, the transistor is shorted. Open transistors normally are the result of a bonding wire failure after the short has caused excessive current. This failed die would (probably) show the base junction as intact in a resistance test.

A close-up view (right) with greater damage. A large section of the die has exploded from the failure point outwards, and molten silicon has been sprayed all over the die. This failure would almost certainly indicate a short on all terminals (provided bonding wires are intact).

It is a sobering thought that these failures would have taken place in a matter of milliseconds - once the second breakdown region has been reached, the transistor will enter a negative resistance state, and there is nothing that will prevent total failure once the process has started. (Negative resistance is probable, but not a certainty - it depends to some extent on the fabrication method.)

(The above photographs kindly provided by ON Semiconductor)

4.0 Protection Schemes

A great many protection systems have been used over the years, in the hope of protecting the transistors from damage under all conditions. Power opamp ICs have the most comprehensive protection, but often at the expense of sound quality. By necessity, the protection must operate while the transistors are still quite safe, so their maximum power is never available.

Discrete amplifiers usually employ a simplified system, that will afford protection from most mishaps. A completely foolproof system is usually quite complex, and considerable care is needed to ensure that it does not activate during normal operation. This also applies to simplified systems, and a great many that I have seen do not provide complete protection at all - some are incapable of protecting against short circuits, unless at the end of a length of speaker cable (i.e. the cable's resistance forms part of the protection).

A further problem is that a full protection system will switch the power transistors off very fast, and given that the loudspeaker load is reactive, a "flyback" voltage can be developed that can easily destroy the transistors anyway. Many amps use a pair of diodes from the output to each supply - they are designed to ensure that the output voltage can never exceed the supply rails (except by the diode voltage drop).

In nearly all cases, it is necessary to either use additional output transistors, or tailor the protection circuit to ensure that excessive fault currents are not possible. This invariably means that there will be regions of the signal waveform where the protection circuit will operate when it should not do so. It is thought by a great many people that protection circuits degrade sound quality, and from tests I have done, this is certainly the case if (when?) they operate on any normal loudspeaker load. The *only* way to avoid problems is to use more output devices than you planned to, and even this is no guarantee that the amplifier will survive every abuse that it will face in a typical domestic or professional application.

Few discrete protection circuits monitor the temperature of the output devices (or the average power level) and thus adjust to suit the conditions. This means that a hot amplifier has a lower level of protection than a cold one, and it is no surprise that amplifiers fail most when driven hard for long periods (commonly as a result of a speaker failure). Thermal tracking is "automatic" in IC power amplifiers, since all devices are on a common piece of silicon.

Probably the best protection of all is a monitor that will operate if the output transistor SOA is approached, and removes the input signal. This is unfortunately much more difficult than it may seem at first, and the signal switching circuit is another candidate for sound quality degradation. Relays cannot be used, as they are not fast enough - remember, faults lasting only one millisecond can be sufficient to cause failure.

Fuses are used in amplifiers to prevent fire and further damage - no fuse is fast enough to protect an amplifier against fault currents, unless it is so low in value that it will blow during normal use (and even that is very doubtful).

Some "high end" amplifiers (where cost is no object) use a vast number of transistors, and ensure that at no normal (or abnormal) load will they ever exceed perhaps 1/2 their maximum rating. For typical consumer amplifiers (and most professional amps as well) the cost is a primary consideration, and transistors are run to their limits - to do otherwise would make the amplifier uncompetitive in the market.

A typical protection circuit is shown in Figure 5.1 and it is representative of the majority of those in common use. By sensing the current through the emitter resistor, the circuit detects an over-current fault, and removes base drive from the Class-A driver. As the voltage across the output transistor(s) is reduced, the sensitivity of the protection circuit is also reduced, allowing the maximum current at lower collector-emitter voltages where second breakdown is not a problem.

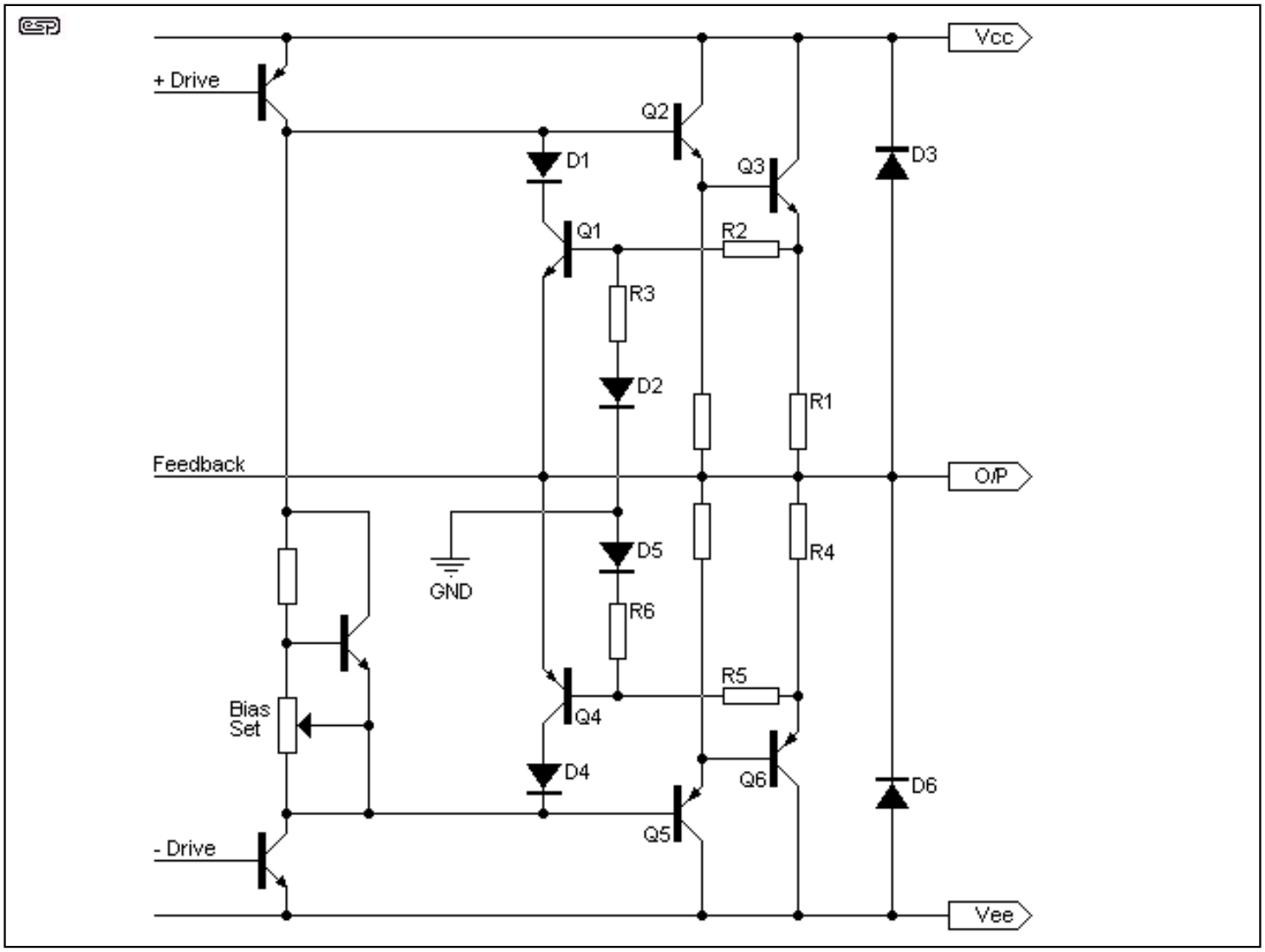


Figure 5.1 - Typical Output Protection Circuit

How does the circuit work? It is fairly straightforward to explain. Looking only at the upper section, R1 is used to sense the current through Q3, and if it exceeds about 0.65V, Q1 will turn on, "stealing" base current from Q2 (and thence Q3). D1 isolates the circuit from the drive circuit under normal operation. At zero volts output, the current through Q3 must be limited to below the danger level on the SOA curve. With a 35V supply, there are no major problems, but above that, the SOA is severly limited. For example, at 50V, the maximum current is 4A, so R1 must be chosen to provide 0.65V at (or below) 4A - remember that the SOA is limited further with increasing temperature, so 2A would be safer.

As the output voltage increases, R3 (via D2) shunts some of the current sense voltage to ground, reducing the effect, and allowing more current. The danger zone (and the cause of most of the problems) remains at 0V though - with a typical loudspeaker reactive load, a significant current is needed at 0V. Even with 35V supplies (as described above), we need as much as 6.5A when the output voltage is at

zero volts (see Fig 2.3). The problems caused by phase angle are such that it is almost impossible to design a current limit circuit that will allow maximum power, but still provide protection for shorted speaker leads, unless there are more output devices than appear to be required.

Note: At ±35V, an amplifier such as P3A is completely happy, and with high power (200W) transistors is operating within the SOA curve at all times and with any load (down to a typical 4 ohm nominal impedance). There is sufficient reserve power capacity to enable the amp to withstand full power into 4 ohm loads even at reasonably elevated temperatures.

However, once the supply voltage is increased, much of the reserve will be used up, making the amp liable to failure. This applies to any amplifier of similar ratings operating into typical loads - P3A has been used as an example, but the same constraints apply to any other design, used in the same way.

With careful component selection, circuits such as the above can work quite well. A good design will be conservative (and will therefore need additional output transistors), and will be reasonably effective in all cases. If the designer tries to get as close as possible to the transistor's ratings, the safety margin is reduced, and protection is less effective, Some legitimate signals will cause limiting, and other loads (especially at elevated temperatures) are likely to cause the transistors to exceed their ratings. There is a very good chance that the amp will survive regardless for many years, as the danger point may never be reached in some installations - others will destroy amp after amp, until one is found that can handle the abuse (or the abuse is removed).

6.0 Conclusion

It is obviously imperative to avoid second breakdown, and there are many ways that various designers have selected to do so. Protection circuits, Class-G (using two or more supply rails), variable supply voltages, and even switched supply voltages - these are common in many home theatre amps, and a switch is used to select the voltage to suit the load impedance.

There is also the "brute force" method, where there are so many power transistors that the cables will melt before any one transistor's ratings are exceeded, but this is uncommon except in high end amps where the added cost is not considered a problem. Many amps provide no protection at all, other than ensuring that dissipation limits are observed, but a shorted speaker lead (or a lower than recommended load impedance) can cause the amp to fail.

Regardless of the method used, it is important to ensure that the designers' recommendations are followed - good output transistors are expensive, and few of us can afford the luxury (??) of blowing up amplifiers for the hell of it. While a design that exceeds the transistor ratings may last for many years, there will eventually be a combination of circumstances that will cause failure. Parties are a prime cause of blown amps and speakers, and if they feature regularly in your activities, a cheap system (that can play loud, but very basic and with passable fidelity) is highly recommended. Its failure is not something you would cry over, and the main system remains intact always.

Finally, it is important to stress the importance of the SOA curve for any transistor used in an output

stage (including driver transistors). Any design that appears to be able to get more power from smaller transistors has almost certainly pushed the devices to (or beyond) their limits, and when driven hard into a difficult load, it will most probably fail - this is an expensive exercise if it takes the loudspeaker with it (not at all uncommon). Ultimately, a "worst case" design procedure assumes that the amp will be driven hard into a difficult load, and with undersized or barely adequate heatsinks. Such a design will survive - others will not.

7.0 References

I am indebted to ON Semiconductor for reference material, die fabrication details, photographs, semiconductors and data sheets used in preparation of this article. Photos and other material provided by ON Semiconductor are used with their permission.

Further information is available in the ON Semiconductor application note AN1628-D in PDF format.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams (other than that material which is Copyright © ON Semiconductor), is the intellectual property of Rod Elliott, and is Copyright © 2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 08 August 2003

Elliott Sound Products — Cables, Interconnects & Other Stuff - The Truth

Rod Elliott - Copyright (c) 1999 - 2004

Page Last Updated - 29 Oct 2004

Main Index

Articles Index

- **Part 1**
  - Introduction
  - Preamble
- **Part 2**
  - Speaker Leads
- **Part 3**
  - Interconnects
- **Part 4**
  - Power Leads
- **Part 5**
  - Audiophile Capacitors
  - High Current Amplifiers
  - Monoblock Amplifiers
  - Power Supplies
  - "Special" Designs
  - Opamps
  - Valves (Tubes)
  - Speakers
- **Part 6**
  - Conclusion

- Further Reading
- References

- **External Sites**
  - The ABX Data Page

When it comes to cable constructions, everything makes a difference. Most (but not all) of these differences are measurable. What is at issue is whether these differences are audible ... or not, when tested properly using a blind A-B test. Sighted tests are at best unreliable, and at worst cause people to believe things that are simply untrue. The vast majority of all cable claims have no basis in reality, and rely on the placebo effect.

Introduction

"I refuse to prove that I exist," says God, "for proof denies faith, and without faith, I am nothing."

> The Hitchhikers Guide to the Galaxy, Douglas Adams

The above could just as easily be re-phrased - for example ...

"I refuse to prove that my cables will make your system sound better", says the snake oil vendor, "for proof denies faith, and without faith, you will hear nothing."

The tenets of faith are an absolute requirement for many of the claims that are made for many (probably most) of the "esoteric" hi-fi additions that you will find everywhere on the web. There is no real information, technical, scientific or otherwise, and the only terms you will hear will be of a subjective nature - for example "solid, sparkling, sweet, musical" will be contrasted with "muffled, veiled, grainy, harsh" - the very selection of the words is designed to sway you to their position, preferably subconsciously.

The marketing is often very subtle, extremely persuasive, and there is no confusing techno-talk in there to confuse the non technical reader. While it may seem like Nirvana, the claims are nearly all completely false.

Faith (in the religious sense) is based on the premise that faith is God's proof that God's existence is truth and does not rely on facts. Indeed, if facts were available, then faith is

not required - so in a sense, faith can be seen to be based on an absence of evidence - a fiction.

Believers may also qualify faith as either representing truth or they will represent it as being above and beyond our understanding. Truth becomes a consequence of faith which is the believer's recognition of the absence of evidence. Truth is therefore defined according to a circular perception.

I am not about to dispute the religious beliefs of anyone - these are sacrosanct, and belong to the individual alone. When the same arguments are used for audio, this is a different matter. Audio (unlike religious beliefs) is based on science. Without the efforts of scientific work and studies over many years by a great many people, we would not have audio as we know it. Now, we have charlatans and thieves claiming that science is ruining audio, and that we have to get back to the basics to enable real enjoyment.

You need, *nay!* ***must have!*** the latest shiny rock on top of your CD player, lest the sound be harsh, grainy, and lacking bass authority, and without the latest cables at only US$200 per foot, you are missing out on half of the music. But ... you must believe, for the magic will surely be dissipated instantly should you attempt even the most rudimentary scientific test, or even request any technical information.

Now, consider the situation with watches. Has any ultra-high-priced watchmaker ever claimed that the "quality" of the time told by their watch is superior to that from "ordinary" watches, or that the "sense" of the time has greater depth and more "chi"? Maybe they just haven't thought of that angle yet, but I expect that this is unlikely. The simple fact is that these pieces of jewellery are finely crafted and superbly executed timekeepers, but are usually no better or worse that "lesser" brands that do exactly the same job.

The situation with cables is no different - you may *choose* to pay outlandish prices to get something that looks amazing, and demonstrates to everyone how much money you have, but it will not make a magical difference to the sound, there will be few (if any) real differences in the electrical characteristics, and it will sound much the same as "lesser" cables, selling at perhaps 100th of the price.

If image is important to you, and you can afford it, then that is your choice - just don't expect that it will make your system better, and don't try to convince others that without "it", they are missing half their music or their sounds are being mangulated in some mysterious way that can only be "fixed" by spending vastly more than they may be able to afford.

Note: It must be considered that there are some people whose hearing acuity is far greater than the average, and they may well hear things that we 'mere mortals' cannot. For such individuals, a particular cable might indeed show an improvement (or at least a difference), but this does not mean that the same improvement/difference will be audible to anyone else.

The majority of this series of articles is directed at the majority of listeners - no surprise there. Just because some rare person with hearing that is well above average can hear a difference does *not* mean that everyone will do so, although it is unlikely that anyone will admit to being unable to distinguish one from another. No-one wants to be classified as being 'tin-eared', and especially so if they have spent a lot of time and money on their system.

Yes, there are a (very) few people who can genuinely be considered to have 'golden ears', just as there are a few musicians who have perfect pitch, and various other individuals with a particular skill in some area that most of us lack. Just as no-one will normally reject the photographs taken from a camera (for example) that one person can see are ever so slightly flawed (but look fine to us), then nor should we reject a cable that sounds just fine.

Indeed, the variations in different recordings (even of the same material - and especially so with vinyl!) will be far greater than the variations of any cable with reasonable construction and sensible design.

Despite what you may read in various forum pages, this entire series of articles is not intended as a "beat up the subjectivists" tale, but rather a discourse based on research that I, and a great many others before me, have done. The idea is not to ruin anyone's enjoyment of audio, but to make sure that the facts are available, without the hype and BS so commonly associated with high fidelity.

The major (and well respected) audio companies did not develop their equipment using only their ears as a guide. Without exception, all the big (and very expensive in many cases) brands have been measured, probed, simulated, then measured some more - before anyone actually gets to hear one. How much of this pure research has gone into most of the overpriced cables and "accessories" currently available? I don't think I need to answer that, as we all have a pretty good idea.

So much has been said about cables over the past few years that there couldn't possibly be any more to discuss. Nice theory, but the wheel has turned a full circle, and there are now people claiming that there is no difference at all between any speaker cable or interconnect. In exactly the same way as the claims that there were "huge differences" were mainly false, so too are claims that there are none.

There is no "black and white" in this topic, but a great many shades of grey, and the latest update to this article attempts to clarify the position. Speaker cables in particular are still a major topic of conversation on many forum sites, and remain one of the more contentious issues.

A quick summary of the topics to follow (in the cable discussion, at least) would be ...

- Power leads will rarely (if ever) have any effect on the sound, provided they are of reasonable construction and are not inducing noise into (unshielded) interconnects. The only exceptions are those that use filters of some sort, which will reduce the noise floor in areas where interference is a problem.
  Some leads are of flimsy construction, and may reduce the available power for sustained loud passages, however, the difference will rarely exceed 1dB in most cases.
- Speaker cables *can* (and sometimes do) sound different with a given amplifier and loudspeaker combination, even where they are well designed and of reasonable gauge. Excluded are very thin or extremely silly combinations - these will always do something to the sound, rarely good.
- Interconnects *might* sound different, but only if they use odd construction techniques. Generally speaking, all properly (sensibly) designed and well made interconnects will sound the same - excluding noise pickup which is common with unshielded designs.

This is not to say that some people will not derive great enjoyment from the fact that they have spent as much on their cables as mere mortals can afford for their whole system, but this is "enjoyment", and has nothing to do with sound quality. This is about prestige and status, neither of which affect the sound.

Try This Next Time Someone Tries to Sell You Something ...

Thanks to a reader for the suggestion, this is a wonderful way to prove something to yourself. Next time a salesperson tries to flog you the latest and greatest (and of course most expensive) cable they have on offer, just use this technique ...

Suggest that you would like to hear the cable in action before committing yourself. As you walk to the demo room with the salesperson, come up with 'spontaneous' bright idea - suggest that you swap the cables, and if the salesperson can correctly identify the 'super cable' that s/he so desperately wants you to purchase, then you will do so. Naturally, you will want to make the swap several times, and the salesperson will have

to get it right at least 75% of the time.

There is every chance that the packet will never be opened, the comparison never done, and you will save a bunch of money. There is nothing dishonest about what you are doing - you simply want (and are entitled to) verification that the cable will make a difference, and if the salesperson is unwilling to participate in the test, s/he knows something that s/he hasn't told you!

Beware! If there is any suggestion that the cable needs to be 'broken in' before you hear the difference, the salesperson is lying! At this point, you should immediately let them know that you know that they are lying, and leave the shop. Cable 'break-in' is a myth, and is perpetuated by those with something to hide - no-one has *ever* been able to show that there is any scientific justification to the claim, nor shown that the performance has changed in any way whatsoever. Cable break-in is real, and occurs between the ears of the listener - nowhere else (most certainly not in the cable).

Preamble

The last link entry for the ABX Home Page has been included so you can have a look at some actual ABX double blind tests that have been carried out. The listing at the ABX site is not extensive, but is excellent reference material. You will find some of the results surprising, and when viewed and interpreted sensibly, they tend to support the comments I have made in this article.

In some cases, the results surprised me, in that I was expecting the listener panel to declare various items as different, and they instead thought they were the same (which is to say that the two items under test could not be identified with certainty, so any choice was pure guesswork).

In this article, I shall attempt to explain some of the misconceptions and untruths that are rife in the audio industry. This article is bound to offend some, but the information is based on fact, scientific data and the results of my own (and others') testing, plus the help I have received from readers, who have provided more information on a number of topics.

In contrast, much of the disinformation comes from the rantings of Hi-Fi reviewers, most of whom know so little about electronics that it is shameful (and fraudulent) for them to be in a position to tell the unsuspecting public what to buy, based on entirely subjective criteria.

In almost all other areas of human interest, objective measurements are paramount. A domestic vacuum cleaner's performance is based on how much dirt it collects from the carpet - any philosophical discussion about the type of motor used, or it's rotational direction having a subtle effect on how clean the carpet feels is at best a pointless and tiresome exercise, and (I hope) has never been entered into.

Discussion - indeed, heated debate - on parameters not dissimilar to those above are commonplace in the high end audio industry, and have been raging since the late 1970's. The majority of people who listen to music generally listen to a few systems at a non-specialist retail outlet, and buy a combination that sounds good (to them), has the features they want, and fits their budget. They are no more interested in the great audio debate than they would be in the philosophy of the rotating mechanical components of their vacuum cleaner.

In his article "Science and Subjectivism in Audio", Douglas Self [1] wrote

> A short definition of the Subjectivist position on power amplifiers might read as follows:
>
> - Objective measurements of an amplifier's performance are unimportant compared with the subjective impressions received in informal listening tests. Should the two contradict the objective results may be dismissed out of hand.
> - Degradation effects exist in amplifiers that are unknown to engineering science, and are not revealed by the usual measurements.
> - Considerable latitude may be used in suggesting hypothetical mechanisms of audio impairment, such as mysterious capacitor shortcomings and subtle cable defects, without reference to the plausibility of the concept, or gathering any evidence to support it.
>
> I believe this is a reasonable statement of the situation. Meanwhile the overwhelming majority of the public buy conventional hi-fi systems, ignoring the expensive and esoteric high-end sector where the debate is fiercest.

In the following articles I shall dissect some of the claims made on many of the components in the audio chain, and show why they are misleading, false, and in many cases downright dishonest. See Further Reading for ... well, further reading.

Preamble Part 2

A fairly well known person (rampant on certain forum pages) has claimed that I consider *all* conductors and insulators to be "perfect", and that "all engineers who design in the

real world know this is not the case". Oh really! ... and where **exactly** did I say that all conductors and insulators are "perfect"? Where did I *imply* that they are perfect? These questions remain unanswered (of course) because I have never claimed, assumed or implied that they are perfect.

No insulator or conductor is perfect - in fact, no "anything" is perfect. The simple fact of the matter is that these imperfections are not significant at audio frequencies, except perhaps in "unusual" cable constructions (of the type often suggested by the lunatic fringe). This is one of the typical "red herrings" that raving psychotics will bring up time and time again, to bolster their unsubstantiated and flawed "reasoning". Claims like that are typical of delusional thinking, and the delusional only have to claim that I (or someone else) said that "all conductors and insulators are perfect" (for example), and it somehow makes it "true" that these words were in fact used.

Well, I have some news that may come as a shock - anyone can *say* anything they like, but the saying does not make it so! I have *never* claimed that all conductors or insulators are perfect, but I have challenged anyone who claims that the imperfections are audible to please do so. So far, there has not been one shred of evidence that indicates that Teflon™ (wonderful stuff that may well be) is **audibly** superior to PVC in a properly controlled double-blind (or ABX) test. Differences are measurable (with the right equipment) but are not relevant to the audio range unless the "facts" or cable topology are manipulated to influence the test.

I have asked every person and/or company named in the Mad As Hell articles for any information they have that substantiates their outrageous claims, and not one, not a single **one**, has supplied anything more than some useless promotional material or "satisfied customer" e-mails. Why is "satisfied customer" in quotes? How do I, or anyone else, know that they are genuine? For all we know, they are fabricated (i.e. lies), without an iota of truth in any of them. Oh, but I am so *negative*!

Of course I am, these people are liars, charlatans and thieves, either by accident (they may actually *think* they are realistic because of mental illness [such as delusion or psychosis] or some other mitigating circumstance) or by design - they simply have one goal ... to separate people from their money. The actual "mechanism" is unimportant - the fact that they are wrong does not enter into their equation of life, so whether their claims are due to mental illness or greed makes no difference to the consumer, who is being ripped off and lied to either way.

I recently had an e-mail exchange on the topic of interconnects, and the "conversation" started out innocently enough. I was advised that by using the tape loop on a preamp, I could listen to the effects of different interconnect cables, simply by switching to/from tape monitor.

I firstly suggested the test methodology suggested was flawed, since any additional circuitry used to make up the tape loop circuit would have some influence. In addition, the feedback to the brain (knowing which switch setting was which) means that a genuinely objective (double blind) test was impossible. The test method does not even qualify as single blind - it is an open test, and the experimenter expectancy effect will confer non-existent attributes to the material being tested, based on preconceived ideas and expectations.

The e-mails went back and forth for a while, and eventually I was finding that it took up too much of my time, and the topic is not all that interesting anyway - after all, how excited can one get over ordinary signal leads.

This is doubly true when the other party invents reasons that ABX tests are "invalid" for audio - something about the signal complexity, and the psychological effects of the music was mentioned. This is exactly why we must use ABX or similar double blind tests - anything else will fail to properly eliminate feedback cues, and these will be used (albeit subconsciously) to determine whether the "standard" or "test" item is currently in circuit. Any test where there is any possibility of identifying the components under test is completely invalid.

It is interesting that in a relatively non-demanding application such as an interconnect, a material such as aluminium would likely be sneered at by any audiophile, yet this very same material is used regularly in loudspeaker voice coils. I am reasonably sure that sonic performance of an aluminium interconnect would be deemed to fall way short of excellence, yet I hear (or read) no highly critical comments about using it in a voice coil. This is an extremely demanding role, and the performance of aluminium is (or can be) audibly and measurably worse than copper. *

My (almost) final e-mail pointed out that no metallic conductor introduces distortion. Now, I must admit that I did not qualify this, but when I speak of distortion I refer almost invariably to non-linear distortion (i.e. the type introduced by all active components, that generates harmonics and intermodulation products not present in the original signal). A simple question would have cleared this up, but ...

The response I received astonished me - suddenly, my statement that "no metallic

conductor introduces distortion" was utterly misconstrued, and became "all metals are perfect conductors"! It was inferred (of course) that this was the reason that my tests and experiences are simply invalid, while those of my correspondent were reasoned and obvious.

This is absolutely the sort of thinking that got everyone to this impasse in the first place. I never suggested that all metals are perfect conductors - I said that they don't generate (non-linear) distortion. By means of misinterpretation, the subjectivist camp will now think it has another "weapon" against the enemy - the fact that it is the result of a gross mangling of the original statement is of no consequence ... "Never let the facts get in the way of a good story".

The fact of the matter is that no metallic conductor causes (non-linear) distortion. There are various resistances depending on the metal, but it's basic conductivity is completely linear. Check things like thermal coefficient of resistance for any metal - it is linear. There are no curves or "fudge factors" to be taken into account. While it may be possible to make an alloy that exhibits some degree of non-linearity, this would not be used as an electrical conductor, and would certainly not be suggested as an alternative to copper. Even then, within the very limited range of acceptable temperatures in the listening room, such non-linearities could easily be less than that of air - the medium that carries the sound from the speakers to our ears..

None of this has anything to do with skin effect, velocity factor or any of the other seemingly strange behaviours of all conductors at high frequencies (none of which are non-linear distortions), we are interested in the simple ability to conduct current from point A to point B without any form of rectification or other non-linear effect. All metallic conductors in common use will do this perfectly well, and will not add harmonics or change the waveshape in any non-linear way.

Harmonics can of course be removed - this is a filter effect (a completely passive linear function), and is caused by capacitance and inductance. All cables have these parameters as a fact of life - a silver wire and an aluminium wire of the same length and diameter have different resistance, but inductance and capacitance are the same.

The degree of hostility I experienced towards ABX testing was equally puzzling. I don't know of any designer who will claim that listening tests are invalid - only that they may not reveal the entire truth of the matter, and that additional "technical" evaluations may be needed to find out why the listening tests did (or did not) correlate with the measurements.

On the other hand, many subjectivists claim that anything ***other*** than a listening test is invalid, and commonly and even vigorously eschew ABX testing - possibly because they know in their hearts that they will be unable to find any difference. This is very confronting, and to have one's beliefs shattered is not a pleasant experience.

What is the most interesting to me is the "head in the sand" behaviour. I was automatically wrong in my thinking, and I suspect that anything that I said would have been twisted around to make sure that I stayed wrong. I could (of course) have simply agreed with the subjectivist's position, however to have done so would have been a lie on my part.

The issues at stake here are the crux of the on-going debate between the two "camps". While I will admit that not all designers will take any subjective opinion seriously, I do know from my own testing and from a huge amount of reader feedback that some of my designs sound better with different transistors or power supply configurations (for example). Most of these differences can be quantified, although some are elusive, and that is something that I live with, knowing that many of the further "tweaks" are assessed by purely subjective methods. There is every chance that ABX testing would reveal no audible difference.

*** More on Aluminium (Aluminum)**

I mentioned above that aluminium interconnects would generate scorn and derision from the audiophiles. Well, it seems that for some, even using it for shielding is bad ...

> "Unused RCA inputs on the back of [amplifiers] are prone to pickup stray RF Interference and EMI. This can cause a higher level of background noise, haze and grain. For years audiophiles have used shorting plugs or (gag!) aluminum foil, to remedy the situation. Unfortunately, many preamps do not like to have their inputs shorted. What to do?"

Wonder what "gag!" implies - I think I can guess. Needless to say, the answer was in the product line for the site in question - I shudder to think how much their little RCA "hats" cost.

I saw remarkably few references to aluminium even being used (let alone sounding "bad") in interconnects, and no adverse comments at all about it's use as a voice coil winding wire. I must confess that I did not spend a vast amount of time on this, partly because as I said early in this section - cables are just not very interesting :-)

How does this thinking occur? An excellent article on the human belief system is The Belief Engine, which is to be found at http://www.csicop.org/si/9505/belief.html. The

article describes the mechanisms we use to generate beliefs, and the ways that these beliefs are reinforced as we go along. One tiny quote from the article ...

> Our brains and nervous systems constitute a belief-generating machine, a system that evolved to assure not truth, logic, and reason, but survival.

What does survival have to do with interconnects - nothing at all, of course. But this does not change the way we think, and especially does not change the way we think we think. Beliefs are extremely powerful, and can be almost impossibly difficult to shed once they have become entrenched ... I have no expectations at all that this article will change that one little bit, but if it helps others (not yet contaminated) to stay well clear of pseudo science, then I have done what I could.

NEXT >

Main Index

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999 - 2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Last Revision: 28 Oct 2004 - 'Try This Next Time ...'

The Audio Pages

Elliott Sound Products White Paper on Speaker Cables and Interconnects

By Roger Sanders
Reprinted By ESP with kind permission of InnerSound

Contents

- Introduction
- Speaker Cables Used With ESLs
  - Inductance
  - Capacitance
  - Impedance
- ESL Speaker Cable Design
- Woofer Speaker Cable Design
- Interconnects
- Interconnect Design
- Conclusion

Introduction

This article is a republished version of that originally written by Roger Sanders (InnerSound), one of the "fathers" of the electrostatic loudspeaker. While it may appear that this is an advertisement for InnerSound, this is not the case at all, but to remove the name would be to unfairly remove the appropriate references to the source of this material. I have no affiliation with InnerSound or Roger Sanders, and references should not be seen as an endorsement or criticism of their products - I have not heard the ESLs, preamps or amplifiers, so any further comment on this is not appropriate.

This material is not published without considerable thought and soul-searching on my part. It reflects many of the things I (and many others before me) have already written on the subject. It is presented here as a public service - the hi-fi fraternity needs to be

aware that a great deal (the vast majority - OK, *all*) of the hype about 'cable sound' is pure, simple, and unadulterated bulls**t!

There are a very few things that make a difference, and these are explainable, repeatable and measurable.

There is no basis in reality for most of the claims made, and a search of the ABX site will confirm this. It is to be hoped that this article (which has apparently already caused some flak) will help to clear up some of the gross exaggerations and dis-information that abounds on the WWW and in magazines. Most of us may think that the dissemination of information should be factual, but unfortunately there are many people who publish 'information' purely for their own benefit.

Note: I have removed the pricing details from the original article, as these are not relevant to the discussion. Otherwise (apart from substituting Australian spellings) the article is untouched in terms of the content.

Speaker Cables Used With Electrostatic Loudspeakers

Electrostatic loudspeakers (ESLs) are different. The load they present to an amplifier and speaker cables is quite unlike that of conventional magnetic speakers. To a speaker cable, they appear as a capacitor, while magnetic speakers appear as a combination of a resistor and inductor. It therefore is not surprising that cables for ESLs have different requirements from those for magnetic speakers.

Cables have inductance, capacitance, and impedance. Cable manufacturers juggle these parameters to get the cables to sound the way they want. Let's look at these elements more closely and see how they should be optimised for ESLs.

An ESL is driven by a high-voltage, step-up transformer. This transformer is inside the speaker and converts the relatively low voltage of an amplifier to the several thousand volts needed to drive an ESL. Unfortunately, all transformers have leakage inductance. This inductance interacts with the capacitance of an ESL to form an L/C (inductance/ capacitance) resonant circuit. This produces an undesirable, high-frequency peak in the frequency response of the ESL.

It is essential that this resonance be kept well above the audio spectrum to prevent the sound from being excessively 'bright'. Since the capacitance of the ESL is fixed, the only way to get the resonance high is to build a transformer with very low leakage inductance.

Designing and building very low leakage inductance transformers that will operate over a wide frequency range and at high voltages is extremely difficult. One of the reasons that some ESLs sound better than others is the design and quality of their transformers.

Inductance is a big problem with ESLs due to the L/C resonance described above. ESL manufacturers expend great effort to obtain transformers with low inductance. So it is vitally important that the cables have low inductance too. If the cables add a lot of inductance to the circuit, they can undo the transformer designer's best efforts.

Inductance

In a speaker cable, inductance is largely determined by the area between the conductors. Most speaker cables have conductors that run side by side ('twin-lead') and that are separated by a small distance, so have moderate inductance. They do not have the low inductance desired for the best performance when driving ESLs. Some cables use many small wires that are woven together. This reduces inductance greatly, but at the cost of increased capacitance.

Capacitance

The capacitance also should be low. This is not as critical as inductance, but it is important. Remember that an ESL is a capacitor, and amplifiers find capacitors very hard to drive. If the cable adds more capacitance, it only makes things that much worse for the amplifier. Capacitance is highly affected by how close the conductors are to each other. So to keep the capacitance low, the conductors must be widely spaced. Note that this is just the opposite of what we need for low inductance.

Many cable manufacturers deliberately add a lot of capacitance to their cables. For example, you will find a box at the end of MIT cables, which contains capacitors. Alpha Core (Goertz) cables are made as a sandwich with two ribbon conductors very close together, and this type of construction produces high capacitance and often, amplifier instability. Woven wires are close together so have high capacitance. These types of high-capacitance cables are best avoided when operating ESLs.

Impedance

Impedance is the resistance to the flow of current in a cable. Most cables are designed

to have low impedance so that they don't significantly reduce the damping factor of the amplifier. But some manufacturers deliberately use high impedance cables to alter the sound of the speaker by both interacting with the speaker's crossover and reducing the damping factor. When the damping factor is reduced, the amplifier cannot keep the woofer under good, tight control. The result is that the bass becomes 'loose'.

In the case of an ESL, it is best to use a medium impedance cable as this will damp the L/C resonance and reduce its magnitude. Since the L/C resonance should be supersonic, this damping effect may not be audible. But reducing the resonance will make life much easier for the amplifier. Of course, if the ESL's transformer is poor, the L/C resonance will be in the audio range and damping it with a medium impedance cable will help smooth out the high frequencies.

ESL Speaker Cable Design

InnerSound's cables are uniquely designed to meet the needs of ESLs in three ways. They have low inductance, low capacitance, and moderate impedance. How is this done?

Because the conductors need to be close together for low inductance, but wide apart for low capacitance, simultaneously obtaining low inductance and low capacitance seems impossible. But surprisingly, there is a solution to this problem. Coaxial cable construction runs one conductor inside the other. So electricity 'sees' the conductors in the same place. This results in very low inductance.

InnerSound's coaxial, low-inductance design is enhanced by spiral-winding the conductors in opposite directions. This further cancels inductance.

But what about capacitance? Doesn't a coaxial design place the conductors close together forming a high-capacitance cable?

Not necessarily. The conductors can be physically separated by a significant distance using a thick, high-value dielectric to produce very low capacitance while maintaining ultra-low inductance.

The impedance is determined by the size and length of the conductor. InnerSound sizes the conductors to obtain medium impedance in the typical range of cable lengths used by most audiophiles.

Woofer Speaker Cable Design

For driving the conventional, magnetic woofers used in hybrid ESL/woofer systems, the demands for low capacitance and low inductance are relaxed, although maintaining these parameters at low levels is still desirable. At the same time, the impedance needs to be low to maintain a high amplifier damping factor to achieve tight control of the woofer.

InnerSound's bass cables meet these criteria by using dual pairs of coaxial cables. This technique drops the impedance to very low levels while maintaining low inductance and capacitance.

Interconnects

All interconnects are NOT equal. There are some very specific features that interconnects should have. InnerSound offers the excellent interconnects with all the finest features. But all the hype surrounding interconnects makes it very confusing to know what is important. The purpose of this paper is to explain the facts so you can make intelligent decisions. And the facts can be quite surprising as you will soon see.

There is no doubt that speaker cables can exert a small influence on the sound of your audio system. But interestingly, all well designed interconnects sound identical.

The above statement sounds absurd, since interconnect manufacturers all claim that their products will make your system sound better. They also claim that different types of wire (copper, silver, oxygen free copper, etc.) sound different, how skin effect causes transient smearing, and how dielectrics change the sound. So the idea that all interconnects sound identical is outrageous.

Or is it? Have you actually done a well controlled test to verify their claims? I strongly urge you to do your own testing rather than taking my word for it. It is very simple and easy to evaluate interconnects. Let me show you how.

The idea behind the test is to make it possible for you to switch back and forth between interconnects instantly and repeatedly while all other components in your stereo system remain the same. You can then listen very critically for any difference in sound between the interconnects you wish to test.

You cannot accurately test interconnects by listening to one for awhile, then unplugging them, connecting another set, and listening again. Our 'audio memory' for subtle details is too short to accurately remember any differences in sound in such a test, and we cannot check repeatedly to be sure of what we hear -- so we are easily deceived. You must be able to switch instantly and repeatedly to hear any real differences between interconnects.

You do not need any test equipment. You can use your preamplifier to do the switching. You will need a Y connector so you can connect the two interconnects under test (let's call them "A" and "B") to the same component -- probably your CD player.

Note that the Y connector is the same for both interconnects, so even if you believe that the Y connector somehow corrupts the sound (they don't), the same corrupted signal will pass through both interconnects so the test will still be valid. Remember that we are only listening for any difference between the interconnects, and you can hear that difference (if present) on any signal, even a corrupted and distorted one. Inexpensive Y connectors can be obtained from Radio Shack. If you want audiophile grade Y connectors, Sound Connections International (phone 813-948-2907) sells beautifully built, gold plated units at reasonable prices.

Connect one end of interconnects "A" and "B" to the Y connector. Do so for both channels.

Connect the other end of interconnect "A" to one of your preamp line level inputs (such as "CD"). Connect the other end of interconnect "B" to your tape monitor input. Do so for both channels. Be sure you don't reverse the channels. All line level inputs on a preamp are identical, so it doesn't matter which ones you use.

You could connect the interconnects to any other line level input on your preamp instead of Tape. But the tape monitor inputs allow to switch back and forth between interconnects by toggling the tape monitor switch instead of having to press different input switches, or rotating a knob. Toggling a single switch is more convenient and makes it easy to do the test 'blind' so you don't know which interconnect you are listening to. Doing the test blind is desirable so your personal prejudices don't influence the test results.

If your preamp doesn't have a tape monitor function, then use any two line level inputs. If you have to use a rotary selector switch, use two inputs that are next to each other on the rotary switch so you can easily move back and forth between them.

The test is done by simply listening to music while switching back and forth between the two sets of interconnects as much as you wish. The idea is to try to hear any difference between the interconnects. There is no time limit, you may switch whenever you wish and take as long as you want.

The test is easiest to do if you have a remote control preamp so you can sit in your listening chair and simply push the Tape Monitor button on the remote whenever you want to switch to the other interconnect. If you don't have a remote control preamp, then you may need an assistant to switch for you whenever you signal them to do so.

To do the test blind, press the tape button several times quickly so you get confused and don't know which interconnect you are listening to. If your preamp has an indicator light showing what you are listening to, then either put a piece of black electrical tape over the light or close your eyes while you do the test.

After doing this test, you will discover that all the hype surrounding interconnects is just that. The fact is that all well designed interconnects sound identical.

But please carefully note that I said all ***well designed*** interconnects sound identical. Some interconnects are badly designed and do indeed sound different. So just what is a "well designed" interconnect?

First, the interconnect must be shielded. Shielding prevents RFI (Radio Frequency Interference) and EMI (ElectroMagnetic Interference) from corrupting the sound. RFI can take several forms with the simplest being a buzzing sound (usually caused from light dimmers), to actually hearing radio or TV program transmissions faintly in the background of your music.

EMI is caused by magnetic flux lines cutting across the interconnect and inducing currents in it. This can take the form of hum if the interconnect is near an electrical transformer or motor, or will be crosstalk if the interconnect is near another interconnect that is active with a different signal.

Shielding is usually done by braiding a fine wire mesh around an internal conductor(s), making the interconnect coaxial in design. Although this mesh is usually adequate, there are small spaces between the wires in the mesh so that there is not 100% coverage. To obtain the greatest shielding, some interconnects are designed with a solid foil shield. This foil is prone to cracking and breaking if it is flexed, so the foil (usually aluminium) is often deposited on Mylar film that is wrapped around the wire to improve flexibility. But still, foil-shielded cables should only be used in stationary applications since frequent

flexing will eventually crack the shield. Braided-mesh shielding should be used for interconnects in home audio systems.

The second requirement is that the interconnect have low impedance. High impedance can cause loss of output at both high and low frequencies depending on the loads presented by the components connected to the interconnect. And when the frequency response is restricted in this way, the effects are indeed audible. Buy why would you want to limit your system's frequency response?

The third requirement is that the connectors at the ends of the wire be practical and trouble free. This encompasses several factors:

- They must not oxidise or corrode as this eventually will cause a high impedance contact and restrict the frequency response. Most connectors are gold plated to meet this criterion, although solid nickel connectors also work well. If they are gold plated, it is important that the plating be of high quality so it doesn't easily chip or flake off.

- The outside contacts of an RCA connector should NOT be tapered. If they grip only on the their tips, they can put great pressure on your components' jacks and can gouge or scratch their gold plating. Along these lines, it is best to avoid connectors that have clamping mechanisms that you tighten after insertion. These can put enormous pressure on your components' jacks and then the slightest motion can tear off the gold plating. And it is virtually impossible to tighten them without moving them.

  The best contacts are those that have precision machined, parallel walls in the shape of a perfect cylinder. These produce smooth, even, firm pressure on the jack without damaging the gold plating.

- The connector should have a strain relief. The purpose of the strain relief is to prevent tension on the interconnect cable from being transferred to the delicate connections inside the connector. This requires some kind of clamping mechanism so that the connector is solidly anchored to the outer covering of the interconnect cable, while the wires inside the connector are slack.

  Most RCA connectors don't have any strain relief. Some have springs around the cable near the connector to prevent excessive cable bending, but it doesn't prevent tension from damaging the internal connections. Some RCA connectors have a small metal strap inside the connector that is pinched around the cable, but this is weak and grips very poorly. The best connectors will have a clamp that can

be screwed down and that gets a really solid grip on the wire's outer cover, but these are very rare.

- The connector should have a tough, scratch resistant, attractive exterior surface. Most are painted, and paint is easily damaged. Some are gold plated. But gold is soft and easily scratched. The best have industrial hard chrome surfaces. This type of coating has an attractive, matte silver colour and is very durable.

- The connector, particularly small RCA connectors, should have a "grippy" surface so that you can grasp it firmly to remove it so you don't have to pull on the cable. A thick ring is nice, but often interferes with other connectors that are close by. So the best option is deep knurling on the surface that produces a rough, easily gripped surface, without increasing the size of the connector.

- The connector needs to be highly conductive to keep the impedance low. Steel is not a suitable material. Brass or copper should be used.

Amazingly, many very expensive interconnects fail to meet these basic criteria. In particular, many have no shielding at all! This is inexcusable in an expensive interconnect. The manufacturers of such poor interconnects only get away with this because most home environments have little RFI and EMI. But this isn't always the case and there are many systems that are plagued with buzzing and other noises due to the lack of shielding. The owner is very frustrated that he can't get the noise out and never suspects that his exotic interconnects are the cause.

Some interconnects have very high impedance. This is because the interconnect uses extremely tiny wire. The manufacturers of such interconnects claim that very small wire prevents "transient smearing" due to "skin effect" or some other arcane reason. But the fact is, wire size and type does not affect the sound (unless the impedance is too high). There is no such thing as "transient smearing" in interconnects and "skin effect" does not alter the sound at audio frequencies. You discovered this in your listening tests. But some of these interconnects have several thousand ohms of impedance and can adversely effect the frequency response of your system.

Very few interconnects have connectors that meet the "practical and trouble free" criteria outlined above. There are too many connector types to discuss here, but if you will examine them, you will see that few meet the criteria outlined above.

Interconnect Design

At InnerSound, we make no extravagant or false statements about our interconnects. We don't claim that they sound better than any other well designed interconnect. What we DO claim is that they are the very finest quality and are superbly engineered. And we sell them at a reasonable price.

Specifically, the cable itself is coaxial, low impedance design, with a braided mesh shield. The mesh is of unusually high quality and has a very tight weave. The shielding is so well done that we use a transparent covering over it so that you can actually see the quality. The bright copper braid is also looks very elegant. The cable is quite flexible and a medium size (5.5mm). Although it doesn't really matter (as proven in your listening tests), the metal used in the wire is oxygen free copper.

Our RCA connectors have precision machined, parallel jaws, in the shape of a cylinder. They grip firmly and so perfectly that you can actually feel a suction and "pop" as you remove them from the jack. All contact surfaces are gold plated over brass. Insulation is Teflon. The jaws are protected by a strong outer cylinder that is separated from the actual contact jaws. This prevents any damaged to the precision contact jaws. The outer surface of the connector is deeply knurled for a good grip and hard chrome plated for superior wear resistance. It has a superb strain relief with tapered jaws that clamp down on the outer coating of the cable as a ring clamp is tightened at assembly.

We also supply balanced interconnects that use professional studio cable with a black covering. It is 7mm in diameter, heavily shielded, and flexible. XLR connectors have anodised aluminium housings and gold plated, brass connector pins. Excellent strain relief clamps are used.

Conclusion

Much as I would like to hope otherwise, I know that this will not make a great deal of difference to the believers of the "Cable Gods", who postulate that the use of unshielded pure silver interconnects make a difference (they do, because they pick up noise), or that exotic mains or speaker leads will change the character of their systems (they won't).

If you have vast amounts of money and want to impress your friends with your $5,000 speaker leads, then far be it from me to deny you this (dubious) pleasure. However, if you are like most of us, and don't have that sort of money to throw around frivolously, then don't for an instant think that you are missing out on musical "Nirvana", because it just isn't true. As I have suggested before, make your own leads, and use the money to

buy more music! This is infinitely more satisfying in the long run.

For the original article, see the original InnerSound white paper, and have a look at their home page too.

I strongly suggest that the disbelievers visit the ABX site, and look at some of the tests that have been performed on all manner of equipment.

Further information can also be found in some of my own articles, in particular Cables, Interconnects and Other Stuff - the Truth. Also see how you can make your own AB switch box, which can be used to test amplifiers, cables, capacitors and most other audio components.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Roger Sanders (InnerSound), and is Copyright © 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Roger Sanders) and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allow that one (1) copy may be made for reference. Commercial use of this published material is prohibited without express written authorisation from Roger Sanders and Rod Elliott.

Page created and copyright © 22 Jul 2001

The Audio Pages

Elliott Sound Products — Loudspeaker Cable Characteristic Impedance

Copyright © 2003 - Rod Elliott (ESP)
Page Published 17 Oct 2003

Articles Index

Main Index

Contents

- 1.0 Introduction
  - 1.1 Impedance Matching
- 2.0 About the Simulations
  - 2.1 Amplifier
  - 2.2 Cables
  - 2.3 Loudspeaker
- 3.0 Simulations
  - 3.1 "Special" Cable
  - 3.2 12 Gauge "Zip" Cable
  - 3.3 DIY Cross Connected Coax Cable
- 4.0 Further Investigations
  - 4.1 Aussie "Figure 8" Cable
  - 4.2 Transmission Line or Cable?
- 5.0 Frequency Response
  - 5.1 Amplifier Vs. "Ideal" Source
  - 5.2 Response Simulations
- Copyright

1.0 Introduction

What is the effect of the characteristic impedance of a loudspeaker cable? Is it important to match the cable to the speaker load impedance, or is this simply a marketing ploy?

Much has been said by many cable vendors about loudspeaker cables characteristic impedance, with claims that it should match the speaker impedance for "optimum results". Likewise, there is also a great deal that we are *not* told, and this is of much greater concern.

One thing the cable vendors have completely neglected to point out, is that the characteristic impedance is *only* important (and relevant) when the source impedance, cable impedance and load impedance are all matched. Having an extremely low impedance at one end (the amplifier) and a variable impedance at the other (the majority of all loudspeakers) makes true matching impossible.

Having said that, even with a very low impedance at one end of the cable, most cables can be made to have a passable match at low to mid radio frequencies (RF) by terminating at the far end only. (Note that at the frequencies we are looking at they are not really transmission lines, although I may use the two terms interchangeably.) Some degree of "matching" may be (to a degree) because the amplifier's output impedance rises as the frequency increases, but all amplifiers will be different in this respect.

To improve the impedance match and reduce the reflections that are caused by an unterminated transmission line, some vendors recommend a Zobel network - a resistor and capacitor in series, typically 10 ohms and 100nF. This, they tell you, should be installed at the speaker end of the cable. The usefulness of this may be dubious as will be shown later in this article, but in most sensible cable constructions, it will do no harm. For some of the more exotic constructions, a far-end Zobel is essential, however most are incorrectly designed.

Another "minor" detail that the cable vendors fail to mention is that the characteristic impedance of a cable varies with frequency. At DC, the characteristic impedance of all cables is infinite (for all intents and purposes), and the rated impedance is usually not reached until the signal frequency is well above the audio band - typically around 100kHz or more, depending on the cable's construction and length. Any cable consisting of parallel or concentric conductors (including flat conductors, multi-wire ribbon cables and Litz cables) acts as a transmission line at high frequencies, or (at least to a degree) if the line is extremely long (usually several to many kilometres). All transmission lines have a characteristic impedance, and this is a basic principle of physics - at issue is the

flippant way many vendors handle the truth.

But wait! What *is* characteristic impedance anyway? The characteristic impedance of a cable (Zo) is a complex function of the diameter (or dimensions if other than round) of the conductors, their relative spacing, and the insulation material. Simplified, it is determined (for high frequencies) as Zo = $\sqrt{}$L/C, where Zo is characteristic impedance, L is inductance and C is capacitance.

Note that Zo is a constant, and is independent of the length of the cable. An ideal cable (for a high powered audio system) will have low inductance, low capacitance and low DC resistance (DCR), but it is important to understand that the Zo of the cable is completely unimportant at audio frequencies.

In order to obtain a low characteristic impedance, it is necessary to have very low inductance and relatively high capacitance, and the high capacitance may impose serious constraints on the amplifier. Indeed, many amplifiers will become unstable if there is sufficient capacitance connected directly to the output, causing oscillation which may damage the amplifier. As described above, regardless of anything else, the cable does *not* act as a true transmission line at audio frequencies, and claims to the contrary are fallacious.

For these esoteric cables, their high capacitance dictates that there will be the opportunity for the insulating material to contribute its "sound" to the overall signal fed to the loudspeakers. It must be pointed out that this is hotly disputed by many engineers, and there is no conclusive evidence that any one dielectric material is "sonically superior" to any other. I know of no properly conducted Double-Blind Test (DBT) where the listening panel was able to pick the difference with greater than 50% accuracy - i.e. pure chance. However, the possibility cannot be discounted, so it is worth mentioning.

1.1 Impedance Matching

For RF applications, or for extremely long signal runs, such as telephone circuits, impedance matching is essential. Matched impedances mean that the source, cable and load impedances are all the same. For video, the standard impedance is 75 ohms, so the output impedance of a video line driver will be 75 ohms, 75 ohm coaxial cable is used for connections, and the receiving end is also 75 ohms. There are nearly always mismatches because of the use of RCA connectors (typically about 40 ohms impedance), but in a domestic installation this usually does not cause a problem, since the cable runs are relatively short. It is generally accepted that if the transmission line

length is less than around 1/10th of the shortest wavelength, then impedance matching is not overly important - again, we will see that this is not necessarily the case, since an audio amplifier's bandwidth can easily exceed the audio bandwidth by a factor of 10 or more - the effects may not be directly audible, but amplifier oscillation is to be avoided under any circumstances.

Matched impedances ensure maximum power transfer from source to load, and this is obviously very important for RF transmitters and telephony applications. It is completely irrelevant for a solid state audio power amplifier however, since the drive principle (known as voltage drive, or constant voltage) does not rely on maximum power transfer, but relies instead on the amplifier maintaining a low output impedance with respect to the load.

The ratio of amplifier output impedance to load impedance is called "damping factor", and with modern amplifiers it can easily exceed all normal (real life) requirements. Power amplifiers usually have an output impedance of between 10 and 100 milliohms, giving damping factors of between 800 and 80 (respectively). Valve amplifiers may have a damping factor as low as unity (i.e. the amplifier's output impedance is equal to the load impedance). No amplifier manufacturer quotes damping factor with cables attached, and in reality it is always less than claimed.

The damping factor (DF) figures are theoretical, and are rarely (if ever) achieved in practice. For the remainder of this article, I will use the output impedance of my simulated amplifier for reference (about 24 milliohms). This figure is passably realistic in a real world amplifier, but the internal wiring will increase it somewhat.

Even though most power amplifiers are limited to at most a few hundred kHz or so, there can still be some energy at higher frequencies - typically noise. What often happens is that an amp can be quite stable with a capacitive load and no signal, but as soon as it is driven it "excites" the whole system, and it then bursts into sustained oscillation.

It is almost impossible for any amp to reproduce high levels at extremely high frequencies, and they are not present in the source material. This has never stopped an amp from oscillating though, usually at a frequency high enough to cause simultaneous conduction of the power transistors, since they cannot switch off quickly enough, and both will be turned on at the same time.

This simultaneous conduction is what causes damage, since the output devices heat up very quickly and may go into second breakdown - if that happens, then it's all over - the amplifier will fail with blown output devices. Anyone who has had an amplifier on a test

bench and supplied it with an input signal at 100kHz or more will have seen this - even with no load, the amp will draw a lot of current even at low output levels. If maintained for any period of time, the amp will fail.

2.0 About the Simulations

The details of any test or simulated test are imperative for a full understanding. With this in mind, the following section describes the simulated amplifier that was used, the unloaded frequency and phase response, and the simulated cable and loudspeaker details.

2.1 Amplifier

The simulated amplifier schematic is shown in Figure 1. The phase angle at unity gain - at round 10MHz - is 150° (or a phase margin of 30° - i.e. a very stable amplifier). Phase is relatively unaffected by load resistance, or the presence of the amplifier's internal Zobel network. The latter is to ensure stability with an inductive load, but is ineffectual against capacitive loading.

To verify that the simulation represents reality to an acceptable degree, a 100nF capacitor placed across the output will cause instability, and this is substantially in agreement with empirical data on real amplifiers. It is worth noting that the simulated transistors have a much wider bandwidth than the majority of real life transistors, and this shifts the results up in frequency. Most of the amplifier effects seen at 10MHz will actually occur at perhaps 1MHz. All transmission line effects (in particular the peaks caused by impedance mismatches are exactly as shown - these are determined by the cable and its length, and are independent of the amplifier. Also note that many amplifiers will oscillate with a great deal less capacitance than the simulated version.

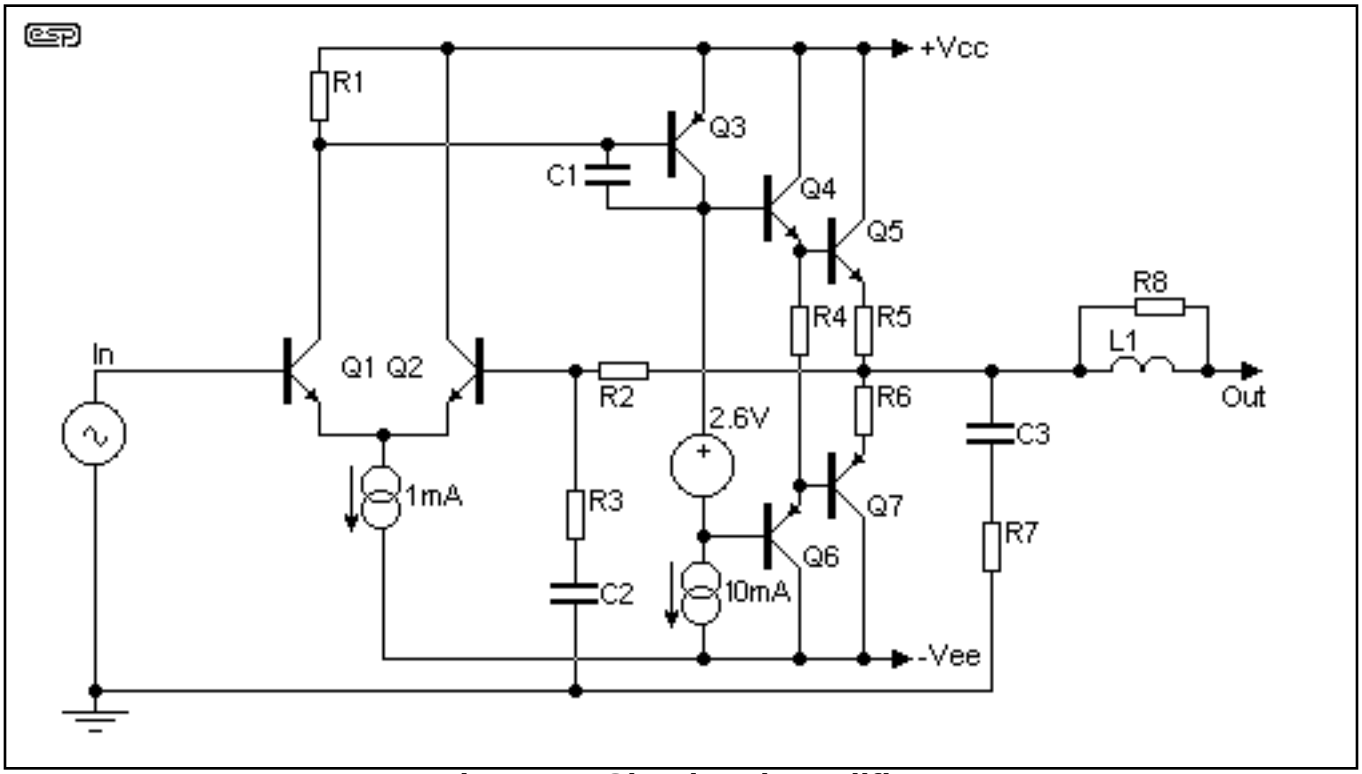


Figure 1 - Simulated amplifier

The amplifier used for these simulations was as shown in Figure 1 and it is a completely conventional (if somewhat simplified) circuit, typical of those used for hi-fi applications (in fact, fairly typical of the majority of amplifiers for any purpose). The closed loop frequency and phase response are shown in Figure 2. The phase margin (the number of degrees of phase shift between the actual unity gain frequency (Ft, or transition frequency) and 180 degrees. At 180 degrees phase shift, an amplifier's negative feedback is reversed in phase, so becomes positive feedback. If the amplifier has gain greater than unity at that frequency, it will oscillate.

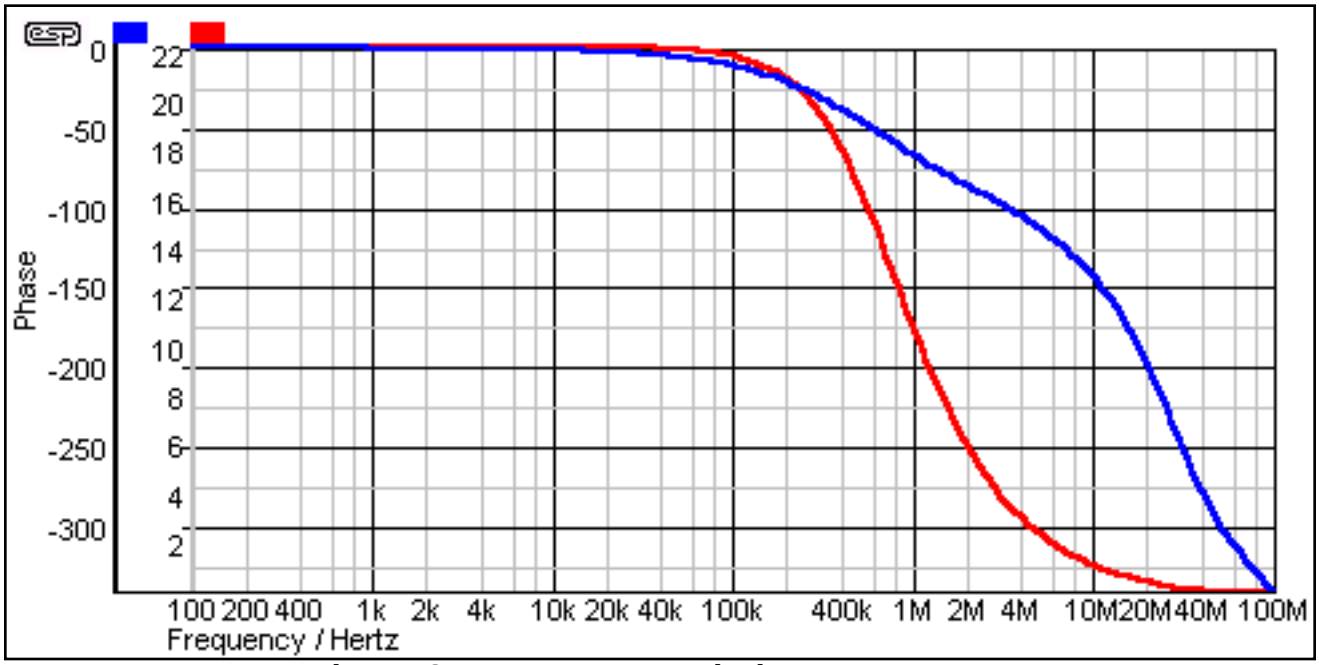


Figure 2 - Frequency and Phase Response

Neither frequency response nor phase response are affected by a resistive load. Inductive loads (of less than a few hundred uH - micro Henrys) are compensated by the Zobel network C3 and R7, with typical values being 100nF and 10 ohms, although this varies to some degree with the particular design. For reference, the simulated amp's output impedance at selected frequencies is shown below.

Frequency	1 kHz	20 kHz	100 kHz	1 MHz	2 MHz
Impedance	24 mΩ	25 mΩ	45 mΩ	444 mΩ	515 m&Omega

Table 1 - Output Impedance Vs. Frequency

In many amplifier designs, the speaker line capacitance is "decoupled" from the amplifier by the combination of L1 and R8. Again, fairly typical values are 800nH in parallel with 10 ohms, but as always, this can vary. Inductance above 1 or 2uH is rare, as it will have an audible effect on the overall frequency response, especially with low impedance speakers (for example, a 10uH inductance causes a 0.4dB loss at 20kHz and a 4Ω load, and that's with no cable at all). The resistor damps the Q of the inductor to prevent (or at least minimise) the possibility of the inductor and cable forming a resonant circuit. This network can be replaced with an 0.1 ohm resistor, but that approach is rare. Simulations were performed without the series inductor, since its presence swamps the very effects we are looking for so that amplifier instability can occur. The amplifier's RC Zobel network was retained for all simulations - very few amplifiers will be stable with any normal speaker and load without this network.

Note that in some of the simulations that follow, the phase margin may *appear* adequate to maintain a stable system. This is a direct result of simulations and real life failing to coincide, and it is imperative that you understand that the results shown may appear to "trivialise" the effects. Cables such as Sample #3 *will* cause most amplifiers to become unstable - the phase results shown also appear to be static, but in fact they change with amplitude, and anything that creates a radical phase shift at the amp's output will almost certainly cause oscillation in real life.

2.2 Cable

Figure 3 shows the schematic of the simulated cable, and the "far-end" Zobel network. Also shown is a "near-end" (i.e. at the amplifier) Zobel, but this is not needed for the most part. It is recommended if you use a series inductor at the amplifier output (or if you know that one has been included by the manufacturer). Leaving it out of circuit will normally have little or no effect on amplifier stability, and only serves to terminate the cable properly at very high frequencies.

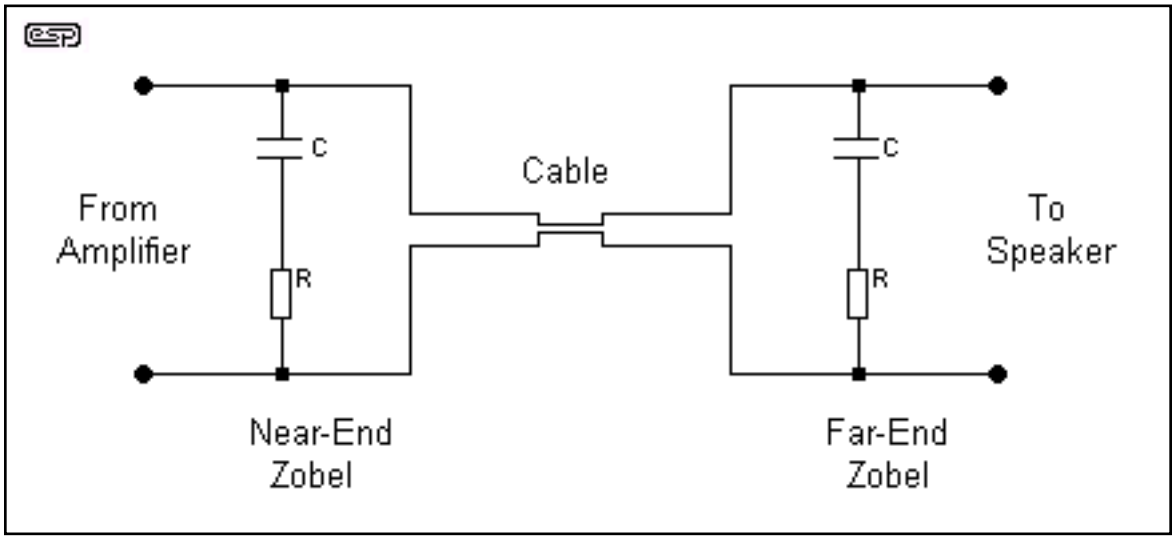


Figure 3 - Simulated Cable & Termination

Although most people (myself included) will not like the idea of a ceramic capacitor in the audio path (for the C element of the zobels), ceramic is actually the best choice for this application. It is important that the inductance of the capacitor (primarily lead and capacitor body length) is as low as possible, or response will be limited at exactly the frequencies where it becomes important. Even if the capacitor is non-linear (which is a characteristic of ceramics), it is in parallel with the load and amplifier output, and will have no audible effect on the signal. Remember that we are dealing with a small capacitance (100nF is typical), and its reactance (or impedance) is 100Ω at about

16kHz. A non-linear device with that much impedance will not affect any known amplifier, and the cap is also in series with a resistor, further reducing its already negligible contribution.

The use of "audiophile grade" polypropylene or other film caps is strongly discouraged, since their performance at several MHz is degraded by internal inductance (which determines the 'self resonant frequency of the cap). The capacitor(s) should be rated at a minimum of 50V AC (or 100V DC), although the voltages actually developed across the caps should be much lower than this at any frequency.

The resistor should be a carbon or metal film type, and a rating of 0.5 - 1W is normally quite sufficient. The power developed with normal signal will be a lot less than 0.5W, even with powerful amplifiers. If you feel that you *must* use a wirewound resistor, then it must be a non-inductive type.

2.3 Loudspeaker

The loudspeaker is a two-way, fully impedance compensated design. It uses a Zobel network to null the woofer's rising impedance caused by voicecoil inductance, and a tweeter resonance compensation circuit. This is based on a simulated loudspeaker system used by Jon Risch (the designer of Cable Sample #2) for some of his measurements and simulations, and is a reasonable approximation of a real speaker system. A resistance (or a much simpler simulated loudspeaker) could have been used, but this would not provide a "real life" experience for the simulations.

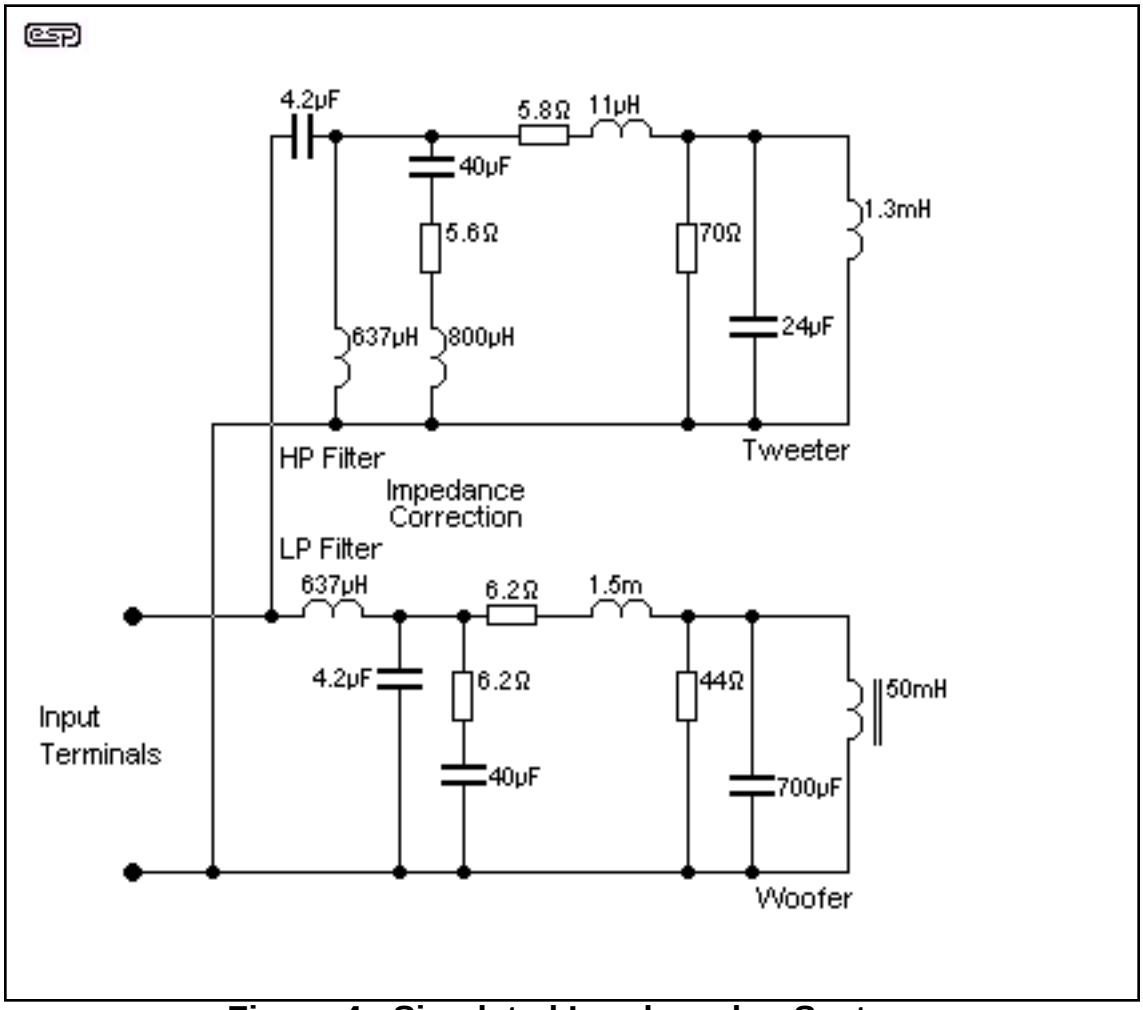


Figure 4 - Simulated Loudspeaker System

3.0 Simulations

Three different cables were used in the simulations that follow. It is important to note that these are simply used as representative, and no conclusions should be formed when comparing to manufacturer's stated (or claimed) data. No endorsement is implied for any configuration, manufacturer or anything else, other than a final recommendation below.

Sample	Resistance	Inductance	Capacitance	Impedance	Reference
1	11.15 mΩ	626.64 nH	68.90 pF	95.37 Ω	12# Zip
2	14.57 mΩ	219.82 nH	114.83 pF	43.75 Ω	Jon Risch
3	14.44 mΩ	32.81 nH	1640.42 pF	4.47 Ω	Goertz MI 1

Table 2 - Simulated Cable Parameters (per metre)

All values are per metre of cable length, and the simulations were performed using a 4

metre length (a little over 13 feet), as this is fairly typical of most domestic installations. As cables are made longer, the effects described occur at lower frequencies.

In the following sections, a great many possibilities are looked at, and graphs of the response are provided. By necessity, these are smaller than optimum, otherwise the page would take 3 weeks to load. The effects of different combinations are very clear, and additional descriptions point out the areas of interest.

3.1 "Special" Cable

The first graph shows Cable 3. I selected this for the first simulation, as it is the most likely to cause amplifier instability with no termination Zobel. Although the radical changes in phase are quite visible, it is a little difficult to see the effect on the amplifier. It transpires that this cable produces an output phase on the amplifier of that is well outside its phase margin, at around 192°.

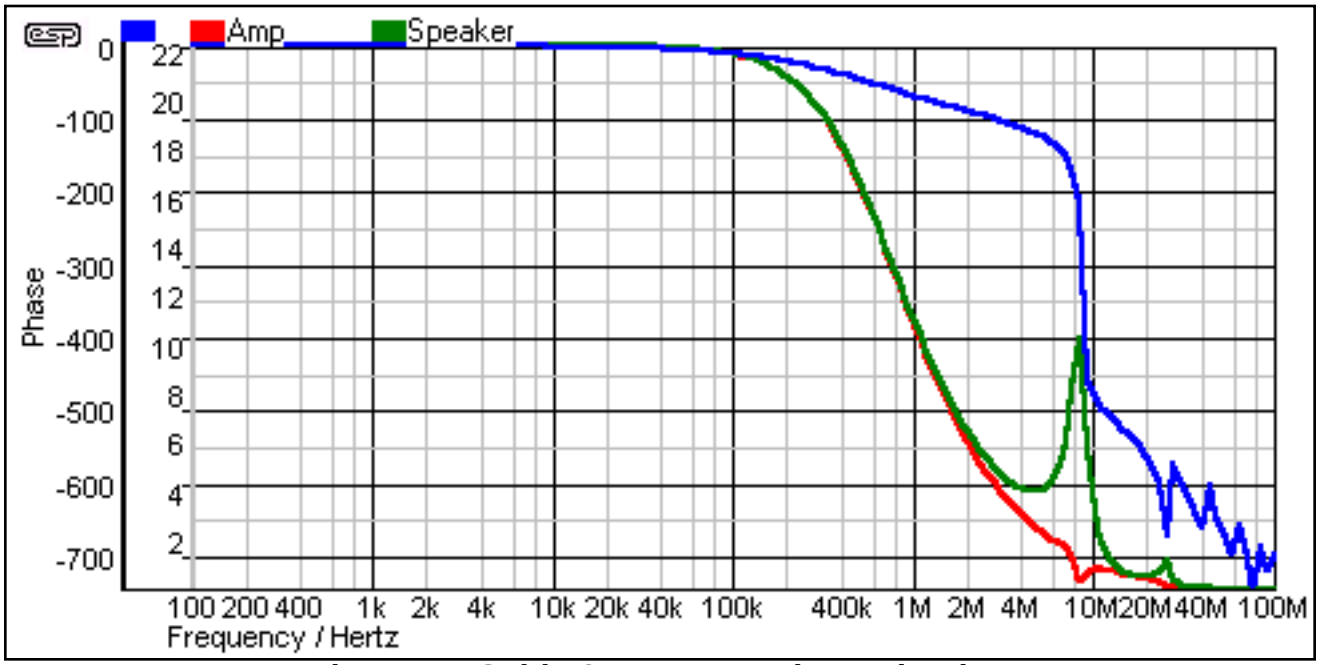


Figure 5 - Cable 3, No Far End Termination

As you can see, the cable has a large peak in the response at a little under 10MHz, and the phase response is savage. Each kink or discontinuity in the plot indicates a reflection, and note the phase angle - it shows 700° of phase shift at 100MHz!

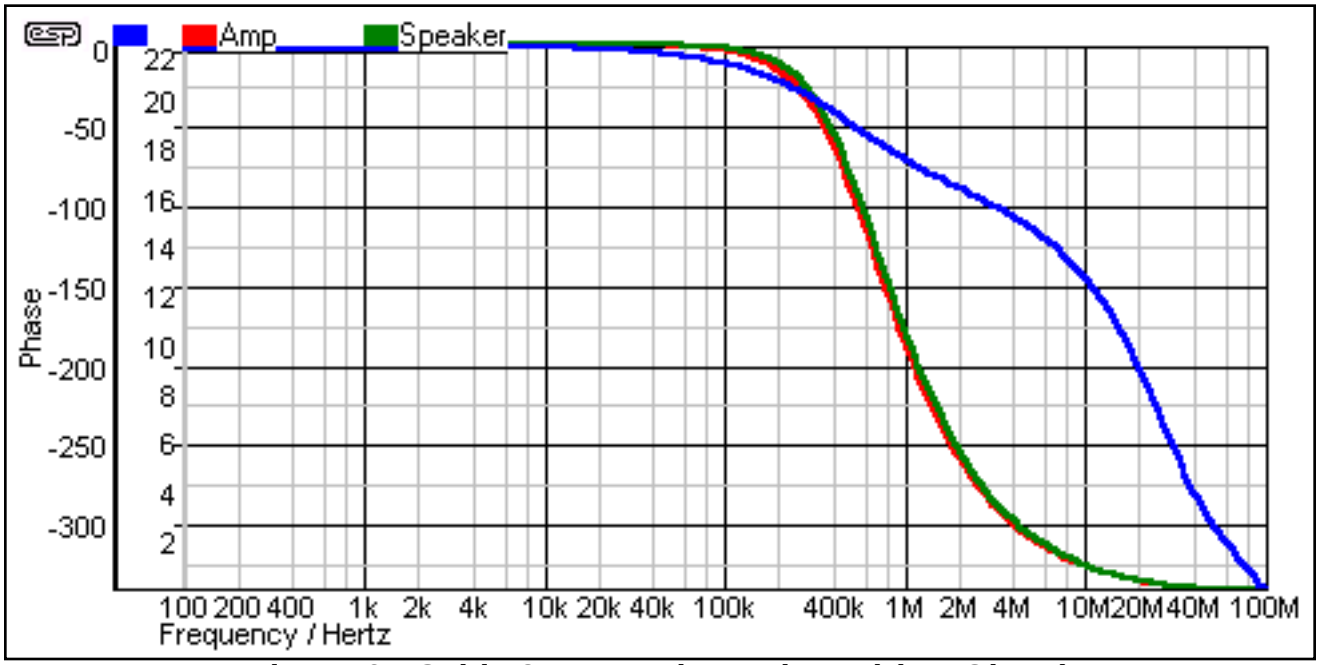


Figure 6 - Cable 3, Far End Terminated (4.7 Ohms)

Adding the far end Zobel network, as you can see here the cable's response decay is perfectly smooth with a 4.7 ohm resistor. This is the optimum match, and is the value that *should* be used - *not*10 ohms as supplied (see below).

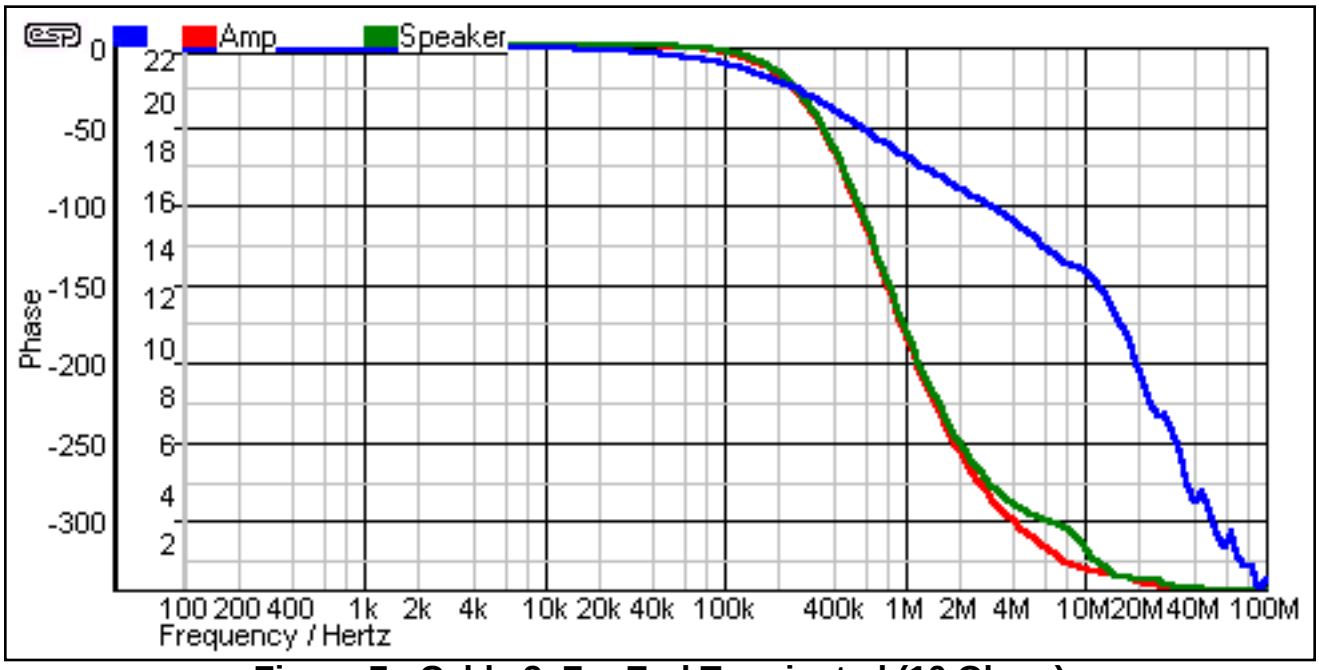


Figure 7 - Cable 3, Far End Terminated (10 Ohms)

Even 100nF in series with 10 ohms restores the amplifier phase margin to normal (150°). As seen above, 4.7 ohms is preferable, but the phase margin is barely affected.

The speaker end response has a small "lump" with 10 ohms, and phase goes "wobbly" at above 20MHz. This is probably not a concern, and you will almost certainly get away

with it. It is very evident that this particular cable should *never* be used without a Zobel at the speaker end, and it is equally obvious that the vendor does not really understand transmission line theory, since the zobels supplied with the cable (and you have to ask!) are incorrect. This is not difficult to get right, and if they are off base with something a simple as a resistor value, I would be disinclined to believe their other material.

While the results aren't actually appalling, it is quite obvious that the performance is not as good as with the correct termination resistance. At this level (and since the impedance of the cable is quoted on the web site), I find it difficult to understand how they could have made such an error. Not that this manufacturer is alone by any means - "experts" will emerge from the woodwork, suggesting a that 10 ohm, 100nF Zobel is the panacea - it is, but *only* for 10 ohm cables!

3.2 12 Gauge "Zip" Cable

This seems to be the standard against which all other cables are judged, so it is next on the list. As you can see, there is a pronounced reflection at almost exactly the same frequency as before. This must be, since the cable (transmission line) is the same length, and the first reflection will occur at the same frequency. Small variations do occur, and are the result of differing velocity factors. Velocity factor refers to the speed at which an electrical signal passes through a cable. Typically, this is between 0.6 and 0.8 of the speed of light (3 x 10^8 metres per second.)

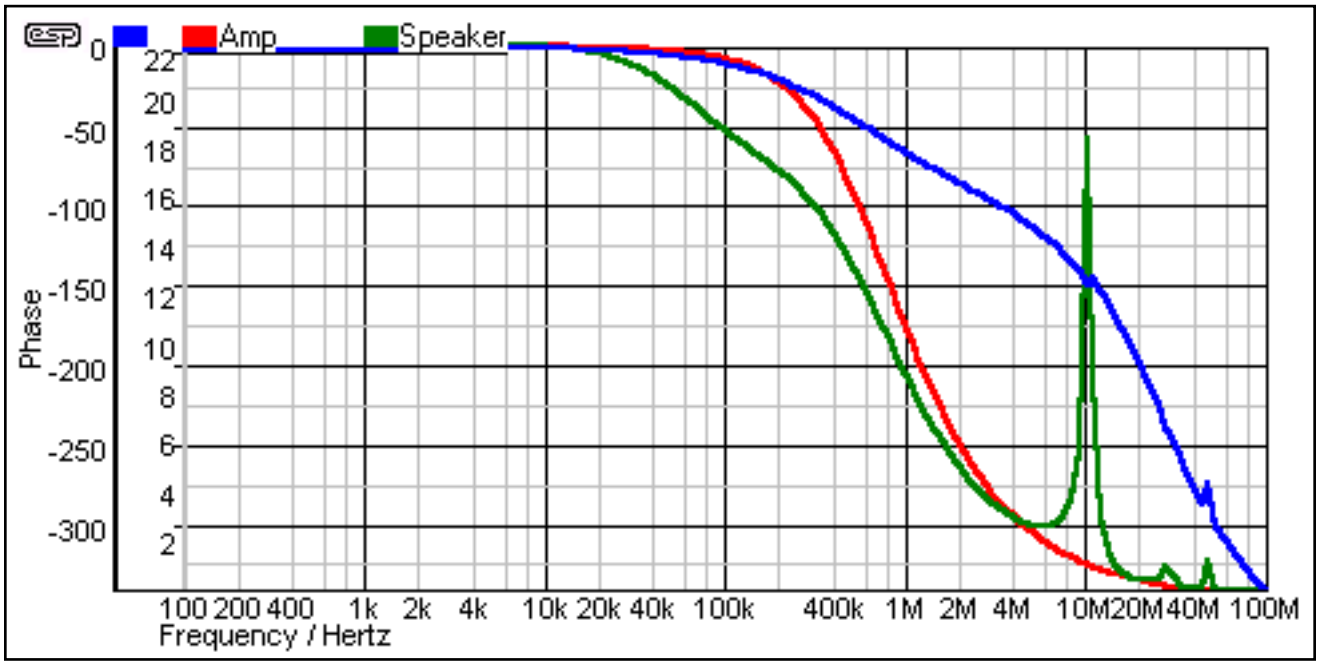


Figure 8 - Cable 1, No Termination

The response is relatively benign, despite the quite large peak at 10MHz. The reflection

spike causes no change to the amplifier's phase response until it is above 40MHz and will have no effect. There is a very slight reduction of level at 100kHz (and down to 20kHz), but this is measured in fractions of a dB, so can safely be ignored. A greater disturbance to the in-room response will be experienced by moving the listening chair or a nearby coffee table.

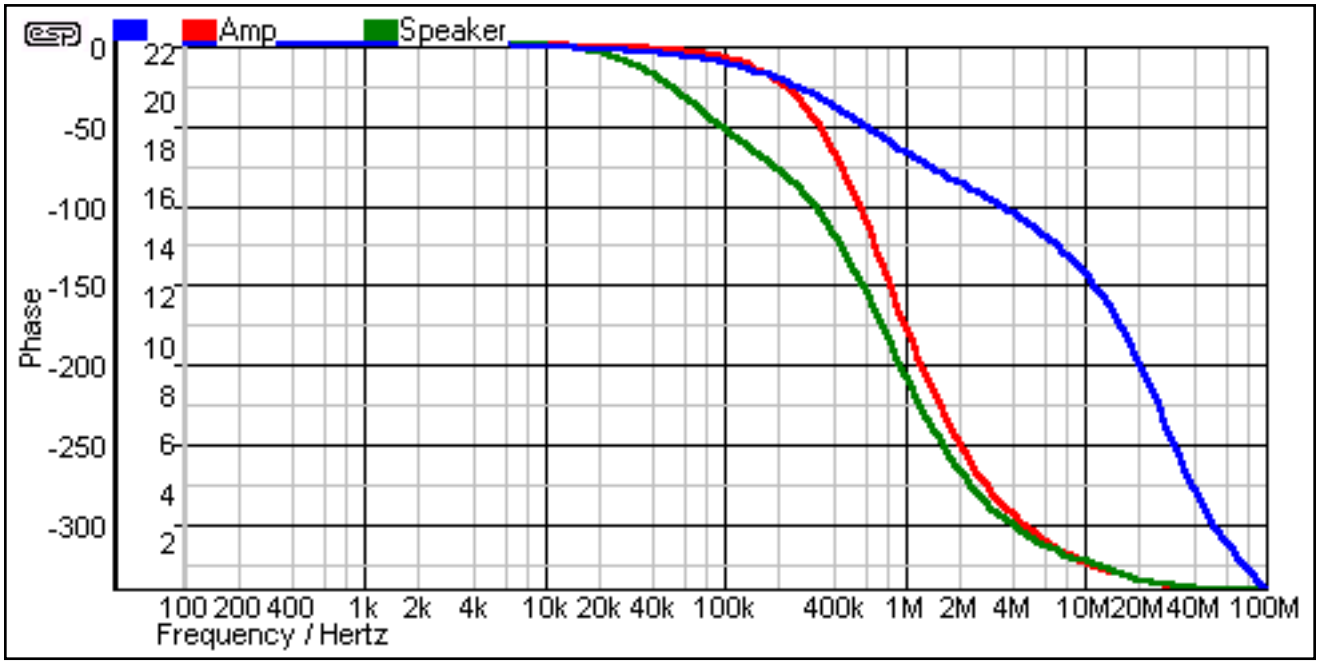


Figure 9 - Cable 1, Far End Terminated (100 ohms)

This is a perfect result. There are no spikes, no response anomalies, and the amp's phase margin is unchanged. For the cost of a 100 ohm resistor (close enough) and a 100nF ceramic capacitor, the cable is nicely terminated, and although virtually any amplifier will drive this cable with no ill effects even when unterminated, there is the potential to reduce RF pickup.

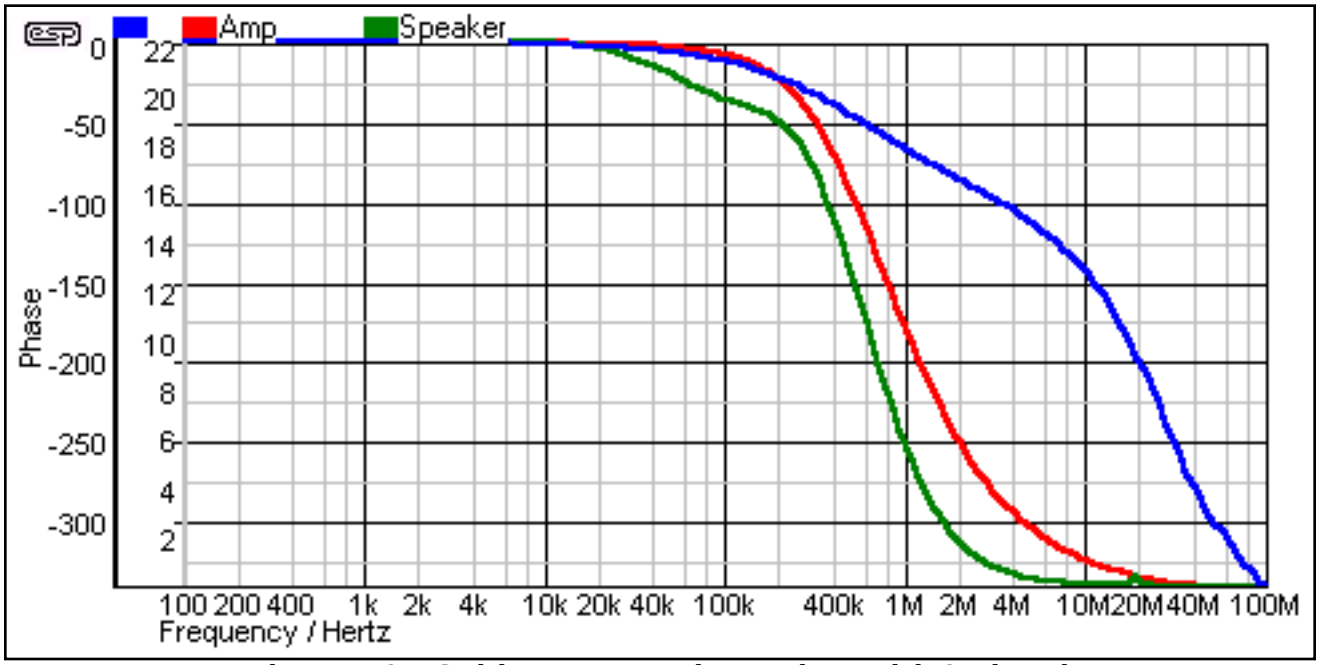


Figure 10 - Cable 1, Far End Terminated (10 ohms)

It is apparent that a 10 ohm termination impedance is "sub-optimal". In reality, it is not that bad, and the effects will almost certainly be inaudible. Something that is not obvious is a peak at around 10kHz - it is not large (about 0.1dB relative to 1kHz), and the top-end response is down by about 0.2dB at 23kHz (note that the effect will be a little worse with a 4Ω speaker). As a compromise, a 47 ohm resistor will cause no major peaking, and presents a passable match to a wide range of "zip" cables. Because these cables are already benign, they are much more tolerant of mismatch.

3.3 DIY Cross Connected Coax Cable

It was actually a reference to this particular cable on the Audioholics website that caused this investigation in the first place. Comments were posted and re-posted, as were claims and counter-claims. This design is also relatively benign, and I do not propose to cover all the issues raised previously - I am interested only on the cable's effect on the amplifier, and whether it is likely to cause instability.

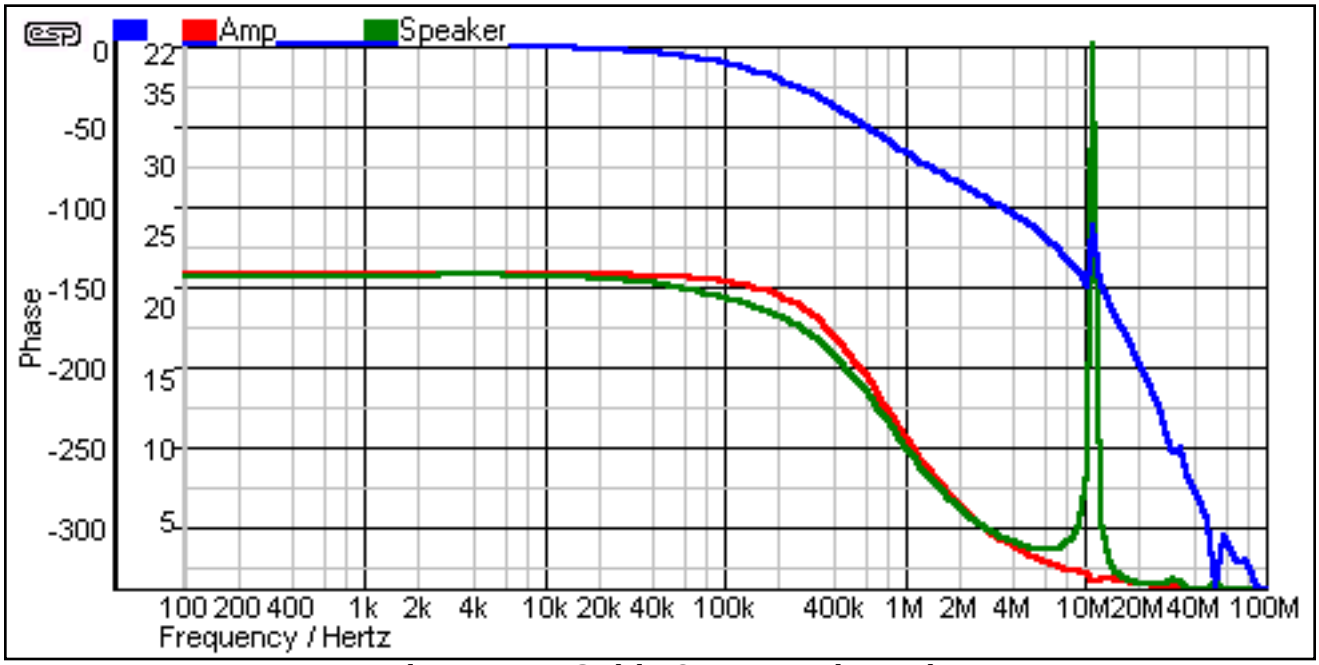


Figure 11 - Cable 2, Unterminated

As can be seen in Fig. 10, even when unterminated, the result *looks* pretty severe (note the height of the spike at 10MHz), but is not very much worse than cable #1. There is a phase anomaly at 10MHz, but the amplifier is well within its phase margin, and few (if any) amplifiers will be affected by this. Not as visible on the chart, but present nonetheless, is a slight broad peak in response centred on 5kHz. The amplitude is well under 0.1dB (0.04dB to be precise) and is insignificant. Amplifier phase margin is not affected, but cable response extends to over 100kHz - I can't hear it and nor can anyone else, but this also means that there is less rolloff with a lower impedance speaker.

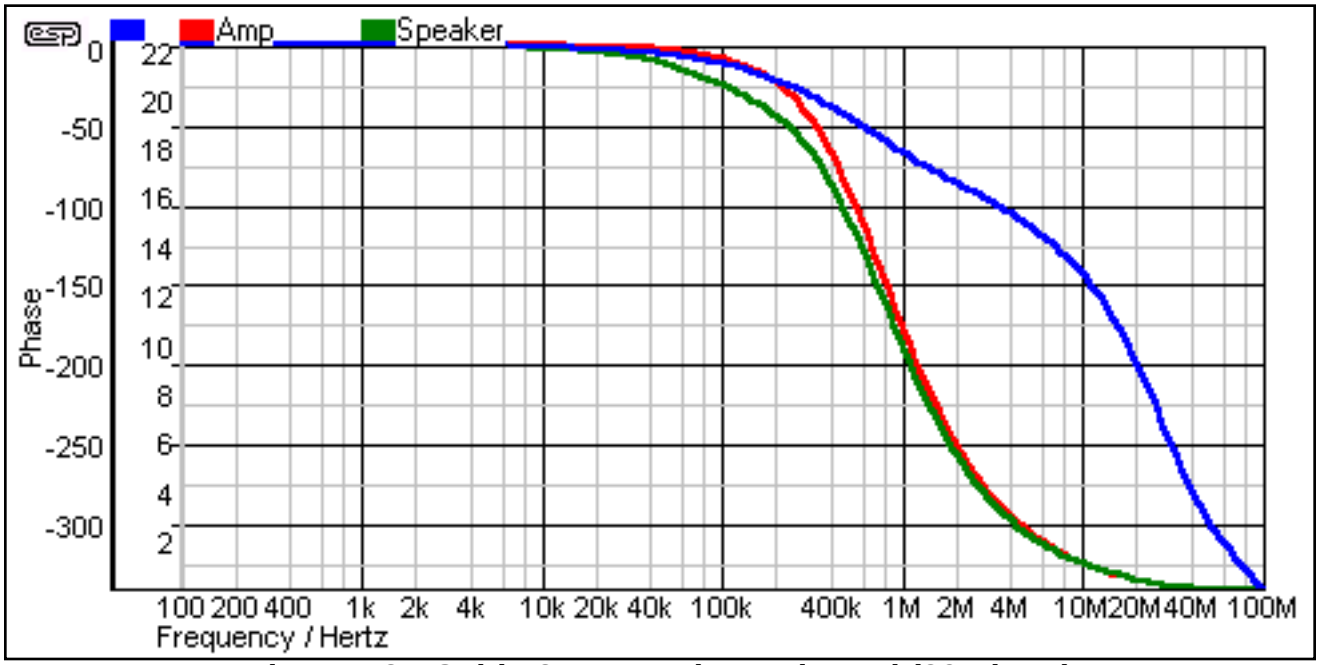


Figure 12 - Cable 2, Far End Terminated (39 ohms)

Since this cable has a Zo of 37 ohms, a 39 ohm termination resistance is close to optimum. The response again is virtually flawless, and reflections are eliminated completely.

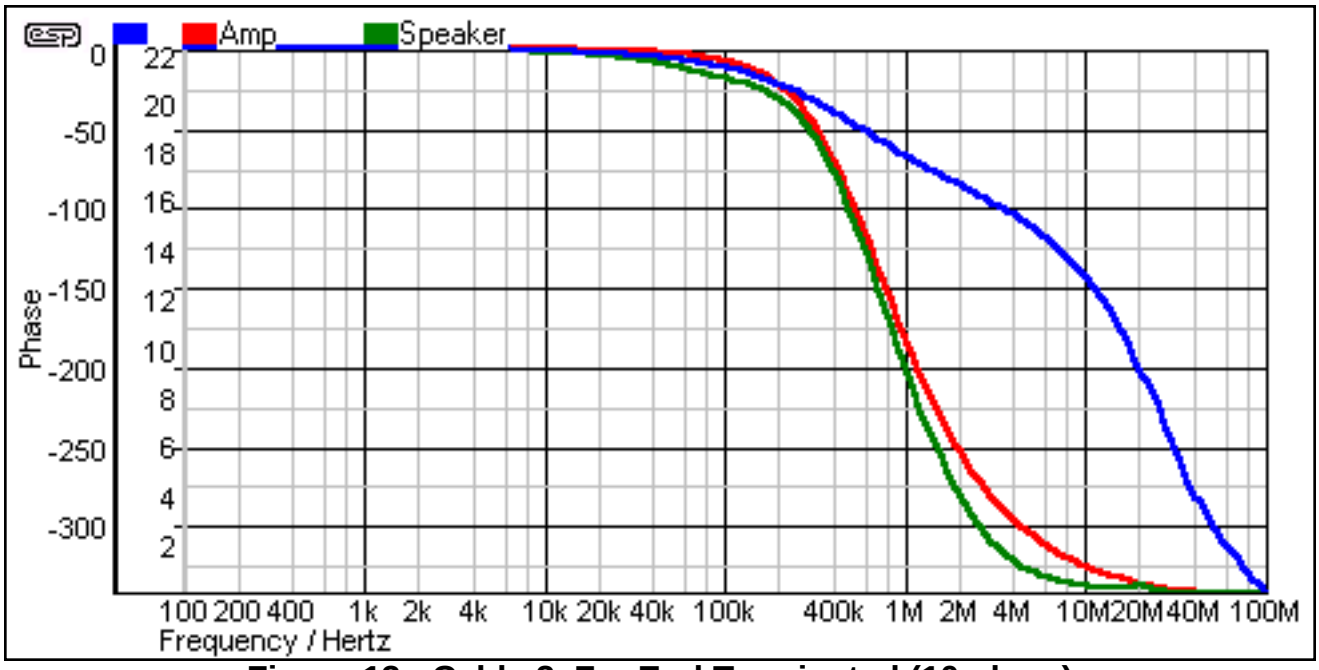


Figure 13 - Cable 2, Far End Terminated (10 ohms)

Overall, this cannot be considered a bad result. In fact, the mismatch actually improves the response, although it is well outside the audio spectrum. That it is less than optimum is not at all obvious (although there are some tiny ripples on the phase response at above 20MHz), but amplifier stability is not compromised, and there is no peaking. Also obvious is the divergence of the amplifier and loudspeaker-end responses (also visible on the other graphs where impedances were mismatched), and this is a clear indicator of a mismatch - even if it is well outside the audio bandwidth.

My personal choice would be to use a 39 ohm resistor, since that is a much closer match to the line impedance, but the resulting differences will certainly not be audible.

4.0 Further Investigations

Using the decoupling RL network at the output of the amplifier will provide protection against all but the most radical of speaker cable designs, but it must be noted that it is a rather pointless exercise to spend a great deal of money to get cables with the lowest possible inductance, and then have to add an inductor so your amplifier doesn't oscillate

The obvious (and recommended) cable constructions are samples 1 and 2, and both of

these will give good performance for a very reasonable outlay. Cable sample #3 cannot be recommended, not only because of its high capacitance, but because it is much more expensive than the other two, and the actual benefits are minimal. If you can afford to purchase such cables, then that is up to you, but bear in mind that the manufacturer cannot even manage the correct value for the Zobel network that *they* supply. In my books, that does not qualify them for anything more than a shake of the head, a mild chuckle and "Not today, thanks".

4.1 Aussie "Figure 8" Cable

Since I had the simulations all set up, I figured that a quick test of what is commonly known in Australia as "Figure 8" cable (basically the same as zip cable) was in order. The standard "hardware store" offering is not especially robust, being 22 strands of 0.022mm wire per conductor (0.75mm^2). It is 240V insulated wire, commonly used for lamps and the like ("lamp cable"). I use this for basic test leads in my workshop, and the figures for a 4 metre length are ...

Resistance = 150 mΩ
Inductance = 3.9 uH
Capacitance = 217 pF

Impedance works out to about 134 Ω. This is a very benign cable, and I have never seen an amplifier oscillate because of it, so, what do the simulations say?

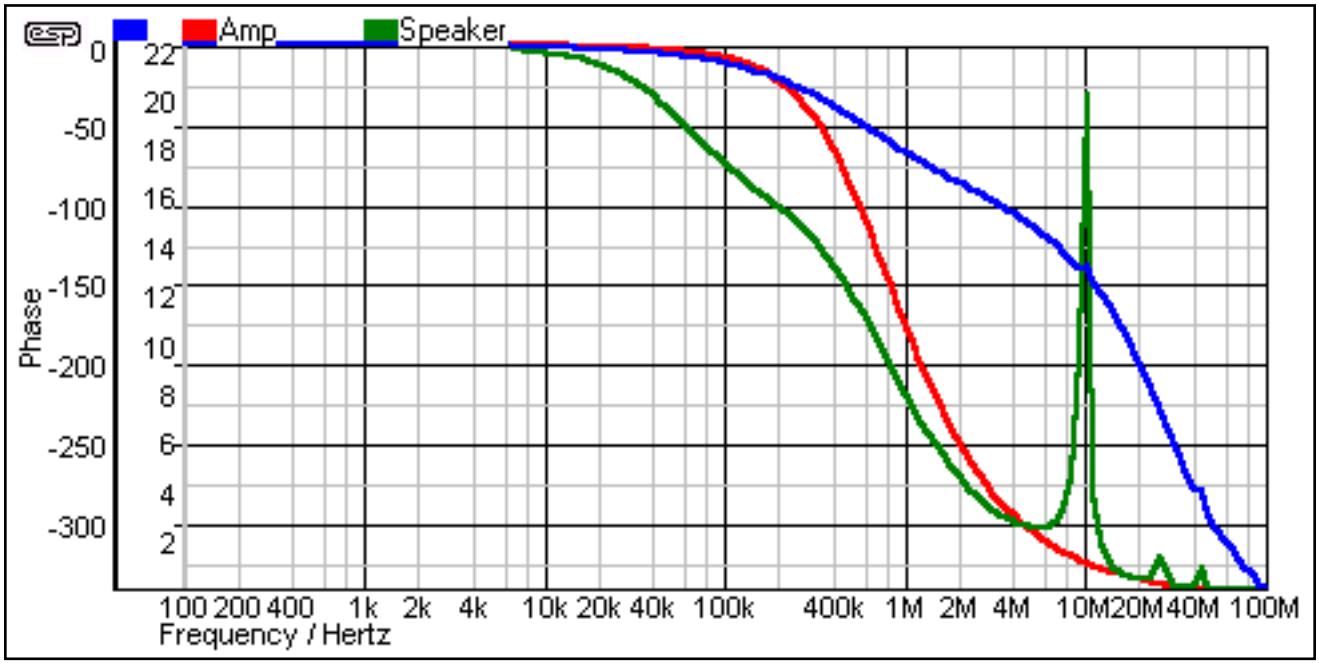


Figure 14 - "Figure 8", No Termination

No surprises at all really. There is an obvious loss at 20kHz (still less than 0.5dB though), the spike that we now expect, and zero phase anomalies.

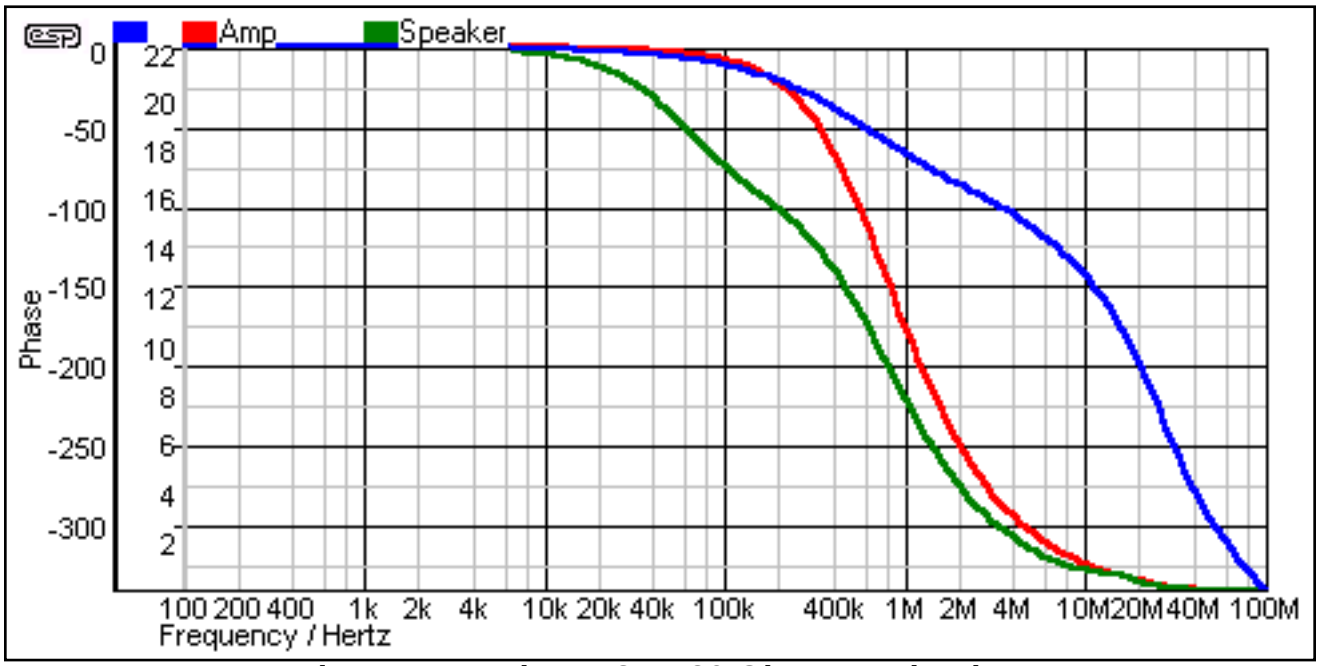


Figure 15 - "Figure 8", 120 Ohm Termination

Again, this is to be expected. Response is -3dB at 67kHz at the speaker end, which actually will do me just fine, since I certainly can't hear the loss (less than 1dB at 20kHz). The termination again removes the spike in the response, and there are zero phase effects at the amplifier. It goes without saying (but I will anyway ;-) that termination causes zero audible difference, however, in the presence of strong RF fields, a terminator *should* reduce or even eliminate RF interference. This is difficult to test because my workshop is nowhere near any RF sources powerful enough to cause problems, but a friend is going to run a test for me when he gets the opportunity, and I will add the results when known.

4.2 Transmission Line or Cable?

It is worth noting that any of these cables will exhibit further anomalies and discontinuities if driven from an amplifier with infinite bandwidth. These have not been invented yet, so the amplifier-end mismatch is not a major problem, but if an inductance isolation network is used in an amplifier, then a second Zobel following the inductor is recommended. This ensures that the transmission line is terminated with the correct impedance at both ends.

In addition, it was stated at the beginning of this article that characteristic impedance is

irrelevant at audio frequencies. The frequency at which a pair of wires line starts to act like a transmission line is determined primarily by its length. Fortunately for the purposes of research, simulators can provide a signal source with zero impedance and infinite bandwidth, so the limitations of the physical world need not concern us. Using a 50 ohm transmission line (purely for convenience), the following points are of interest ...

- When driven and terminated by 50 ohms, a line with (essentially) zero DCR has a flat response from DC to daylight. It is true transmission line, and behaves exactly as one would expect, regardless of length.
- A line with "typical" DC resistance is dominated by the DCR up to a frequency that is dependent upon its length. A transmission line is not influenced by DC resistance, so as long as the DCR is dominant, the line is *not* a true transmission line. This is hard to comprehend, but it is so, and any reference book on the topic will verify this "strange but true" fact.
- At audio frequencies, speaker cables are *not* transmission lines. They are merely cables, with inductance, capacitance and resistance. Despite popular belief, they are bereft of any magical properties, only physics.

Now, let's look at each of these claims. If a cable acts like a transmission line, is properly matched at each end, and has essentially zero resistance, then 1V from the generator will result in 0.5V at the input of the cable, and 0.5V at the output, since the system behaves like a simple resistive voltage divider. The cable is irrelevant. I will spare you the tedium of looking at a graph with a straight line at exactly 0.5V from 1Hz to 100MHz, since it has zero interest value.

In order to highlight the resistive effects, I simulated a 100 metre cable, with 1 ohm per metre DCR. This is obviously not a useful cable (at least not as a speaker cable), but it shows the effect very clearly. Figure 14 shows the test circuit, and current was monitored at the output of the AC generator.

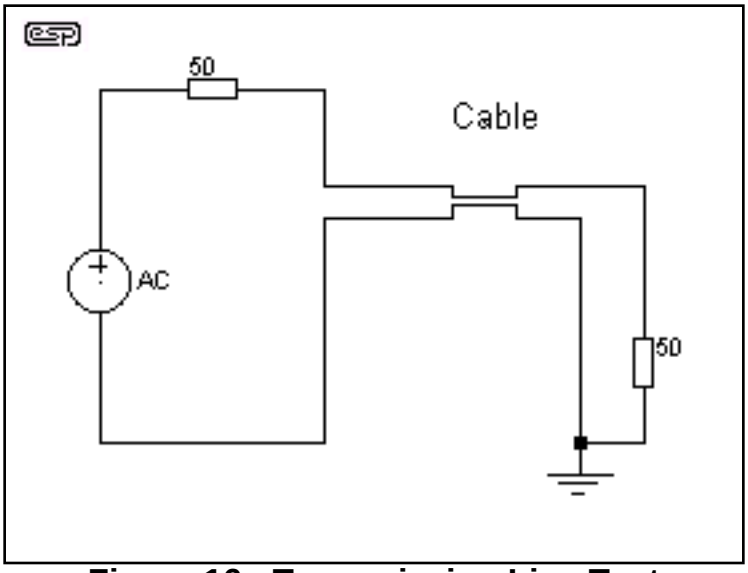


Figure 16 - Transmission Line Test

In Fig. 17, you can see that the output current is clearly dominated by the resistance until the cable behaves like a transmission line. The current from the generator (AC) should be equal to V/R ... V is 1 Volt, and R (for the 100 metre line at 1 ohm/ metre) is 200 ohms (including the two external resistors). This works out to 5mA, and indeed, that's exactly what is measured - until the "magic" frequency of 100kHz, where the current increases.

Above 1MHz, the current is 10mA (1V /100Ω), meaning that the cable resistance has disappeared! In the space of one decade in frequency, the cable has transformed itself into a true transmission line, where the signal is not conducted as such, but transferred by a waveguide (waveguides are usually just tubes for UHF signals [e.g. microwave], however a wired transmission line is also a waveguide). Naturally, the resistance never *really* disappears, but its influence is greatly reduced.

It is worth noting that a cable will *never* act as a true transmission line with a defined (and maintained) Zo unless its source and load impedances are equal to the line impedance. This means that no audio cable will ever be a transmission line, (almost) regardless of length, unless the amplifier output impedance, cable impedance and load impedance are all equal at all frequencies within the desired range. No known amplifier or loudspeaker system can meet these criteria. Alternatively, the cable may be infinitely long, however this is usually impractical in a domestic environment.

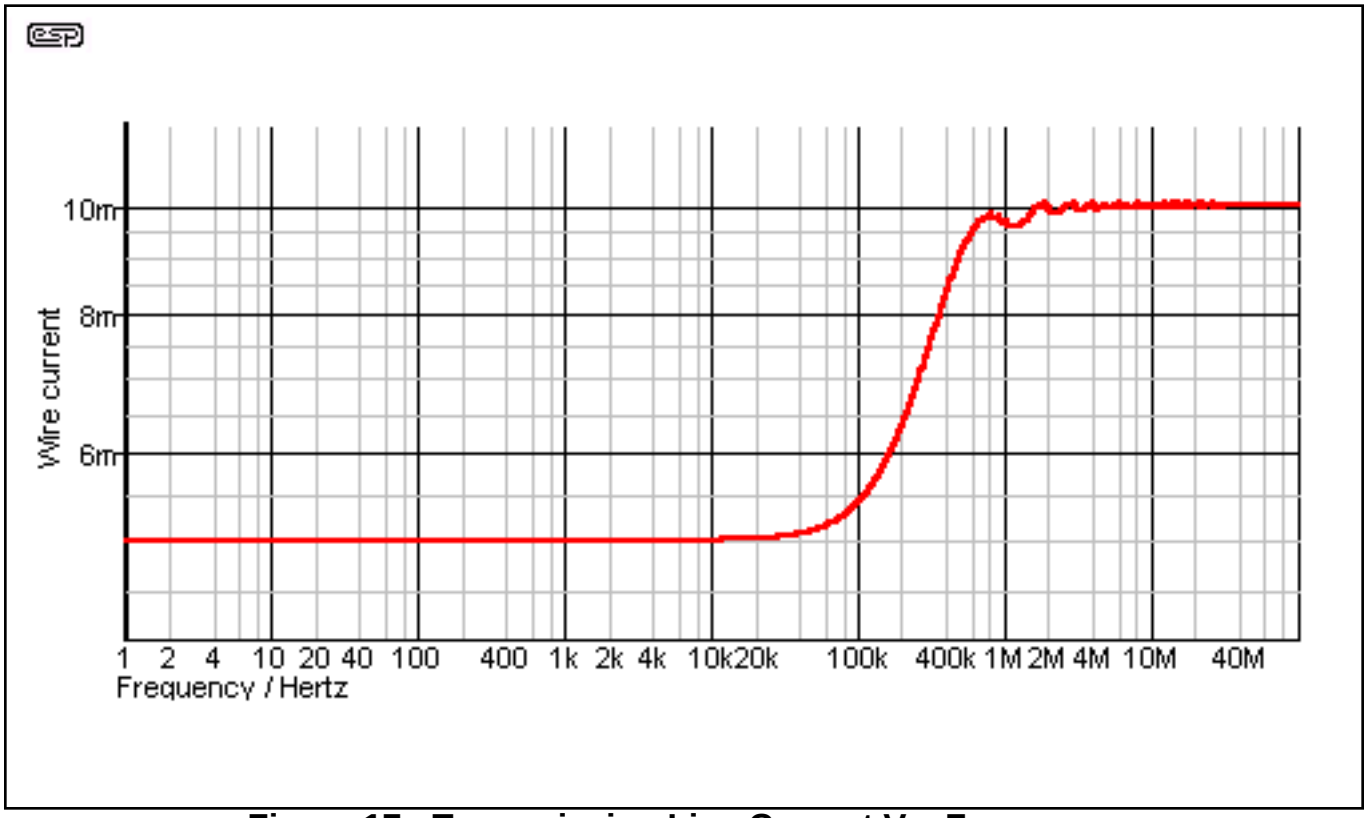


Figure 17 - Transmission Line Current Vs. Frequency

Obviously, the cable is just a pair of wires below the transition frequency, and *your* speaker cables (regardless of claims or cost) are exactly the same. In case you may want to simulate this effect yourself, I used DCR=1Ω, L=200nH and C=100pF per metre, and used a 100 metre line. A shorter line does exactly the same thing, but the frequency where it becomes a transmission line (rather than a couple of wires) increases as the line length is reduced. For a 10 metre cable, the frequency is 1Mhz, and for 3 metres this increases further to 3.3MHz. These effects were present in all the previous simulations, but were masked by the amplifier's rolloff.

5.0 Frequency Response

To be fair, it is unreasonable to investigate these cables without looking at the frequency response, so the following tests were done. In each case, response at the far end (loudspeaker) was plotted in dB, with only the cables resistive component (red graph) and with inductance and capacitance included (green graph). Response plots were done from 10Hz to 100kHz, and it is obvious that there are some differences. The same simulated amplifier was used for all tests, so its influence on the response is included. Again, all cables were measured using a 4 metre length.

Even with the very worst cable (the Aussie "Figure 8" lamp cable), response is

dominated by resistance at low frequencies. The addition of inductance and capacitance actually improved matters for the 12 gauge zip cord and Figure-8, and the far-end Zobel makes no difference (it was included for all tests, and was 100nF in series with 100Ω).

5.1 Amplifier Vs. "Ideal" Source

For comparison purposes, I checked the 12 gauge zip cable with a "perfect" zero ohm source. The maximum deviation was 0.106dB, and the predominant frequency was at 7.8kHz. This shows quite clearly that for most cable constructions, the amplifier's output impedance is a factor at low to medium frequencies. Cable inductance affects the signal at above 20kHz, while capacitance is not really an issue at all - other than its potential to make an amplifier oscillate if high enough.

The variation between a "perfect" (or ideal) source having zero ohms impedance and response from DC to daylight, compared to an amplifier with 24mΩ output impedance is not great with any cable - the 12 gauge zip shows a variance of only 0.03dB (close enough) between the two sources, and this has the lowest DC resistance of the cables tested. Obviously, the lower the DCR of a cable, the more influence the amp's output impedance has on the overall result ... until inductance becomes the predominant factor.

Note that all response measurements were done using a nominal 8 ohm load, and that inductance will have a greater effect if the load is 4 ohms (or less). The following table shows the -1dB frequency for 8Ω and 4Ω loads, with a range of inductances from 1uH to 10uH.

Inductance	8Ω -1dB	4Ω -1dB
1 uH	647 kHz	325 kHz
2 uH	324 kHz	162 kHz
3 uH	216 kHz	109 kHz
4 uH	162 kHz	82 kHz
5 uH	130 kHz	65 kHz
6 uH	108 kHz	54 kHz
7 uH	92 kHz	46 kHz
8 uH	81 kHz	41 kHz
9 uH	72 kHz	36 kHz

10 uH	65 kHz	32 kHz

Table 3 - Frequency Rolloff Vs. Inductance and Load Impedance

It is quite obvious that a total inductance of up to 10uH will be quite acceptable for the highest of fidelity with any loudspeaker that has a benign impedance at the high end of the audio spectrum, at either 4 or 8 ohms. Most speaker systems are reasonably consistent at the high frequency end, but obviously there are exceptions, and these *will* cause audible differences.

5.2 Response Simulations

These results are all simulated (rather than measured), since this is the fastest way to achieve the results, and the simulations will agree with "reality" very well. While there are other factors that are not taken into account (such as skin effect or insulation material "soakage"), these are generally considered to be inaudible, and no proof has ever been offered that anyone can distinguish the difference in a DBT. That there *are* differences is undeniable, they can be measured quite easily with the right equipment (which I don't have), but any such effects are well below the noise floor and/or resolution of even the best amplifiers and speakers. Indeed, atmospheric changes will cause *far* greater variations in the signal you hear.

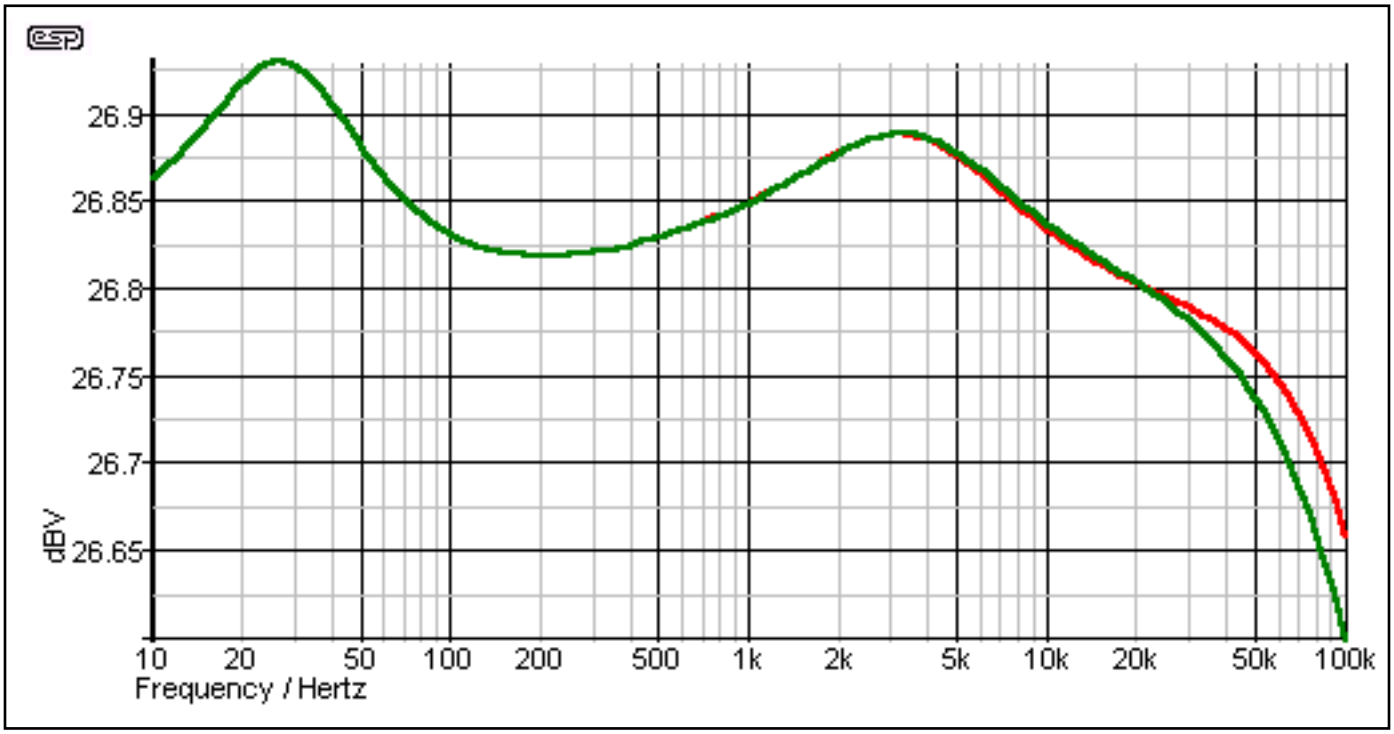


Figure 18 - Cable #3 Frequency Response

This looks terrible, until you notice the dB range. Total variation is from a low of 26.8dB

at 20kHz to a peak of 26.93dB at about 27Hz (woofer resonance, and the highest impedance the loudspeaker presents). A total variation of 0.13dB. Midband level is 26.826dB at 200Hz.

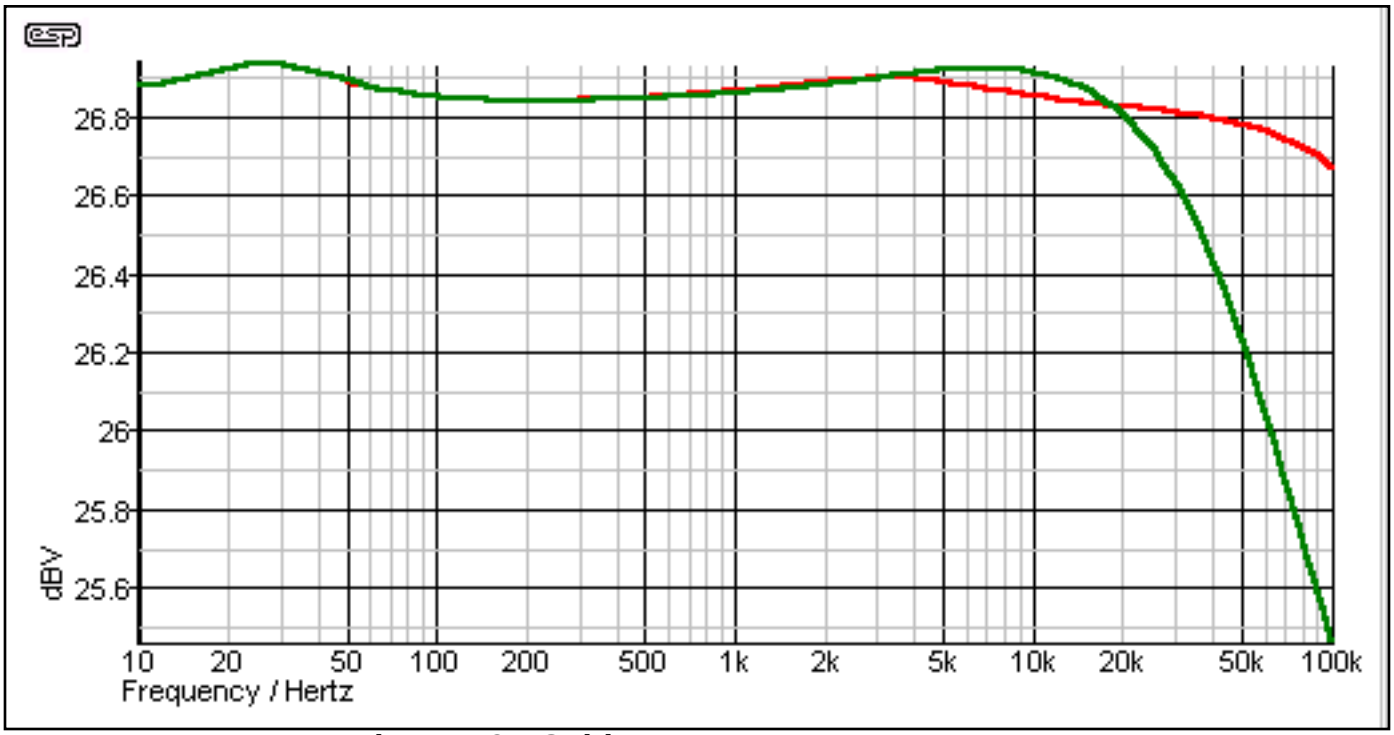


Figure 19 - Cable #1 Frequency Response

The zip cable looks better, and at a very small fraction of the cost. Minimum is 26.8dB at 20kHz and maximum is 26.93dB at woofer resonance - a total deviation of 0.13dB. The HF rolloff between 10kHz and 20kHz is 0.1dB (hardly woeful - find a tweeter that good!) Midband level is 28.84dB at 200Hz.

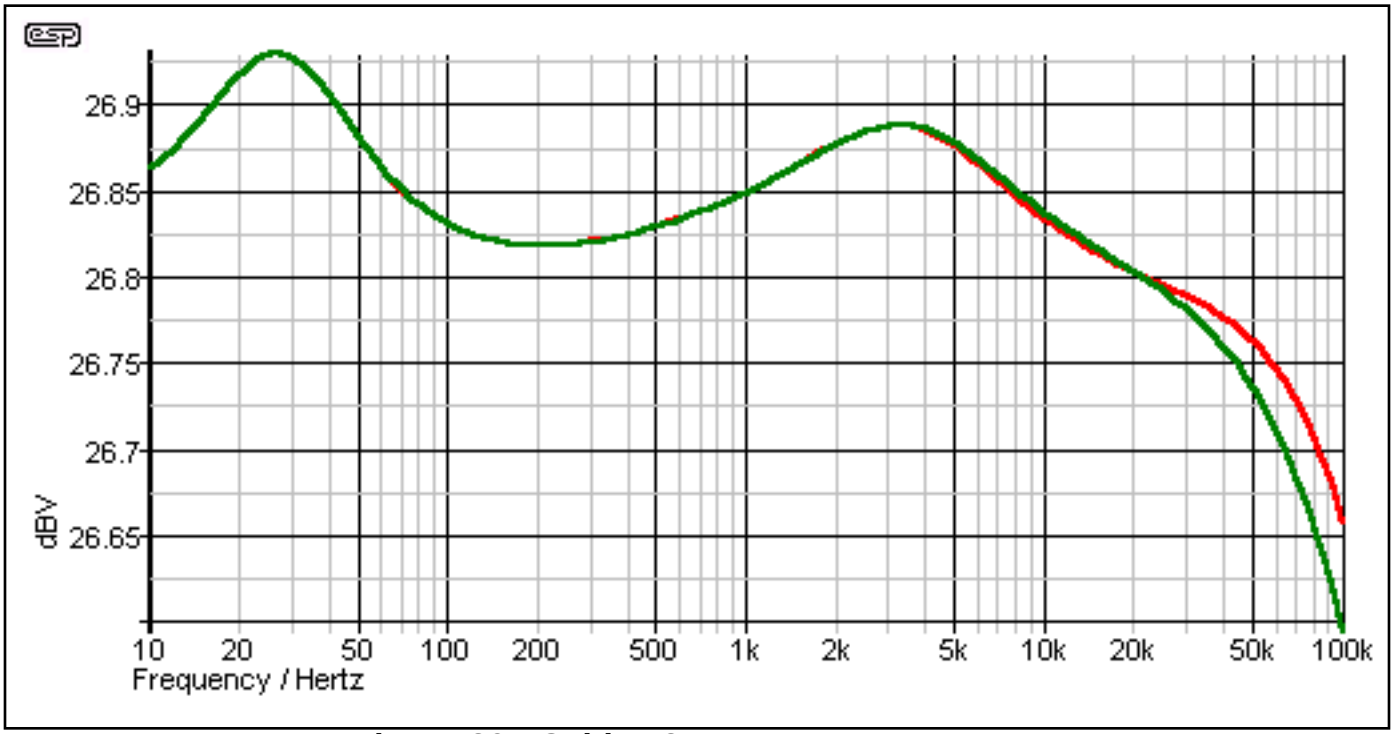


Figure 20 - Cable #2 Frequency Response

Cable #2 is again at 26.8dB at 20kHz, peaking at 26.93dB - a total variation of 0.13dB. Midband level is 26.82dB at 200Hz.

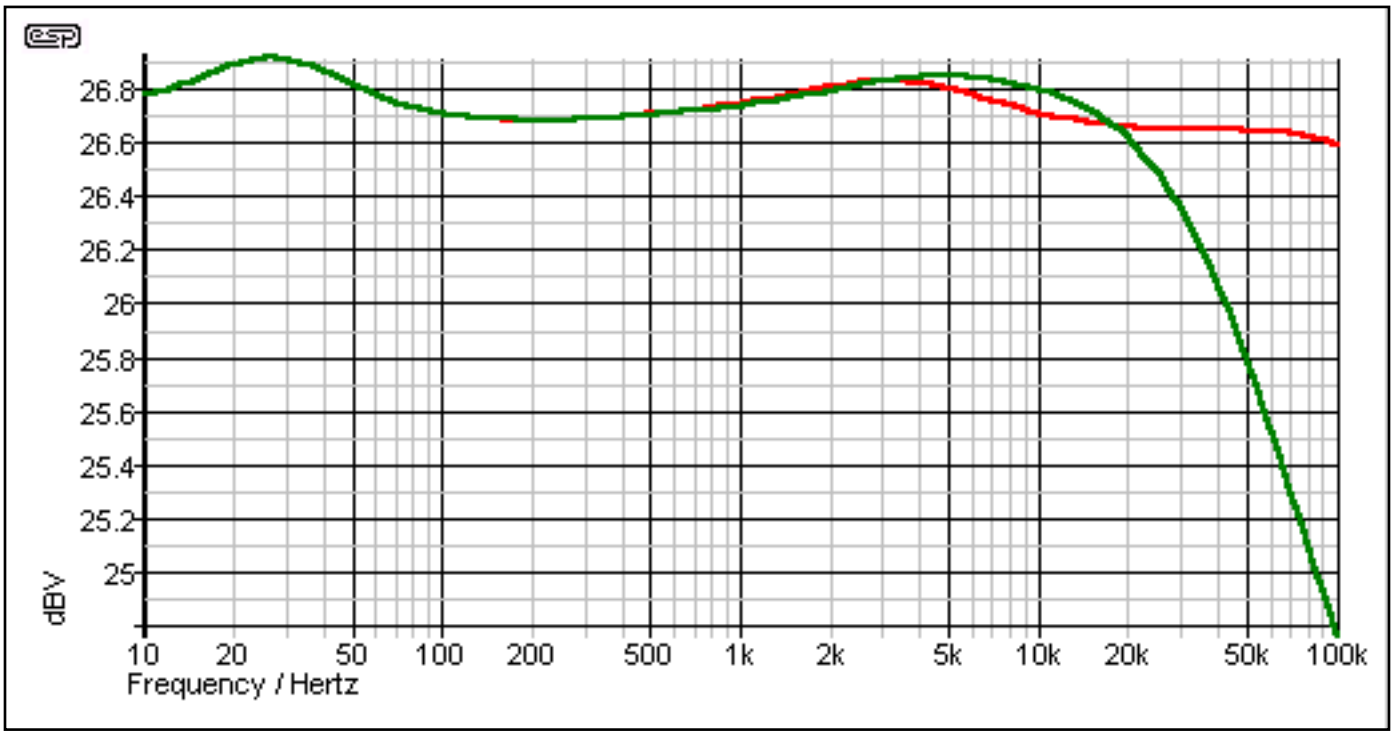


Figure 21 - "Figure 8" Frequency Response

Last (in all respects) comes the Figure-8 cable, but at the equivalent of perhaps 14 gauge (US gauge numbering), one expects it to be less than impressive. Minimum level is 26.6dB at 20kHz, and maximum is around 26.91dB - a total deviation of 0.31dB.

Midband level is about 26.68dB at 200Hz.

It is worth noting that none of these cables is 1dB down on midband level at any frequency up to 20kHz, and even the Figure-8 lead is only 1dB down at 50kHz - it is a rare tweeter indeed that will be anywhere near as good, regardless of price.

So, what do you get for your money? Quite obviously, very little improvement is afforded by any of these cables over another - 12 gauge zip cable is cheap and easy to make into a speaker lead, Jon's cable is a little more expensive, and there is a fair amount of work involved, and the Goertz cable will set you back about US$200 a pair for 4 metre lengths (and maybe make your amplifier oscillate). Naturally, you can spend a great deal more (and still make the amp oscillate), but I don't see much point.

If it makes you feel better to have sexy looking cross-connected coax leads, then far be it for me to attempt to deny you that pleasure, besides, it might be fun to do (which is far more important). I still can't recommend the Goertz cable, as its capacitance is just too high. A Zobel tames that, but I would be reluctant to use it anyway, and I certainly wouldn't pay their prices for it!

The Web has a great many examples of over-the-top cable pricing (although some are a lot cheaper than others), claims and mistakes. This is not to single out any manufacturer - it is simply to point out that a great many examples can be found of "high-end" cables with claims that cannot be substantiated by DBT listening test or simulation. There are (of course) so many that they are too numerous to mention.

Decisions, decisions ....

6.0 References

Online Radio and Electronics Course

Principles of Transmission Lines

All simulations were carried out using SIMetrix, and a free demo version is available from Catina Software Ltd in the UK. This is (IMO) an outstanding simulator, and regularly produces results that can be transferred directly from a simulation to a working circuit, without the need for any changes at all.

Cable parameters were taken from the Audioholics website, and I thank Gene DellaSala for permission and for his support for this article.

The simulated loudspeaker is based on a version published by Jon Risch, and his DIY cross-connected coaxial cable was used for sample #2 in the simulations. For those who wish to experiment with "exotic" designs, those published by Jon are high performance and cheap to build - certainly a major departure from the US$1000/metre offerings that seem to proliferate in the market. Jon's designs are also far less likely to cause amplifier instability than many of the commercial offerings!

Finally, the parameters for cable #3 were obtained from the Alpha-Core website.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 04 May 2003./ Published 17 Oct 2003

Elliott Sound Products **Amplifier Sound - What Are The Influences?**

Rod Elliott (ESP)

Articles Index

Main Index

Introduction

The sound of an amplifier is one of those ethereal things that seems to defy description. I will attempt to cover the influences I know about, and describe the effects as best I can. This is largely hypothesis on my part, since there are so many influences that, although present and audible, are almost impossible to quantify. Especially in combination, some of the effects will make one amp sound better, and another worse - I doubt that I will be able to even think of all the possibilities, but this article might help some of you a little - at least to decipher some of the possibilities.

I don't claim to have all the answers, and it is quite conceivable that I don't have any (although I do hope this is not the case). This entire topic is subject to considerable interpretation, and I will try very hard to be completely objective.

Readers' input is encouraged, as I doubt that I will manage to get everything right first time, and there are some areas where I do not really know what the answers are. The only joy I can get from this is that I doubt that anyone else can do much better. If you can, let me know.

The Components of Sound

When people talk about the sound of an amplifier, there are many different terms used. For a typical (high quality) amplifier, the sound may be described as "smeared", or having "air" or "authoritative" bass. These terms - although describing a listener's experience - have no direct meaning in electrical terms.

Electrically, we can discuss distortion, phase shift, current capability, slew rate and a myriad of other known phenomena. I don't have any real idea as to how we can directly link these to the common terms used by reviewers and listeners.

Some writers have claimed that all amplifiers actually sound the same, and to some extent (comparing apples with apples) this is "proven" in double-blind listening tests. I am a great believer in this technique, but there are some differences that cannot be readily explained. An amp that is deemed "identical" to another in a test situation, may sound completely different in a normal listening environment. It is these differences that are the hardest to deal with, since we do not always measure some of the things that can have a big influence on the sound.

For example; It is rare that testing is done on an amplifier's clipping performance - how the amp recovers from a brief transient overload. I have stated elsewhere that a hi-fi amplifier should never clip in normal usage - nice try, but it IS going to happen, and often is more common than we might think. Use a good clipping indicator on the amp, and this can be eliminated, but at what cost? It might be necessary to reduce the volume (and SPL) to a level that is much lower than you are used to, to eliminate a problem that you were unaware existed.

Different amplifiers react in different ways to these momentary overloads, where their overall performance is otherwise almost identical. I have tested IC power amps, and was dismayed by the overload recovery waveform. My faithful old 60W design measures about the same as the IC in some areas, a little better in some, a little worse in others (as one would expect).

Were these two amps compared in a double blind test (avoiding clipping), it is probable that no-one would be able to tell the difference. Advance the level so that transients started clipping, and a fence post would be able to hear the difference between them. What terms would describe the sound? I have no idea. The sound might be "smeared" due to the loss of detail during the recovery time of the IC amp. Imaging might suffer as well, since much of the signal that provides directional cues would be lost for periods of time.

Measurable Performance Characteristics

A detailed description of the more important (from a sound perspective) of the various amplifier parameters is given later in this article, but a brief description is warranted first. Items marked with a * are problem areas, and the effect should be minimised wherever possible. The parameters that should normally be measured (although for those marked # this is rare indeed) are as follows:

* Important parameter
# Rarely measured

- **Input Sensitivity :** The signal level required to obtain full power at the amplifier's output. This is determined by the gain and power rating of the amp. A 10W amplifier requires far less gain than a 200W amplifier to obtain full power for the same input voltage. It would be useful if all amplifiers had the same gain regardless of power, but this is not the case. Sensitivities vary widely, ranging from about 500mV up to 1.5V or more.
- **Total Harmonic Distortion (THD) * :** This is a measure of the amount of distortion (modification) of the input signal, which adds additional signal frequencies to the output that are not present in the input signal. THD is commonly measured as a percentage, and can range from 0.001% to 0.1% for typical hi-fi amplifiers. A theoretically perfect amplifier contributes no distortion.
- **Transient Intermodulation Distortion (TIM) * :** Also sometimes called slew induced distortion, this is a form of distortion said to occur when the input signal changes so fast that the output cannot keep up with it. When this happens, feedback ceases to be effective, since the output signal is delayed too long. This remains somewhat contentious, and most modern amplifiers are quite capable of handling the normal programme amplitude and frequency range without difficulty.
- **Crossover Distortion *# :** A form of distortion caused by the power output devices in a push-pull amplifier operating in Class-AB. This occurs in valve and solid state designs, and is caused by one device switching off as the other takes over for its half of the waveform. There are some designs that claim to eliminate this distortion by never turning off the power devices, but in reality, only Class-A amplifiers have zero crossover distortion. This is generally measured as a part of the THD of an amplifier, and becomes worse as power is reduced from the maximum.
- **Frequency Response * :** The amount of frequency versus amplitude distortion in an amplifier. A perfect amplifier will amplify all signals equally, regardless of frequency. Realistically, an amplifier needs a

response of about 5Hz to 50kHz to ensure that all audible signals are catered for with minimal modification.

- **Phase Response :** This indicates the amount of time that the input signal is delayed before reaching the output, based on the signal frequency. Variations in absolute phase are not audible in an amplifier system, but are generally considered undesirable by the hi-fi press. Since it is not difficult to ensure phase linearity, this is not generally a design issue except with valve amplifiers.
- **Output Power :** This is most commonly measured into a non-inductive resistive load. This is not done to improve the figures or disguise any possible shortcomings, but to ensure that measurements are accurate and repeatable. Power should only ever be quoted as "RMS", which although is not strictly correct, is accepted in the industry, and may be measured into 8 Ohms, or other impedances that the amplifier is capable of driving.
- **Output Current # :** Not often measured, but sometimes quoted by manufacturers, this represents the maximum current the amplifier can supply into any load. It is rare that any amplifier will be called upon to deliver any current greater than about 3 to 5 times the maximum that the nominal speaker impedance would allow for the amplifier's supply voltage. Greater variations may be possible with some speaker designs, but (IMO) this represents a flaw in the design of the loudspeaker.
- **Power Bandwidth :** This is usually taken as the maximum frequency at which the amplifier can produce 1/2 of its rated output power (this is the -3dB frequency). A 100W amplifier that can produce 50W at 50kHz will be deemed as having a 50kHz power bandwidth.
- **Slew Rate # :** Closely related to power bandwidth, the slew rate is the maximum rate of change (measured in Volts per microsecond) of the amplifier output. The higher the amplifier power, the higher the slew rate must be to obtain the same power bandwidth.
- **Open Loop Bandwidth # :** The bandwidth of the amplifier with no AC feedback applied. Very few amplifiers will have an open loop bandwidth greater than a few kilo-Hertz, but valve amps and some solid state designs have a comparatively high open loop bandwidth.
- **Open Loop Gain # :** Rarely quoted except for DIY amps (and few of them as well), this is the gain of the amplifier without any AC signal feedback. It is not really a helpful parameter for most people, but can be used to determine the ...
- **Open Loop Distortion #* :** The THD of the amplifier with no feedback applied. This should be as low as possible, but realistically will usually be quite high by normal standards. The open loop distortion is reduced by an amount approximately equal to the feedback ratio.
- **Open Loop Output Impedance # :** The output impedance of the amplifier with no AC feedback applied. This may range from a few Ohms to 10 or more Ohms, depending on the design of the amplifier. Valve amplifiers will normally have an open loop output impedance of 0.7 of the designed speaker impedance.
- **Feedback Ratio # :** How much of the open loop gain is fed back to the amplifier's input to obtain the sensitivity figure quoted for the amp. For example if an amplifier has an open loop gain of 100dB, and a gain of 20dB, then the feedback ratio is 80dB. The application of feedback will
  - Increase bandwidth
  - Reduce phase shift
  - Reduce distortion
  - Reduce output impedance
- **Output Impedance * :** This is the actual output impedance of the amplifier, and has no bearing on the amount of current that can be supplied by the output stage. Valve amplifiers usually have a relatively high output impedance (typically 1 to 6 Ohms), while solid state amps will normally have an output impedance of a fraction of an Ohm. By use of feedback, it is possible to increase output impedance (> 200 Ohms is quite easy), or it can be made negative. Negative impedance has been tried by many designers (including the author), but has never gained popularity - possibly because most speakers react very poorly to negative impedances and tend to sound awful.

Every amplifier design on the planet has the same set of constraints, and will exhibit all of the above problems

to some degree. The only exception is a Class-A amplifier, which does not have crossover distortion, but is still limited by all other parameters.

The difficulty is determining just how much of any of the problem items is tolerable, and under what conditions. For example, there are many single ended triode valve designs which have very high distortion figures (comparatively speaking), high output impedance and low output current capability. There are many audio enthusiasts who claim that these sound superior to all other amplifiers, so does this mean that the parameters where they perform badly (or at least not as well as other amps) can be considered unimportant? Not at all!

If a conventional (i.e. not Class-A) solid state amplifier gave similar figures, it would be considered terrible, and would undoubtedly sound dreadful.

Although all the issues described above are separate in their own right, many can be lumped together under a single general category....

Distortion
Technically, distortion is any change that takes place to a signal as it travels from source to destination. If some of the signal "goes missing", this is distortion just as much as when additional harmonics are generated.

We tend to classify distortion in different ways - the imperfect frequency response of an amplifier is not generally referred to as distortion, but it is. Instead, we talk about frequency response, phase shift, and various other parameters, but in reality they are all a form of distortion.

The bottom line is that amplifiers all suffer from some degree of distortion, but if two amplifiers were to be compared that had no distortion at all, they must (by definition) be identical in both measured and perceived sound.

Naturally, there is no such thing as a perfect amplifier, but there are quite a few that come perilously close, at least within the audible frequency range. What I shall attempt to do is look at the differences that do exist, and try to determine what effect these differences have on the perceived "sonic quality" of different amplifiers. I will not be the first to try to unravel this mystery, and I doubt that I will be the last. I also doubt that I will succeed, in the sense that success in this particular area would only be achieved if everyone agreed that I was right - and of that there is not a chance! (However, one lives in hope.)

In this article I use the somewhat outdated term "solid state" to differentiate between valve amps, and those built using bipolar transistors, MOSFETs or other non-vacuum tube devices.

I have also introduced a new (?) test method, which I have called a SIM (Sound Impairment Monitor), the general concept of which is described in the appendix to this article.

Clipping Distortion

How can one amplifier's clipping distortion sound different from that of another? Most of the hi-fi fraternity will tend to think that clipping is undesirable in any form at any time. While this is undeniably true, many of the amps used in a typical high end setup will be found to be clipping during normal programme sessions. I'm not referring to gross overload - this is quite unmistakable and invariably sounds awful - regardless of the amplifier.

There are subtle differences between the way amplifiers clip, that can make a very great impact on the sound.

Valve amps are the most respectable of all, having a "soft" clipping characteristic which is comparatively unobtrusive. Low feedback Class-A amplifiers are next, with slightly more "edge", but otherwise are usually free from any really nasty additions to the overall sound.

Then there are the myriad of Class-AB discrete amps. Most of these (but by no means all) are reasonably well behaved, and while the clipping is "hard" it does not have significant overhang - this is to say that once the output signal is lower than the supply voltage again it just carries on as normal. This is the ideal case - when any amp clips, it should add no more nastiness to the sound than is absolutely necessary. Clipping refers to the fact that when the instantaneous value of output signal attempts to exceed the amplifier's power supply voltage, it simply stops, because it cannot be greater than the supply. We know it must stop, but what is of interest is how it stops, and what the amplifier does in the brief period during and immediately after the clipping has occurred.

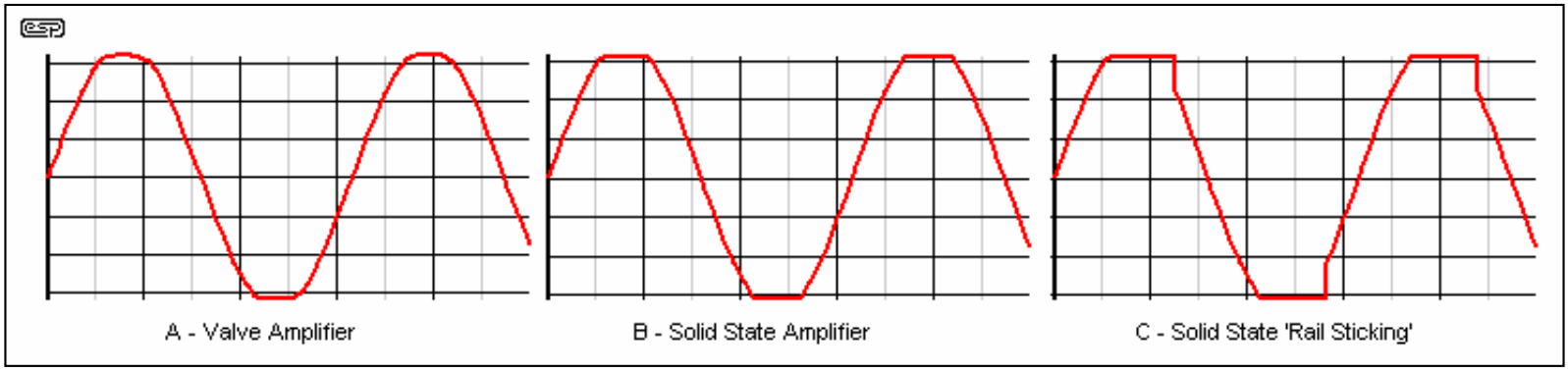


Figure 1 - Comparison of Basic Clipping Waveforms

In Figure 1, you can see the different clipping waveforms I am referring to, with "A" being representative of typical push-pull valve amps, "B" is the waveform from a conventional discrete Class-AB solid state amp, and "C" shows the overhang that is typical of some IC power amps as well as quite a few discrete designs. This is a most insidious behaviour for an amp, because while the supply is "stuck" to the power rail, any signal that might have been present in the programme material is lost, and a 100Hz (or 120Hz) component is added if the clipping + "stuck to rail" period lasts long enough. This comes from the power supply, and is only avoidable by using a regulated supply or batteries. Neither of these is cheap to implement, and they are rarely found in amplifier designs.

Although Figure 1 shows the signal as a sinewave for ease of identification, in a real music signal it will be a sharp transient that will clip, and if the amp behaves itself, this will be (or should be) more or less inaudible. Should it stick to the supply rail, the resulting description of the effect is unlikely to accurately describe the actual problem, but describe what it has done to the sound - from that listener's perspective. A simple clipped transient should not be audible in isolation, but will have an overall effect on the sound quality. Again, the description of this is unlikely to indicate that the amp was clipping, and regrettably few amps have clipping indicators so most of the time we simply don't know it is happening.

To be able to visualise the real effect of clipping, we need to see a section of "real" signal waveform, with the lowest and highest signal frequencies present at the same time. If the amp is clipped because of a bass transient (this is the most common), the period of the waveform is long. even if the signal is clipped for only 5 milliseconds, this means that 5 complete cycles of any signal at 1000Hz is removed completely, or 50 complete cycles at 10kHz. This represents a significant loss of intended information, which is replaced by a series of harmonics of the clipped frequency (if clipping lasts for long enough), or more typically a series of harmonics that is not especially related to anything (musically speaking - all harmonics are related to something, but this is not necessarily musical!)

I think that no review of any amplifier should ever be performed without some method of indicating that the amp is clipping (or is subject to some other form of signal impairment), and this can be added to the reviewer's notes - along the lines of:

"This amplifier was flawless when kept below clipping (or as long as the SIM failed to show any noticeable impairment), but even the smallest amount of overload caused the amp to sound very hard. Transparency was completely lost, imaging was impaired badly, and it created listener fatigue very quickly."

Now, wouldn't that be cool? Instead of us being unaware (as was the reviewer in many cases) that the amp in review was being overdriven - however slightly - we now (all of us) have that missing piece of information that is not included at the moment. I have never seen a review of an amp where the output was monitored with an accurate clipping indicator to ensure that the reviewer was not listening to a signal that was undistorted. I'm not saying that no-one does this, just none that I have read.

The next type of overload behaviour is dramatically worse, and I have seen this in various amps over the years. Most commonly associated with overload protection circuits, the sound is gross. I do not know the exact mechanism that allows this to happen, but it can be surmised that the protection system has "hysteresis", a term that is more commonly associated with thermal controllers, steel transformer laminations and Schmitt trigger devices. Basically, a circuit with hysteresis will operate once a certain trigger point is reached, but will not reset until the input signal has fallen below a threshold that is lower than the trigger point. The typical waveform of an amplifier with this problem is shown in Figure 2, and I believe it IS a problem, and should be checked for as a normal part of the test process. This type of overload characteristic is not desirable in any way, shape or form.

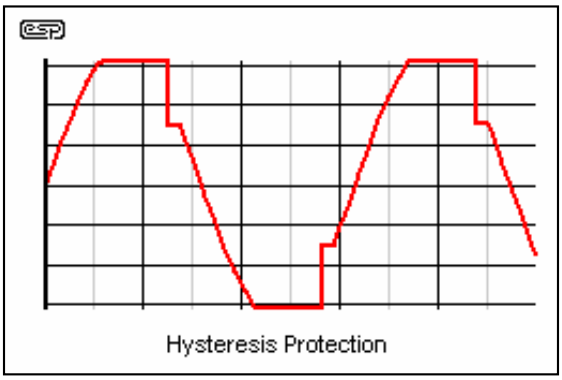


Figure 2 - Hysteresis Overload Waveform

In this case, the additional harmonic components added to the original sound will be more prominent than with "normal" clipping. As before, I cannot even begin to imagine how the sound might be described - all the more reason to ensure that testing includes informing the reader if the amp was clipping or not during the listening tests. The loss of signal with this type of distortion will generally be much greater than simple clipping, and the added harmonic content will be much more pronounced, especially in the upper frequencies.

Clipping Synopsis

Tests conducted as a part of any review would be far more revealing if the clipping waveform were shown as a matter of course. After some learning on our behalf, we would get to know what various of the hi-fi press meant when they described the sound while the amp was clipping, versus not clipping, or what the amp sounded like when it's overload protection circuits came into action.

To this end I have designed a new distortion indicator circuit, which not only indicates clipping, but will show when the amp is producing distortion of any kind beyond an acceptable level. One version has been published

as a project, and I have chosen the acronym SIM (Signal Impairment Monitor) for this circuit.

The SIM will react to any form of signal modification, and this includes phase distortion and frequency response distortion. I do not believe that this approach has been used before in this way. It is not an uncommon method for distortion measurement, but has not been seen anywhere as a visual indicator for identifying problem areas that an amp may show in use. This circuit will also show when an amplifier's protection circuit has come into effect.

Although the detector has no idea what type of problem is indicated, it does indicate when the input and output signals no longer match each other - for whatever reason. Oscilloscope analysis would be very useful using this circuit, as with a little practice we would be able to identify many of the currently unknown effects of various amplifier aberrations. Note that it is unable to reveal crossover distortion unless it becomes quite high, which is extremely unlikely in any modern amplifier design.

Crossover Distortion

Class-A amplifiers have no crossover distortion at all, because they maintain conduction in the output device(s) for the entire waveform cycle and never turn off. Class-A is specifically excluded from this section for that reason.

For the rest, a similar question as the one before - how can one amplifier's crossover distortion sound different from another? Surely if there is crossover distortion it will sound much the same? Not so at all. Again, valve amplifiers are much better in this area than solid state amps (at least in open loop conditions). When valves cross over from one output device to the next (standard push-pull circuit is assumed), the harmonic structure is comprised of mainly low order odd harmonics. There will be some 3rd harmonics, a smaller amount of 5th, and so on.

Solid state amps tend to create high order odd harmonics, so there might be more of the 5th harmonic than the 3rd, and even more 7th (and so on). Transistor and MOSFET amps have very high open loop gains, and use feedback to reduce distortion. In all cases, the crossover distortion is caused because the power output devices are non-linear. At the low currents at which the changeover occurs, these non-linearities are worse, as well, the devices usually have a lower gain at these currents.

This has two effects. The open loop gain of the amplifier is reduced because of the lower output device gain, so there is less negative feedback where it is most needed. Secondly, the feedback tries to compensate for the lower gain (and tries to eliminate the crossover distortion), but is limited by the overall speed of the internal circuitry of the amplifier. This results in sharp transitions in the crossover region, and any sharp transition means high order harmonics are produced (however small they might be).

One method to minimise this is to increase the quiescent (no signal) current in the output transistors. With a linear output stage in a well designed circuit, crossover distortion should be minimal with any current above about 50 to 100mA. Figure 3 shows the crossover distortion (at the centre of the red trace) and the residue as seen on an oscilloscope (green trace, amplified by 10 for clarity) - this is the typical output from a distortion meter, with an amplifier that has noticeable crossover distortion. If measured properly, the distortion is highly visible, even though it may be barely audible. Note that the waveform below would not qualify for the last statement - this amount of crossover distortion would be very audible indeed.

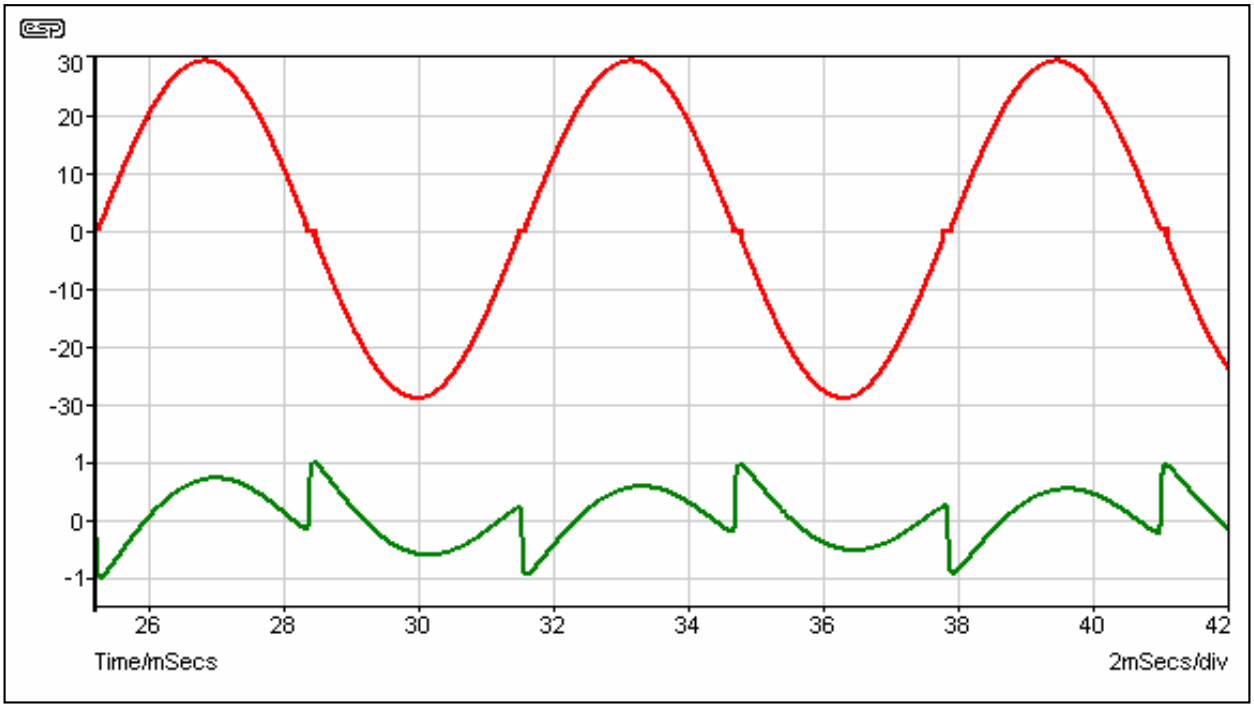


Figure 3 - Crossover Distortion Waveform

If THD is quoted without reference to its harmonic content, then it is quite possible that two amplifiers may indicate identical distortion figures, but one will sound much worse than the other. Distortion at a level of 1W should always be quoted, and the waveform shown. Once the waveform can be seen, it is easy to determine whether it will sound acceptable or dreadful - before we even listen to the amp. Listening tests will confirm the measured results with great accuracy, although the descriptive terms used will vary, and may not indicate the real problem.

Crossover Distortion Synopsis

Although this is one area where modern amplifiers rarely perform badly, it is still important, and should be measured and described with more care than is usually the case. While few amplifiers will show up badly in this test now, crossover distortion was one of the main culprits that gave solid state a bad name when transistors were first used in amplifiers.

I do not believe that we can simply ignore crossover distortion on the basis the "everyone knows how to fix it, and it is not a problem any more". I would suggest that it is still a real problem, only the magnitude has been reduced - the problem is still alive and well. Will you be able to hear it with most good quality amp? Almost certainly not.

Frequency And Phase Distortion

Distortion of the frequency response should not be an issue with modern amplifiers, but with some (such as single ended triode valve designs), it does pose some problems. The effect is that not all frequencies are amplified equally, and the first to go are the extremes at both ends of the spectrum. It is uncommon for solid state amps to have a frequency response at low powers that extends to anything less than the full bandwidth from 20Hz to 20kHz. This is not the case with some of the simple designs, and single ended triode (SET) Class-A - as well as inductance loaded solid state Class-A amps - will often have a less than ideal response.

I would expect any amplifier today should be no more than 0.5dB down at 20Hz and 20kHz, referred to the mid-band frequency (usually taken as 1kHz, but is actually about 905Hz). (My preferred test frequency is 440Hz

(concert pitch A, below middle C), but none of this is of great consequence.) 0.5dB loss is acceptable in that it is basically inaudible, but most amps will do much better than this, with virtually no droop in the response from 10Hz to over 50kHz.

For reference, the octaves included for "normal" sound are:

20 40 80 160 320 640 1280 2560 5120 10240 20480

To determine the halfway point between two frequencies one octave apart, we multiply the lower frequency by the square root of 2 (1.414). The halfway point is between 650 and 1280Hz, or 904.96Hz. You must be so pleased to have been provided with this piece of completely useless information! Just think yourselves lucky that I didn't tell you how to calculate the distance between the frets on a guitar. :-D

Most amplifiers will manage well beyond the range necessary for accurate reproduction, at all power levels required to cater for the requirements of music. So why are some amps described as having poor rendition of the high frequencies? They may be described a "veiled" or something similar, but there is no measurement that can be applied to reveal this when an amplifier is tested. Interestingly, some of the simpler amplifiers (again, such as the single ended triode amps) have poorer response than most of the solid state designs, yet will regularly be described as having highs that "sparkle", and are "transparent".

These terms are not immediately translatable, since they are subjective, and there is no known measurement that reveals this quality. We must try to determine what measurable effect might cause such a phenomenon. There are few real clues, since amplifiers that should not be classified as exceptional in this area are often described as such. Other amps may be similarly described, and these will not have the distortion of a single ended triode and will have a far better response.

We can (almost) rule out distortion as a factor in this equation, since amps with comparatively high distortion can be comparable to others with negligible distortion. Phase shift is also out of the question, since amps with a lot of phase shift can be favourably compared to others with virtually none. One major difference is that typical SET amplifiers have quite high levels of low order even harmonics. Although these will give the sound a unique character, I doubt that this is the sole reason for the perceived high frequency performance - I could also be wrong.

Phase distortion occurs in many amplifiers, and is worst in designs using an output transformer or inductor (sometimes called a choke). The effect is that some frequencies are effectively delayed by a small amount. This delay is usually less than that caused by moving one's head closer to the loudspeakers by a few millimetres. It is generally thought to be inaudible, and tests that I (and many others) have conducted seem to bear this out.

Frequency And Phase Distortion - Synopsis

There must be some mechanism that causes multiple reviewers to describe an amplifier as having a poor high frequency performance, such as (for example) a lack of transparency. There are few real clues that allow us to determine exactly what is happening to cause these reviewers to describe the sound of the amp in such terms, and one may be tempted to put it all down to imagination or "experimenter expectancy". This is likely to be a mistake, and regardless what we might think about reviewers as a species, they do get to listen to many more amplifiers than most of us.

One of the few variables is a phenomenon called slew rate. This is discussed fully in the next section.

Slew Rate Distortion

This has always been somewhat controversial, but no-one has ever been able to confirm satisfactorily that slew rate (within certain limits) has any real effect on the sound. Figure 4 is a nomograph that shows the required slew rate for any given power output to allow full power at any frequency. To use it, determine the power and calculate the peak voltage, and place the edge of a ruler at that voltage level. Tilt the ruler until the edge also aligns with the maximum full power frequency on the top scale. The slew rate is indicated on the bottom scale.

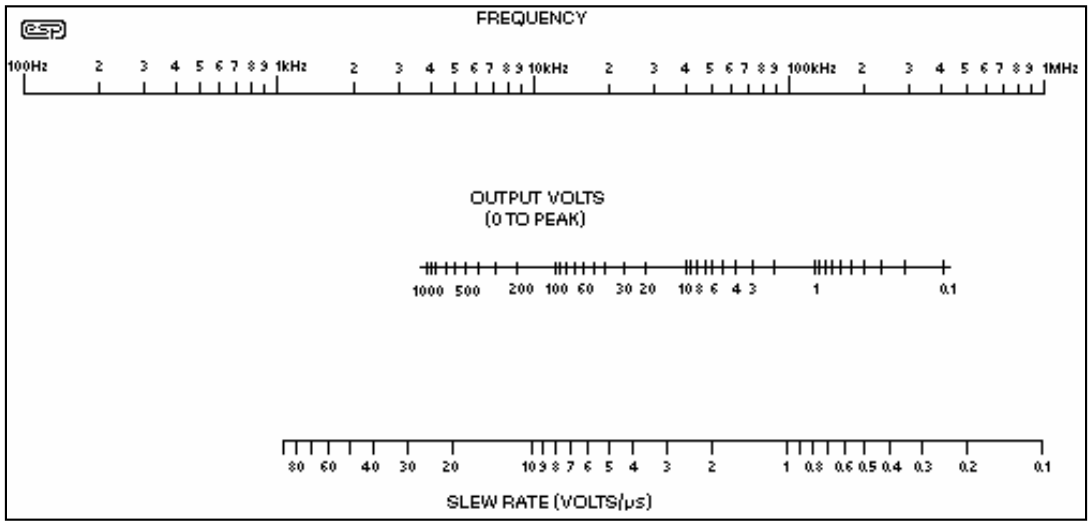


Figure 4 - Slew Rate Nomograph

Slew rate distortion is caused when a signal frequency and amplitude is such that the amplifier is unable to reproduce the signal as a sine wave. Instead, the input sine wave is "converted" into a triangle wave by the amplifier. This is shown in Figure 5, and is indicative of this behaviour in any amplifier with a limited slew rate. The basic problem is caused by the "dominant pole" filter included in most amplifiers to maintain stability and prevent high frequency oscillation. While very few amplifiers even come close to slew rate induced distortion (AKA Transient Intermodulation Distortion) with a normal signal, this is one of the very few possibilities left to explain why some amps seem to have a less than enthusiastic response from the reviewers' perspective.

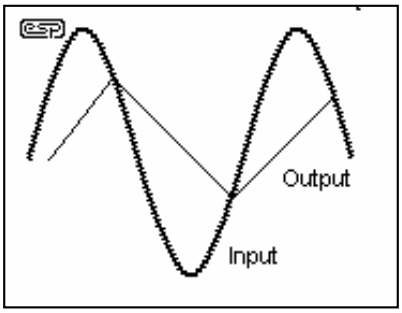


Figure 5 - Slew Rate Limiting In An Amplifier

Is this the answer, then? I wish it were, since we could all sleep soundly knowing exactly what caused one amp to sound the way it did, compared to another, which should have sounded almost identical.

A further test is to apply a low frequency square wave at about half to 3/4 power, mixed with a low-level high frequency sinewave to the amplifier. At the transitions of the squarewave, the sinewave should simply move up and down - "riding" the squarewave. If there is any mis-behaviour in the amp, the sinewave may be seen to be compressed so its shape will change, or a few cycles may even go missing entirely. Either is unacceptable, and should not occur.

This is an extremely savage test, but most amplifiers should be able to cope with it quite well. Those that don't will modify the music signal in an unacceptable way in extreme cases (which this test simulates). Again, this is an uncommon test to perform, but may be quite revealing of differences between amps.

Frequency And Slew Rate Distortion - Synopsis

We need to delve deeper, and although there seems to be little (if any) useful evidence we can use to explain this particular problem, there is an answer, and it therefore possible to measure the mechanism that causes the problem to exist.

Open Loop Responses

The performance of a feedback amplifier is determined by two primary factors. These are

- Open loop performance
- Feedback ratio

If the amp has a poor open loop gain and high distortion, then sensible amounts of feedback will not be able to correct the deficiencies, because there is not sufficient gain reserve. By the time the performance is acceptable, it may mean that the amplifier has unity gain, and is now impossible to drive with any normal preamp.

Many amplifiers have a very high open loop gain, but may have a restricted frequency response. Let's assume an amp that has a gain of 100dB at 20Hz, and 40dB gain at 20kHz. If we want 30dB of overall gain (which is about standard), then there is 70dB of feedback at 20Hz, but only 10dB at 20kHz. As a very rough calculation, distortion and output impedance are reduced by the feedback ratio, so if open loop distortion were 3% (not an unreasonable figure), then at 20Hz, this is reduced to 0.0015%, but will be only just under 1% at 20kHz.

Because these figures are so rarely quoted (and I must admit, I have not really measured all the characteristics of the 60W amp in Project 03 - open loop measurements are difficult to make accurately), we have no idea if amplifiers with poor open loop responses are responsible for so many of the failings we hear about. It is logical to assume that there must be some correlation, but we don't really know for sure.

Ideally, an amplifier should have wide bandwidth and low distortion before global feedback is applied, which will just make a good amp better. Or will it? I have read reviews where a very simple amp was deemed one of the best around (this was quite a few years ago), and I was astonished when I finally saw the circuit - it was almost identical to the "El Cheapo" amplifier (see the projects pages for more info on this amp).

The only major difference between this amp and most of the others at the time was the comparatively low open loop gain, and a somewhat wider bandwidth than was typical at the time, because it does not need a miller capacitor for stability. So the amp was better in one respect, worse in another.

In the end, it doesn't really matter what the open loop response is like, as long as closed loop (i.e. with feedback applied) performance does not degrade the sound. Again, we have the same quandary as before - unless we can monitor the difference between input and output at all levels and with normal signal applied, we really don't know what is going on. The usual tests are useful, but cannot predict how an amp will sound. I have heard countless stories about amps that measure up extremely well, but sound "hard and dry", and have no "music" in them.

Unless these measurements are made (or at least some modified form), we will still be no further in

understanding why so many people prefer one brand of amp over another (other than peer pressure or advertising hype).

One possibility is to measure the amp with a gain of 40dB. This is an easy enough modification to make for testing, and the performance is far easier to measure than if we attempt open loop testing. The difference between measured performance at 30dB gain (about 32) versus 40dB (100) would be an excellent indicator of the amp's performance, and it is not too hard to predict the approximate open loop response from the different measurements. To be able to do this requires that all measurements be very accurate.

Would these results have any correlation with the review results? We will never know if someone doesn't try it - work the techniques discussed here thoroughly, with a number of different amps. It would be useful to ensure that the reviewer was unaware of the test results before listening, to guard against experimenter expectancy or sub-conscious prejudice.

It is very hard to do a synopsis of this topic, since I have too little data to work with. Only by adopting new ideas and test methods will we be able to determine if the "golden-ear" brigade really does have golden ears, or that they actually hear much the same "stuff" as the rest of us, but have a better vocabulary. That is not intended as a slur, just a comment that we have to find out if there is anything happening that we (the "engineering" types) don't know about, or not. Unless we can get a match between measured and described performance, we get nowhere (which is to say that we stay where we are, on opposite sides of the fence).

Speaker - Amplifier Interface

Many is the claim that the ear is one of the most finely tuned and sensitive measuring instrument known. I am not going to dispute this - not so that I will not offend anyone (I seem to have done this many times already), but because in some respects it is true. Having said that, I must also point out that although extremely sensitive, the ear (or to be more correct, the brain) is also easily fooled. We can imagine that we can hear things that absolutely do not exist, and can just as easily imagine that one amplifier sounds better than another, only to discover that the reverse is true under different circumstances.

Could it be the influence of speaker cables, or even loudspeakers themselves? This is quite possible, since when amps are reviewed it is generally with the reviewer's favourite speaker and lead combination. This might suit one amplifier perfectly, while the capacitance and inductance of the cable might cause minute instabilities in other otherwise perfectly good amplifiers. Although it a fine theory to suggest that a speaker lead should not affect the performance of a well designed amplifier, there are likely to be some combinations of cable characteristics that simply freak out some amps. Likewise, some amps just might not like the impedance presented by some loudspeakers - this is an area that has been the subject of many studies, and entire amplifiers have been designed specifically to combat these very problems [1].

Many published designs never get the chance of a review, at least not in the same sense as a manufactured amplifier, so it can be difficult (if not impossible) to make worthwhile comparisons. In addition, we sometimes have different reviewers making contradictory remarks about the same amp. Some might think it is wonderful, while others are less enthusiastic. Is this because of different speakers, cables, or some other influence? The answer (of course) is that we have no idea.

We come back to the same problem I described earlier, which is that the standard tests are not necessarily appropriate. A frequency response graph showing that an amp is ruler flat from DC to daylight is of absolutely no use if everyone says that the highs are "veiled", or that imaging is poor. Compare this with another amp that is also ruler flat, and (nearly) everyone agrees that the highs are detailed, transparent, and that imaging is

superb.

We need to employ different testing methodologies to see if there is a way to determine from bench (i.e. objective) testing, what a listening (i.e. subjective) test might reveal. This is a daunting task, but is one that must be sought vigorously if we are to learn the secrets of amplifier sound. It is there - we just don't know where to look, or what to look for ... yet. Until we have correlation between the two testing methods, we are at the mercy of the purveyors of amplifier snake oil and other magic potions.

The SIM distortion indicator is one possible method that might help us, but it may also react to the wrong stimulus. Perhaps we need to add the ability to detect small amounts of high frequencies with greater sensitivity, but now a simple idea becomes quite complex, possibly to no avail. It is also important that such a device has zero effect on the incoming signal itself, so some care is needed to ensure that there is negligible loading on the source preamplifier.

This is not the only avenue open to us to correlate subjective versus objective testing. Both are important, the problem is that one is purely concerned with the way an amplifier behaves on the test bench, and a whole series of more or less identical results can be expected. The other is veiled in "reviewer speak", and although it might be useful if the reviewer is known and trusted, is not measurable or repeatable. The whole object is to try to determine what physical factors cause amplifiers to sound different, despite that fact that conventional testing indicates that they should sound the same.

Impedance

The output impedance of any amplifier is finite. There is no such thing as an amplifier with zero output impedance, so all amps are influenced to some degree by the load. An ideal load is perfectly resistive, and has no reactive elements (inductance or capacitance) at all. Just as there is no such thing as a perfect amplifier, there is also no such thing as a perfect load. Speakers are especially gruesome in this respect, having significant reactance, which varies with frequency.

A genuine zero impedance source is completely unaffected by the load, and it does not matter if it is reactive or not. If such a source were to be connected to a loudspeaker load, the influence of the load will be zero, regardless of frequency, load impedance variations, or anything else.

Since this is not the case in the real world, the goal (or at least one of them) is generally to make the amplifier have the lowest output impedance possible, in the somewhat futile hope that the amp will not be adversely affected by the variable load impedance. In essence, this is futile, since there will always be some output impedance, and therefore the load will always have some influence on the behaviour of the amp.

Another approach might be to make the output impedance infinite, and again, the load will have zero effect on the amplifier itself. Alas, this too is impossible. Given that the conventional approaches obviously cannot work, we are faced with the problem that all amplifiers are affected by the load, and therefore all amplifiers must show some degree of sensitivity to the speaker lead and speaker.

The biggest problem is that no-one really knows what an amplifier will do when a reactive load reflects some of the power back into the amp's output. We can hope (without success) that the effects will be negligible, or we can try to make speakers appear as pure resistance (again, without success).

A test method already exists for this, and uses one channel of an amp to drive a signal back into the output of another. The passive amplifier is the one under test. It is also possible to use a different source amplifier

altogether, since there is no need for it to be identical to the test amp. Use of a "standard" amplifier whose characteristics are well known is useful, since the source will be a constant in all tests. Differences may then be seen clearly from one test to the next.

The method is shown in Figure 6, and is a useful test of the behaviour of an amp when a signal is driven into its output. This is exactly what speakers do - the reactive part of the loudspeaker impedance causes some of the power to be "reflected" back into the amplifier. Since one amplifier in this test is the source, the device under test can be considered a "sink"

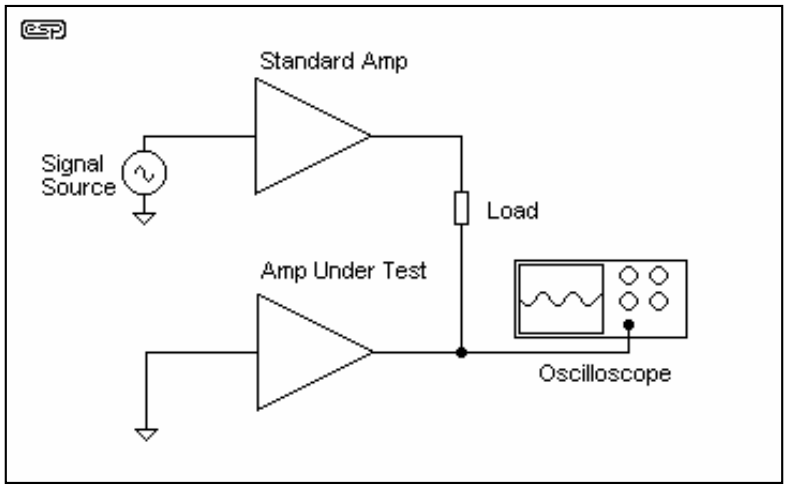


Figure 6 - Amplifier Power Sink Test

I have used this test, and although it does show some interesting results, they are essentially not useful unless used as a comparative test method. The driven amplifier is also subjected to very high dissipation (well above that expected with any loudspeaker load), because the transistors are expected to 'dump' the full current, while they have the full rail voltage across them. There is a real risk of damaging the amplifier, and I suggest that you don't try this unless you are very sure of the driven amplifier's abilities.

We may now ask "Why is this not a standard test for amplifiers, then?". The answer is that no-one has really thought about it enough to decide that this will (or should) be part of the standard set of tests for objective testing of an amplifier. The results might be quite revealing, showing a signal that may be non-linear (i.e. distorted), or perhaps showing a wide variation in measured signal versus frequency.

If either (or both) is the case, then we have some more information about the amplifier that was previously unknown. It might be possible to correlate this with subjective assessments of the amp, and gain further understanding of why some amps sound better (?) than others. We might discover that amps with certain characteristics using this test are subjectively judged as sounding better than others.

If this test became standard, and was routinely allied with the SIM tester described above, we might become aware of many of the problems that currently are audible, but for which there is no known measurement technique.

Conclusion

This article has described some tests that although not new, are possibly the answer to so many questions we have about amplifiers. The tests themselves have been known for some time, but their application is potentially of benefit for all of us. We may be able to finally perform an objective test, and be able to predict with a degree of confidence how the amp will sound. It may also happen that these tests are not sufficient to reveal all the

subtleties of amplifier sound, but will certainly be more useful than a simple frequency response and distortion test.

Any change to the testing methods used is not going to happen overnight, and nor are we going to be able to see immediately which problems cause a difference, and which ones have little apparent effect. Time, patience, and careful correlation of the data are essential if this is to succeed. There are laws of physics, and there are ears. Somewhere the two must meet in common ground. We already know that this happens, since there are amplifiers that sound excellent - according to a large number of owners, reviewers, etc. - now we need to know why.

There is a test method (or a series of methods) that will allow us to obtain a suite of tests that makes sense to designers and listeners alike, so we can get closer to the ideal amplifier, namely the mythical "straight wire with gain", but from the listener perspective rather than the senseless repetition of tests that seem to have no bearing on the perceived quality of the amp. This is not to say that the standard tests are redundant (far from it), but they do not seem to reveal enough information.

For this to succeed, the subjectivists must be convinced, as must the "objectivists". We are all looking for the same thing - the flawless reproduction of sound - but the two camps have drifted further and further apart over the years.

These are my musings, and I am open to suggestions for other testing methods that may reveal the subtle differences that undeniably exist between amplifiers. At the moment we have a chasm between those who can (or think they can) hear the difference between a valve and an opamp, a bipolar junction transistor and a MOSFET, or Brand "A" versus Brand "B", and those who claim that there is no difference at all.

The fact that there are differences is obvious. The degree of difference and why there are differences is not. It would be nice for all lovers of music (and the accurate reproduction of same) if we can arrive at a mutually agreeable explanation for these differences, that is accurate, repeatable, and measurable.

If these criteria are not met, then the assessment is not useful to either camp, and the chasm will simply widen. This is bad news for all - it is high time we all get together and stop arguing amongst ourselves whether (for example) it is better to use one brand of capacitor in the signal path or another.

These testing methods can also be applied to the measurement of individual components, speaker cables, interconnects and preamps, particularly the SIM tester. Using the amplifier power sink test with different cables and speakers might give us some clues as to why so many people are adamant that one speaker cable sounds better than another, even though there is no measurable difference using conventional means.

The greatest benefit of these tests is that they will reveal things we have not been looking at (or for) in the past, and may show differences that come as a very great surprise to designers and listeners alike.

For information on the use of the SIM, and an initial article describing how it works and my results so far, please see "Sound Impairment Monitor - The Answer?".

References

1. Douglas Self - Blameless Amplifier, Electronics World (Refer to The Self Site)

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 26 Feb 2000 Updates - 23 dec - added some more data, and square/sine test info./ 27 Feb, moved SIM description to it's own page.

Elliott Sound Products **Class A Amplifiers - A Brief Explanation**

Copyright © 1999-2005 Rod Elliott
Page Last Updated 02 Apr 2005

Main Index

Articles Index

Class-A Part 2

Introduction

Recently there has been a resurgence of two "ancient" technologies - vacuum tube (valve) amplifiers and Class-A systems. The big question is .... is there a difference? This discussion centres on the Class-A amplifier, and explains (or attempts to) how it is different from a conventional power amplifier.

Why would someone want to build or buy an amplifier which is sooo inefficient? A Class-A power amp will typically draw anything from 1/2 to about 1½ times the peak speaker current in its quiescent state (i.e. while it is just sitting there doing nothing).

To put this into perspective, for a measly 8 Watts into 8 Ohms, the RMS current is 1 Amp. The peak current is just over 1.4 Amps, so a typical 8 Watt Class-A amp will draw anything from 700mA to 2 Amps continuous. This equates to a quiescent (no signal) power dissipation of between 17 Watts and 48 Watts, based on a 24 Volt supply (+/- 12 Volts). At very best, such an amplifier will have an efficiency of less than 35% at full power - at worst, this will be perhaps 15% or less.

The basic premise of a Class-A amp is that the output device(s) shall conduct all the time (through 360 degrees of the signal waveform). This means that in the simplest form, the power devices must conduct a continuous current which exceeds the

maximum peak load (loudspeaker) current. If we use a power level of 20 Watts (hardly a powerhouse) for all further calculations, we can see the whole picture.

In contrast, a typical Class-AB power amplifier's output devices only conduct for about 182 degrees (at full power), which means that for much of the signal's duration, only one or the other device is conducting. The other is turned off. The "crossover distortion" so often referred to is nothing to do with the frequency divider in the speaker system, but is created as the signal "crosses over" the 0 Volt point (see Figure 3).

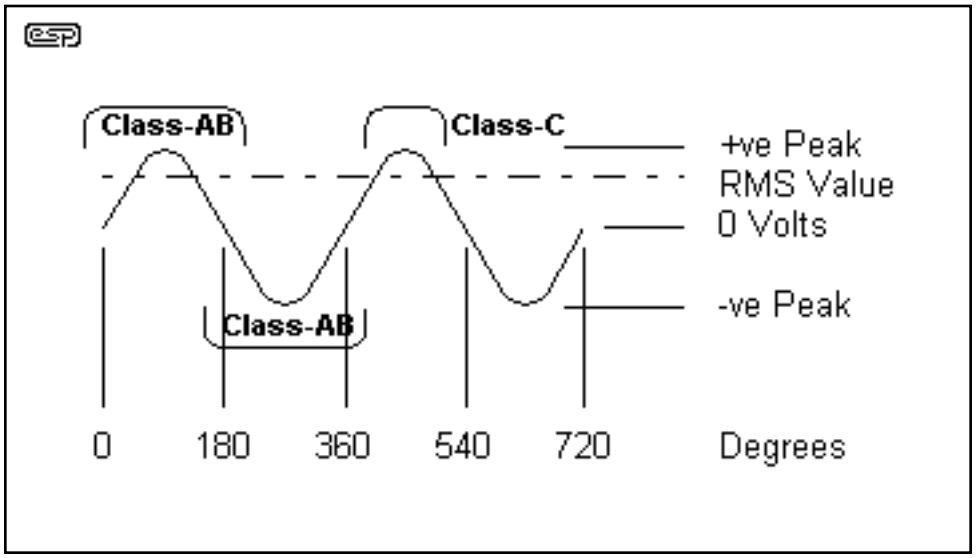


Figure 1 - The Sinewave Cycle

Let's have a quick look at some of the power amp "classes", so we have all the info:

- **Class-A** Output device(s) conduct through 360 degrees of input cycle (never switch off) - A single output device is possible. The device conducts for the entire waveform in Figure 1
- **Class-B** Output devices conduct for 180 degrees (1/2 of input cycle) - for audio, two output devices in "push-pull" must be used (see Class-AB)
- **Class-AB** Halfway (or partway) between the above two examples (181 to 200 degrees typical) - also requires push-pull operation for audio. The conduction for each output device is shown in Figure 1.
- **Class-C** Output device(s) conduct for less than 180 degrees (100 to 150 degrees typical) - Radio Frequencies only - cannot be used for audio! This is the sound heard when one of the output devices goes open circuit in an audio amp! See Figure 1, showing the time the output device conducts (single-ended operation is assumed, and yes this does work for RF).

When I first wrote this article, I had completely forgotten about the Quad "Current-Dumping" amp, which uses a low power "good" amplifier, with a push-pull Class-C type amp to supply the high currents needed for high power. Although these enjoyed a brief popularity, they seem to have faded away. I was reminded of their existence by an article by Douglas Self ("Class Distinction", in the March 1999 issue of Electronics World), in which he quite rightly points out that the current-dumper is (at least in part) Class-C.

- **Class-D** Quasi-digital amplification. Uses pulse-width-modulation of a high frequency (square wave) carrier to reproduce the audio signal - although my original comments were valid when this was written, there have been some very significant advances since then. There are some very good sounding Class-D amplifiers being made now, and they are worthy of an article of their own.

There are many amplifier topologies which I have not mentioned above, mainly because most of them are either too bizarre, not worth commenting on, or are too complex to explain simply. Of these, Class-G and Class-H use power supply switching and modulation (respectively). This provides greater than normal efficiency and lower dissipation, but both are essentially Class-AB designs.

Although many audio amps may be called Class-B, generally they are not. Virtually without exception they are Class-AB, although most will be at the bottom end (conduction for 181 degrees for each device). Most power amps operate in Class-A up to about 5 to 10mW, after which they become Class-B.

In the device department - For the remainder of this paper, I shall use bipolar transistors for the power devices, since they exhibit highly desirable characteristics for this application. They are also far more linear than MOSFETs, and some of the newer bipolar devices are outstanding in this regard. Note that there are two types of MOSFET in common use - Lateral devices are designed for audio, and although less linear than bipolar transistors can make a very good amp indeed (see Project 101). Power switching MOSFETs are (IMHO) not suitable for use in audio except where very high power is needed and extreme linearity is not required.

Power	20W (continuous)
Load Voltage (at Speaker)	12.65 Volts RMS (17.9 Volts Peak)
Load Current (through Speaker)	1.58 Amps RMS (2.23 Amps Peak)
Supply Voltage	+/- 20 Volts (constant)

Supply Current	+/- 2.25 Amps (peak)

Table 1 - Class-A Amplifier Requirements (Approx.) - 8Ω Load

In amplifier design, we are interested in the peak voltage and current, since if these are not met, then the required RMS values cannot be achieved. The ratio of RMS to peak (for a sine wave) is the square root of 2 (1.414), so RMS values must be multiplied by this constant to derive the peak values of voltage and current. (Refer to Figure 1 to see the relationship between peak and RMS voltages.)

This is how the values in the table were determined. The supply voltage needs to be slightly higher than the actual speaker peak voltage because the output devices (transistors) are not perfect, and some voltage will be lost even when they are turned on fully. (If MOSFETs were to be used, the losses may be much greater, unless an additional power supply is employed.)

Ok. We have determined that the peak speaker current is 2.25 Amps, so in the simplest of Class-A designs this will require a quiescent current of 2.25 Amps. Given that the voltage is +/- 20 Volts, this means that the power output stage will have to dissipate 40 x 2.25 = 90 Watts (45 Watts per output device).

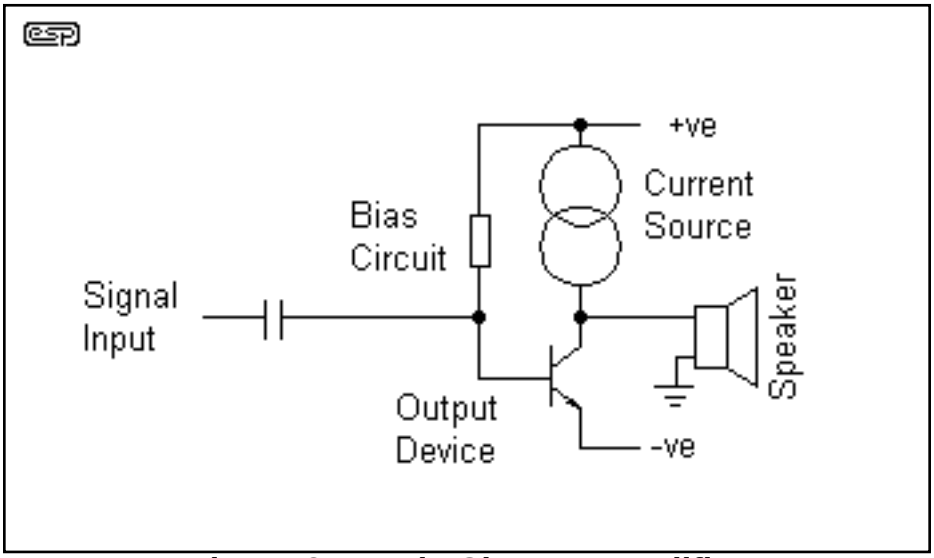


Figure 2 - Basic Class-A Amplifier

Figure 2 shows what a simple Class-A amp looks like. The current source is a simple circuit, which will provide a current which remains constant regardless of the load placed at its output. The output transistor "dumps" any current which is not needed by the load (speaker), so when it is completely turned off, all the current source output flows through the speaker. Conversely, when it is turned on, the speaker current flows through the output transistor (as well as the current from the current source!), so its current will vary

from almost 0 Amps, to a maximum of 4.5 Amps for our example. When there is no input signal, the output transistor's current must exactly equal the output of the current source. If it does not, then the difference will flow through the speaker. It is allowable (generally speaking) for an absolute maximum of 100 mV DC to be present across the speaker terminals - this equates to 1.67 mW of DC for an 8 Ohm system, assuming a 6 Ohm DC resistance for the voice coil. (Power = (V x V) / Impedance)

This simple model is not really appropriate for general use, since it wastes far too much power, although many Class-A amps still use this principle. The next step is to operate the current source at about 1/2 the speaker's peak current, and modulate its current output to ensure that both current source and power amplifier output device conduct during the entire signal cycle, but are able to vary their current in an appropriate manner. This improves efficiency (which remains dreadful, but slightly less so), and lowers the quiescent dissipation to more manageable levels.

The simple Class-A amplifier described by John L Linsley-Hood and the very similar looking Death of Zen (DoZ) amp on these pages use this latter approach, and it is a sensible variant of the various Class-A designs. As an example, the amplifier will only (?) need to dissipate about 50 Watts when idle, since the quiescent current is reduced to around 1.2 Amps.

Another version of the Class-A amp looks exactly the same as a standard Class-AB (Class-B) power amp, except the quiescent current is increased to just over 1/2 of the peak speaker current. This is thought by some (including me up until I was shown the error of my ways) that this is not a "real" Class-A amplifier. It is real Class-A, and is best described as push-pull (as opposed to single ended) operation. If the bias current is not high enough for the actual reactive speaker load (not some quoted nominal resistive load), it is still possible that one transistor or the other will switch off at some part of the signal cycle. This will happen at a much higher power level than is normally the case, but if this happens, then the amplifier ceases to be true Class-A.

As an extension of the above, it is possible to design an amp that looks remarkably like a conventional Class-AB amp, but with additional circuitry is biased in such a way that the output transistors do not turn off - ever. This technique can also be used with Class-AB, and supposedly reduces crossover distortion. I have not used this method, since in my experience the crossover distortion in a well designed output stage should be sufficiently low that the additional complexity is not warranted.

The last three "variants" cause the current to be modulated in each supply rail, so there is not the steady state current one expects from a Class-A amp, but a waveform that

varies with the signal. When properly designed and biased, the output devices conduct at all times, but the power supply has to contend with a varying load. I have not investigated this fully, but it makes the design of the supply a little more difficult (or simpler in some ways) because of the varying load current. Tests I have performed with the DoZ amp do not show any audible effect on the sound quality - provided the supply is designed to handle the variations without any problems.

Actually, the idea that a Class-A amp draws a continuous steady current from the supply is true *in one case only*. A single ended amp using a current source as the collector load will draw a continuous steady current - but ***only*** if it uses a single supply. In the case of a dual supply, the same amp will draw a continuous current from one supply, and a varying current from the other. (My thanks to Geoff Moss for pointing this out - a detail that few published designs have ever mentioned !)

An amp that uses a fixed current source of (say) 2.5A from the positive supply will draw 2.5A regardless of load or signal level, but only from the positive supply. The negative supply current will vary from 2.5A at no signal, but will be almost zero at maximum positive swing, when the lower transistor is turned off, and the current flows from the current source to the load. At maximum negative signal swing, the negative supply current will be close to double the quiescent current, since the lower transistor now carries the current from both the load *and* the current source.

This "small" detail seems to have received scant reference in any of the articles I have read, but it will make a very big difference to the power supply. In this respect, I do not feel that the single ended version should be operated from a dual supply. If it is so important to you to eliminate the coupling capacitor, then I suggest that either a push-pull Class-A design be used, or build separate power supplies for each polarity.

There is some evidence (I refer again to Doug Self) to indicate that the distortion of a "true" Class-AB amp will often be worse than that of a Class-B design, since the switching transients are larger due to the output devices' higher gain at moderate (0.5A to 1.5A) currents. I have not been able to verify this, and the tests I have done indicate that there are definite benefits in the higher quiescent currents, provided the current is chosen reasonably carefully.

One of the biggest problems with Class-A amps is that the simple power supply used with conventional Class-AB amps is usually no good to us. The reason is that the AC ripple on the DC power rails is injected into the amp, and emerges as hum (at 120 or 100Hz, depending on location - US or elsewhere, respectively). The magnitude of this ripple is far greater than with a Class-AB amp, because a considerable amount of

current is being drawn at all times, rather than during signal peaks (etc). A power supply which provides a no-load ripple of perhaps 50 mV for a Class-AB amp may have 1 Volt (or more) of ripple at a current of 1.2 Amps. This will be audible at low signal levels.

Adding capacitance helps, but by the time the ripple is reduced to a reasonable level, you have sold the car to pay for the capacitors, and no longer have a vehicle to carry them home in. You will need a ridiculous amount of capacitance to obtain reasonable hum levels (> 70dB signal to noise ratio) unless a regulated supply is used. The fact is that many Class-A power amps do not have particularly good power supply rejection (Ok, it is not generally too bad, but cannot compete with the likes of an operational amplifier), and a regulated power supply is recommended for all such amps. In case you were wondering, that does indeed mean that you need more transistors, more heatsinks, and it will cost more money. Such is the price we pay for perfection.

There is an alternative (which I have not tried for this application, but have carried out numerous spice simulations) called a capacitance multiplier, which is simpler and cheaper than a regulated supply, but should be capable of reducing the ripple to very low levels. I have had a few e-mails from readers who have built the capacitance multiplier project (see the Projects page), and the results have been very positive, so this makes the Class-A idea far more attractive from a cost and heat perspective. (Capacitance multipliers are not required to regulate, so operate with a much lower input to output differential voltage - therefore, less heat!) Indeed, the design by John Linsley-Hood referenced on these pages uses a capacitance multiplier, although its performance can be improved dramatically.

- John L Linsley-Hood's Class-A amplifier design
- Simple Capacitance Multiplier Power Supply

Decisions, Decisions

The question now is - is this *really* what I want to do? The answer might be a resounding **yes** (after all, there is no good reason that a Class-AB amp cannot be just as good) - but to be sensible, we should apply the Class-A amp for the tweeters in the system, and use conventional Class-AB amps for the low and mid frequencies. To obtain adequate sound pressure levels, most modern speakers need lots of power, since they are not very efficient (i.e. electrical power in versus acoustical power out).

Rather than extend this page to a short text book on the subject, I shall leave you with a

simplified model, which I produced for a reader who had a speaker system which was even less efficient than is common. The table shows the power needed to achieve various peak SPLs (at one metre) for a speaker with an efficiency of 85dB/m/W.

Based on the sensitivity of these speakers, the following shows roughly what you can expect, based on a single amplifier for clarity (i.e. not bi-amped or tri-amped):

dB SPL at 1 metre	Amp Power, Watts, one channel
85	1
88	2
91	4
94	8
97	16
100	32
103	64
106	128
109	256
112	512

Table 2 - Power Vs. SPL

This is not good news for the most part, as it clearly shows that vast amounts of power are needed to achieve a realistic SPL in a typical listening environment. Remember that the figures shown are at a distance of only one metre - the SPL will fall by a further 6dB each time the distance is doubled. (Mind you this is a theoretical figure, which is generally not met in practice - perhaps 5dB would be closer to the truth?)

Realistic SPL in this context is worthy of a page (book?) in itself, but remember that for an average SPL of (say) 85dB, transients will require between 10 and 20dB of headroom. This means that the peak power needed will be between 10 and 100 times the power needed to reproduce the average of 85 dB. At a distance of 2 metres, something around 3 Watts will be needed for this example. To reproduce the transients, the actual power needed must be between 30 and 300 Watts!

In case you were wondering, 85dB SPL is not loud (although "s/he who must be obeyed" will almost certainly disagree). In fact, it is only marginally louder (by about 5dB) than the recognised optimum level for normal speech.

Class-A Benefits

Since Class-A amps are inefficient, generate lots of heat, and require a far more complex power supply than conventional Class-AB amplifiers, there have to be some compelling reasons to use this arrangement. The first is circuit simplicity. In the light of the above discussion, the circuit is not simple, but for the audio signal it can be far less complex than for a conventional power amp.

The benefit of this is that the signal is subjected to comparatively little amplification, resulting in an open loop (i.e. without feedback) gain which is generally fairly low - probably less than 250 (48dB), and possibly as low as 50 or so (34dB). This means that very little overall feedback is used, so stability and phase should be excellent over the audio frequencies. A well designed Class-A amplifier should not require any frequency compensation (or very little), so the open loop gain will remain reasonably constant over the audio range. This results in superior transient response, and dramatically reduced "Transient Intermodulation Distortion" (or TID, aka Dynamic Intermodulation Distortion), which is thought by many designers to be caused by phase and time delays between the input and feedback signals. It may be possible that this is the cause, although the existence of TID is virtually zero in any competently designed amp.

The simple fact is that the more amplifying devices that are introduced into the chain, the more phase shift must be introduced. No amplifying device is capable of responding instantaneously to a change of input - all have some inherent delay (which usually includes different turn-on and turn-off times). With fewer devices in the audio circuit, there must be less delay between a change in the input causing a change in the output. The simplified topology used for most Class-A amps can also be used with Class-AB - often with very good results indeed.

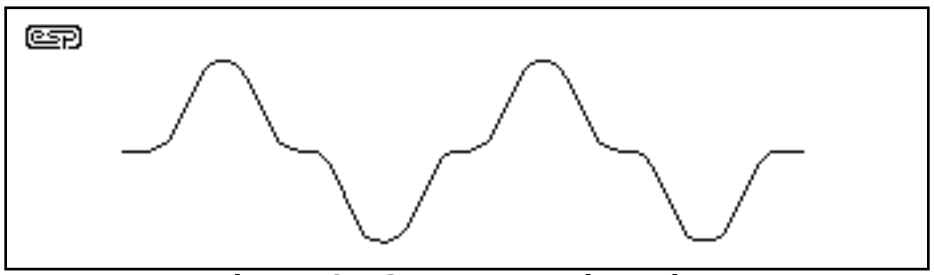


Figure 3 - Crossover Distortion

Figure 3 shows the crossover distortion of a Class-B type amplifier. This is exaggerated for clarity, and the "soft" crossover distortion shown is actually more typical of a push-pull valve (vacuum tube) or lateral MOSFET amplifier (without proper bias adjustment). As can be seen, when the signal is reduced, the ratio of distortion to signal will become

much worse, resulting in an increase in distortion as power is reduced. Indeed, this is exactly what happens in many amplifiers, but it generally is "swamped" by so much feedback that it seems to disappear. It can be seen from the diagram that for this crossover distortion to appear, the amplifier's gain must fall as the signal level approaches 0 Volts.

The point that distortion "seems to disappear" is the operative term here - it does not go away at all, and worse, as the crossover point is reached, the open loop gain of the amplifier is reduced, meaning that there is not as much feedback as at higher signal levels. This will be apparent to readers with an electronics background - note that near the crossover point, the amplitude of the signal is much lower than it should be (this is what causes the problem in the first place!). Since the amplitude is reduced, it is obvious that the amplifier's gain must be lower at this level than at higher levels.

Therefore, if the open loop gain is lower, then the available feedback must also be lower. This is an area that has received some study, and this is illustrated by the very 'flat' gain vs. collector current curves of many of the more desirable audio output transistors. It is certainly a cause for some concern, and indicates that the open loop behaviour of a power amp should minimise crossover distortion before any feedback is added. Simply increasing the quiescent current is not always a complete answer, because this problem is created by the inherent non-linearity of the output devices as they commence (or cease) conduction. Increasing quiescent current will move the "kink" further away from the 0 Volt point, but it will still be there - and may actually be worse than at lower quiescent currents. A major advantage is that the distortion components will be (potentially) somewhat less audible, and will affect the signal while it is comparatively loud - this will reduce its audibility further.

I bagged MOSFETs earlier in this article because they are actually more non-linear in this region than transistors. Since this is the most critical part of the signal, it is important that it is treated with the utmost respect.

This does not mean that MOSFETs are not capable of exemplary performance. A carefully designed MOSFET amp will sound every bit as good as a bipolar amp, whether it is operated in Class-A or Class-AB.

In the light of this, it is a wonder that any Class-AB (conventional) power amplifiers sound any good at all. Historically, it is exactly the problems I have highlighted here which created the term "transistor sound" (used in a derogatory sense of course) when transistor or "solid state" amplifiers first appeared. Despite anything you may read, these problems are caused by the physical and electrical characteristics of transistors, and have never gone away. New devices are far more linear than those of the 60s and 70s,

but they are not perfect. Operation at higher quiescent currents (i.e. more into the Class-AB region) will reduce the non-linearity at crossover, but it can never be eliminated altogether - at least not with any devices currently available.

It is fair to say that although the problem cannot be eliminated, the effects can be reduced to such an extent that many amplifiers have almost unmeasurably small levels of crossover distortion. It is not at all uncommon that to be able to see the distortion residual (after the fundamental has been removed with a distortion analyser), it is necessary to use a digital oscilloscope that can apply averaging. The distortion is buried below the amplifier noise floor, and is not visible without the averaging feature. In tests I have performed, listening to the residual noise + distortion reveals that the distortion component (in isolation) is barely audible over the system noise - itself normally below audibility with typical loudspeakers.

So, it is entirely possible to design an amplifier whose distortion at any level below clipping is virtually unmeasurable. Marginally higher levels are commonplace, and it is thought by many that the typical distortion level in most well designed power amplifiers is inaudible under most listening conditions. There are (of course) others who deny this - either because they have done proper comparisons under controlled conditions, because they have hearing that is far more acute than most of us, or because they have been told that they *must* be able to hear the difference - if they can't, they must have 'tin ears'. Nothing like a bit of peer group pressure to influence one's perceptions.

Where does this leave Class-A? There is an emotional connection with the idea of a Class-A amp, and it has to be considered that sometimes there is simply a 'feel good' aspect to this - technicalities don't even enter into it. Despite my own ambivalence, I was still a bit disappointed in my decision not to use P36 for my own tweeters - and this in spite of the fact that I could hear no difference between the P36 and the high quality power opamp which I *am* using for my tweeters.

Because the transistors in a Class-A amplifier are never switched off, there is obviously no crossover distortion (after all, there is no crossover - where one transistor turns off, and the other supplies the load current). There is distortion though - it is caused by all the normal non-linearities in any active device, and in particular the wide current variation in the output device (in combination with elevated temperature). It is worth noting that crossover distortion is exactly the same as clipping distortion, but with a different phase with respect to the signal. Consequently, it contributes odd harmonics (as does clipping) - 3rd, 5th, 7th, etc.

If properly designed, a Class-A amplifier should be capable of a maximum open-loop

distortion of perhaps 5% at full power, reducing as the input signal (and hence output power) is lowered. This distortion is believed to be predominantly 2nd harmonic, which (in moderation) is far less intrusive than the odd-order distortion created by conventional push-pull Class-AB amplifiers, however this may not be the case. In contrast, most common Class-AB amps will have an open loop distortion of perhaps 10% to 15% at full power, although some will be much lower.

Such amps typically rely on global feedback to reduce this distortion, and usually have very high open loop gains. Another problem is that the open-loop gain is not constant with frequency, so the amount of feedback applied is reduced at the higher frequencies - not at all what is really needed. However, it *does not* mean that all such amplifiers are unlistenable - despite claims to the contrary.

For additional comment on Class-A, the "Death of Zen" (DoZ) article may be an interesting read.

Class-A Myth #1
A Class-A amp maintains the same current through the transistors, therefore ensuring that they remain in their most linear region at all times.

This is *not* the case at all - the current varies widely in the output device in the case of a current source amplifier, and it varies widely in *both* output transistors for other types of Class-A amp. While it is possible to make the current reasonably constant, it is neither practical nor sensible to do so.

Class Comparison

As often happens when writing, I suddenly decided that I just had to run a simulation on a pair of output stages. One is Class-AB (essentially the same as that used in Project 3A) and a Class-A emitter follower circuit. Both were operated with zero feedback, and the Class-AB stage was run at a quiescent current of 14mA vs. 2A for the Class-A circuit.

Rather than make this article longer than necessary, if you want to see the details see Class-A Part 2

Conclusion

Class-A is the most desirable of the amplifier configurations from a purist point of view, but is not suited to high power systems unless outrageous power dissipation is

acceptable (like between 825 to 1500 Watts of pure heat, to get 300 Watts of audio). However, if used for the high frequency amplifier in a tri-amplified system, it is possible to obtain the SPL you desire in your listening room, but without having to install a dedicated air-conditioning system to remove the heat generated.

When used for the frequency range of 3000Hz and above, comparatively little power will be needed, and the sonic benefits should be readily apparent - crystal clean highs, without any harsh distortion components. The distortion generated may be (but is not necessarily) predominantly 2nd harmonic, and will be greatest at high power levels where it is least likely to be audible. Bear in mind though, that a great many Class-AB amplifiers will be capable of performance that is just as good, and in a lot of cases, far better.

Class-A Myth #2

Class-A amps give predominantly 2nd order distortion.

They might, or they might not, depending entirely on the topology. A great many Class-A amps will produce distortion components that are almost identical to those produced by a Class-AB amp. This excludes clipping distortion, which should be avoided in *any* class of amplifier used for high quality audio.

An Alternative?

Where it is not feasible (economically or otherwise) to use a Class-A amp in the tweeter frequency range, a modified Class-AB amp could be used. The modification needed is to increase the quiescent current (to perhaps 1 Amp or so) so that the amplifier operates as Class-A for any signal below about 8 Watts - assuming a well behaved 8 ohm load such as a tweeter. Such a modification to an existing amp is quite simple for an experienced electronics engineer or service person, but will almost certainly require that the heatsinks be upgraded to prevent the destruction of the output devices. It is also probable that additional capacitors will be needed for the power supply - and possibly a regulator or capacitance multiplier circuit, too. Without these, the hum level may become intrusive, which rather negates the whole purpose of the exercise. Some basic experimentation is required for anyone thinking along these lines.

Bear in mind that you can say a fond farewell to any warranty which may exist on your amp - few manufacturers will accept that ripping their product to pieces and rebuilding it as something "new" is a perfectly reasonable thing to do.

Despite the cost of modifying an amp in this way, it is bound to be cheaper than buying

or building a Class-A amp from scratch - even more so if you have a perfectly good (but underpowered) amp just lying about waiting to be put to use. For not a lot of work and relatively few dollars, a potentially fine amplifier can be yours.

Warning

Please be aware that the above section is more in the line of "musings" than established fact with full testing. The theory is (more or less) sound, but one cannot predict the exact behaviour of any amp once modified, and I suggest that if any such mods are to be attempted, they should be done with "before and after" measurements to allow proper comparison. Operation at a higher than normal quiescent current may actually degrade performance with some amplifiers.

Main Index

Articles Index

Class-A Part 2

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 1999-2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page last updated - 03 Apr 2005 - several clarifications, and various updates to bring the article up to date with recent transistors and to reflect additional research./ 21 Mar 2001 -amended section on 1/2 current biasing./ 09 May 2000-added some comments on cap-mult supply and more on "modulated current" Class-A amps./ 29 Nov 1999, some minor changes to the wording

The Audio Pages

Elliott Sound Products **Power Amplifier Clipping**

Copyright © 2004 - Rod Elliott (ESP)
Page Created 13 Oct 2004

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Description
  - 2.1 Test Results
  - 2.2 Sources of Asymmetrical Waveforms
  - 2.3 Other Issues
  - 3.0 Conclusion
- 4.0 Acknowledgments & References

1.0 - Introduction

Although it is recognised that a power amplifier should never be subjected to (overload) clipping, it is equally well recognised that it will happen at some stage. This article will explain how different amplifiers have different characteristics in this area, and the explanations that follow may go some way to showing how some amps *can* sound very different from others when subjected to overload.

First, we need to know what overload clipping actually is. An amplifier is said to be clipping when the output signal attempts to exceed the supply voltage. Since the supply voltage defines the absolute maximum peak output voltage from the amp, the signal will

be clipped or 'cut off' if the input signal level is too high. For normal testing purposes, a sinewave is the most common test signal used, but this only tells part of the story.

A great deal of normal programme material is asymmetrical - it has peaks in one direction that are not duplicated in the other. Because of the signal processing used in the recording chain (including the microphone), the signal always has an effective long term DC level of zero volts, and no normal audio recorder or reproducer is expected to be able to handle a DC voltage. Short term deviations from the zero volts DC level are not uncommon, but can hardly be called 'DC' unless they are of long enough duration to cause a subsonic (in this case less than 1Hz) level shift.

Since it is very rare indeed that anything is recorded below about 16Hz, there is no apparent need for any reproducer to go any lower than that. The reality is that preamps and power amps will often have a -3dB frequency that is much lower than expected, with figures as low as 1Hz being common. I normally aim for around 7Hz in my designs, but even there you may find exceptions.

The following article applies to amplifiers used in any audio application - hi-fi, studio monitoring, professional sound reinforcement systems (live sound PA), or general purpose public address as used in buildings. Each application will present its own set of problems - the amplifier may be driving loudspeakers directly (multi-amping), or there may be transformers involved (electrostatic loudspeakers or building PA systems). In some cases tweeters and midrange drivers will be protected by series capacitance, while in other cases they may also be at risk from DC or very low frequency signals.

2.0 - Description

Within The Audio Pages, you will find several descriptions of amplifier overload, clipping recovery times, and various other associated topics (including Why do Tweeters Blow). This article concentrates on something very different, and a specific part of the topic that is not well covered elsewhere. In this case, it is the behaviour of the amplifier when clipping is asymmetrical - not because of the amplifier, but because of the applied signal.

Figure 2.1 shows the amplifier used for these simulations - there is a remarkable similarity to the P3A, however 'ideal' current sources were used where appropriate. This does not change anything in the simulations. Any amplifier from any manufacturer will do exactly the same given the test conditions described below. The behaviour described is not a function of the amplifier topology, only a function of the DC response. Valve (tube) amplifiers may generally be excluded - with reservations (see Conclusion

below).

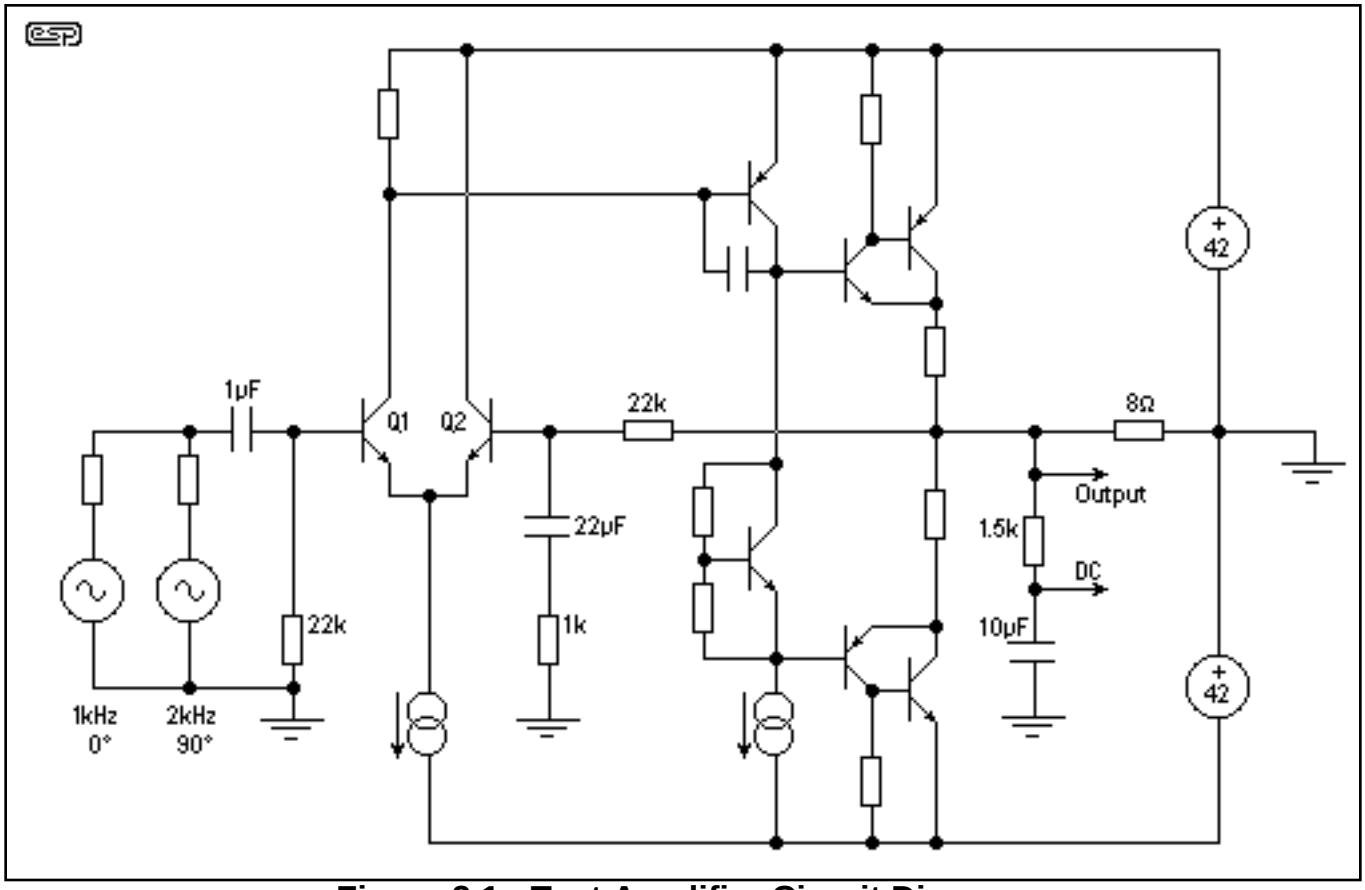


Figure 2.1 - Test Amplifier Circuit Diagram

The important parts of the above diagram are the input coupling cap and resistor, the feedback resistors and capacitor, and the final filter network used to isolate the DC component of the signal. It may also be said that the simulations that follow are by no means the 'worst case' ... any real signal may have far greater effect than that shown.

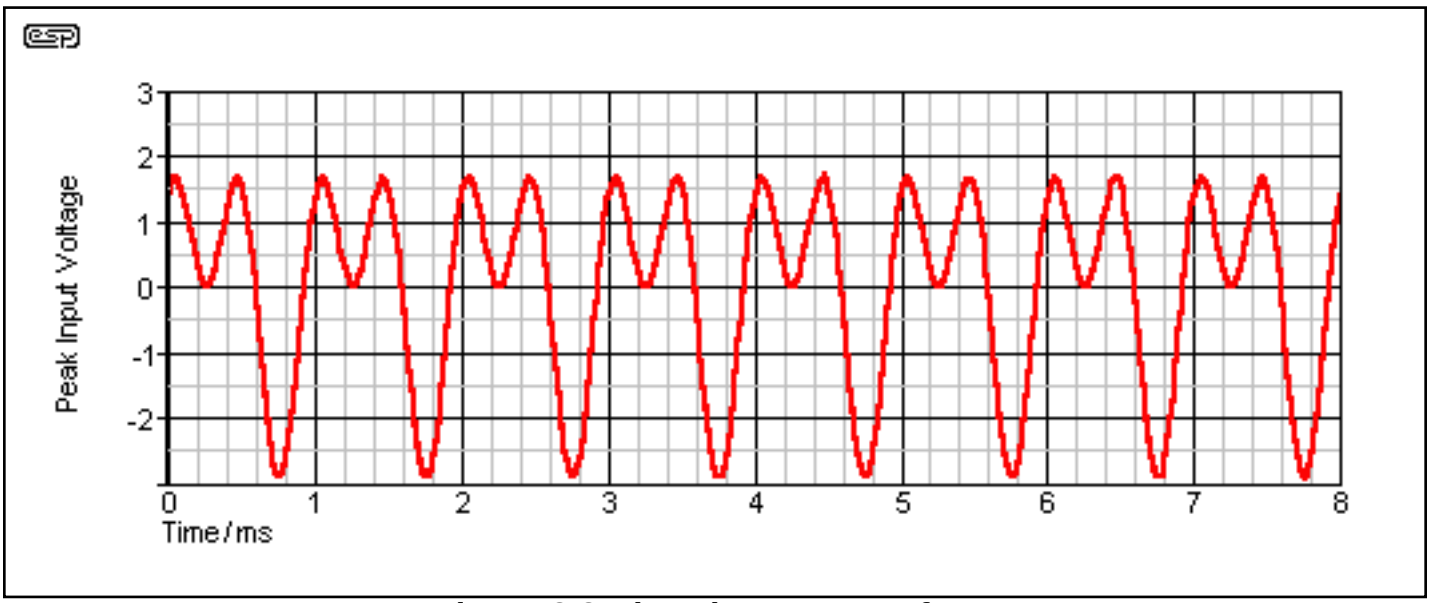


Figure 2.2 - (Raw) Input Waveform

The composite input signal is made up of a 3V peak-to-peak sinewave at 1kHz, added to a 2kHz 3V P-P sinewave whose phase is shifted by 90° to obtain an asymmetrical waveform that has a peak amplitude that is almost twice as great in the negative direction as in the positive direction. The DC component of this waveform is zero - the area of waveform above the zero volt line exactly equals that below, so the net DC must be zero (the simulator claims 3.15mV, but this is insignificant, and is simply the result of the simulator sampling rate).

The input waveform shown above is a bit nasty from the simulator's perspective, so a small delay was introduced to ensure that the signal started at exactly zero Volts. This can be seen in the displays below, where there is a 916.6us gap at the beginning to ensure that the input to the amplifier commenced at the 0V level. To do otherwise skews the results and makes it look a lot worse than it really is. It is also unrealistic, since no music signal can start from other than a zero reference voltage.

Before we continue, it is important to understand just *how* clipping the waveform can introduce a DC voltage. First, look at the basic waveform details (two complete cycles are shown), both before and after it is clipped ...

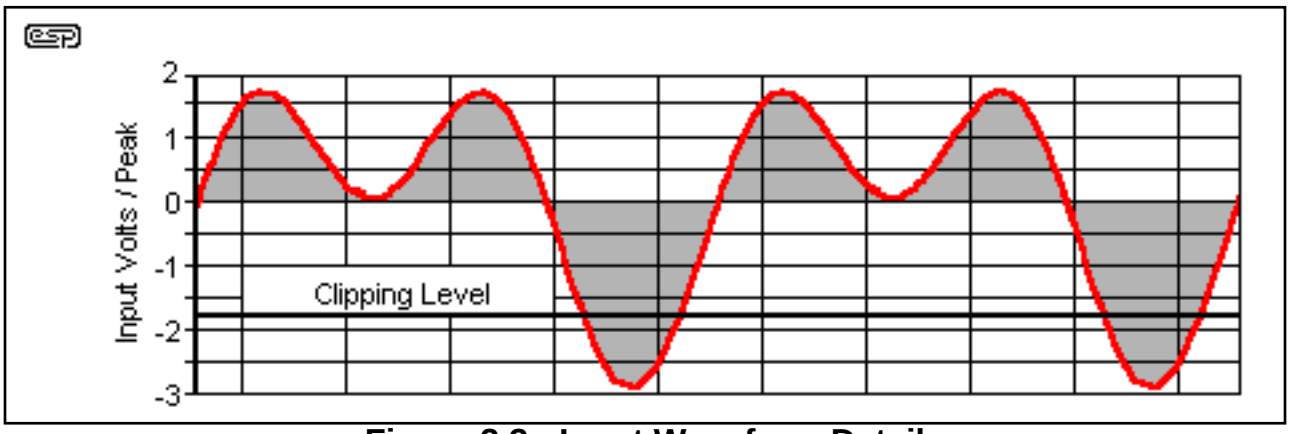


Figure 2.3 - Input Waveform Detail

The total shaded area shows that the area above the zero volt line is exactly equal to the area below, therefore, the signal is perfectly balanced and there is no DC at all. When the amplifier chops off (clips) the section below the line marked 'Clipping Level', the two areas are no longer the same - the area in the lower part of the waveform is smaller because some of it has been removed when the amplifier overloads. Since the upper (positive) section of the waveform now has more area than the lower (negative) section, the overall waveform has a positive bias - this is seen as a DC voltage. No tricks, no special effects, just plain old physics at its most basic level.

Note: - Although it may not *appear* to be the case that the areas above and below the zero volt line are equal, any apparent deviation is the result of conversion of the waveform to an image suitable for publication. Even so, you are welcome to count the pixels if you wish - you will find that the two sections are remarkably similar.

Any amplifier that has full gain at DC will present this instantaneous DC level to the loudspeaker, but it will be shown that an amp that does not extend to DC will eventually average out the DC component, reducing it to a subsonic signal with peak levels that are (hopefully) non-damaging to the loudspeaker. What is *not* shown is the recovery time (after the amplifier stops clipping), where it is understood that a non DC amplifier will produce a subsonic signal that has an equivalent amplitude but opposite polarity of that created when the clipping was present. I shall leave this as an exercise for the reader.

In all descriptions that follow, it is assumed that the amplifier is driving the loudspeaker directly (not using a passive crossover network). While the caps in a passive crossover will save the mid and tweeter drivers from being subjected to the DC, the woofer will still be affected, *regardless of the frequency that clips*.

2.1 - Test Results

Referring to the schematic above, the 1uF capacitor is the input cap. This ensures that any DC applied to the input cannot be amplified, and protects the speakers from DC introduced by the preamp or signal source. The -3dB frequency of the input cap and associated 22k resistor is 7.2Hz.

The 22uF cap is the feedback capacitor - when in place, the amp's gain will also be 3dB down at 7.2Hz. This capacitor ensures that the amplifier can have a maximum gain of unity for a DC input (including offset from the input long-tailed pair). With both capacitors in place, the -3dB frequency is 11Hz, and is about 1dB down at 20Hz.

Note that in the diagrams below, the charts only show the response obtained out to 20ms, but measurements were taken to 100ms of the waveform. The relatively slow rise of the DC component of the output is caused by the integrating filter (1.5k resistor and 10uF capacitor), but this was essential so that the DC levels could be seen easily. In all cases, the red trace shows the clipped output signal, and the green trace shows the DC component of the output. The clipping is easily seen - the bottom of the waveform is flat, and no longer looks like the input signal shown in Figure 2.2.

AC Coupled - When the amp is completely AC coupled (having the input cap and the

feedback cap installed), the clipping waveform shows that the peak DC voltage across the loudspeaker is 1.93V at 20ms, rapidly falling to 540mV at 100ms. While this is not especially desirable, it is considerably better than the alternatives that follow. The AC coupling throughout ensures that significant DC levels can never appear across the voicecoil for any appreciable period of time.

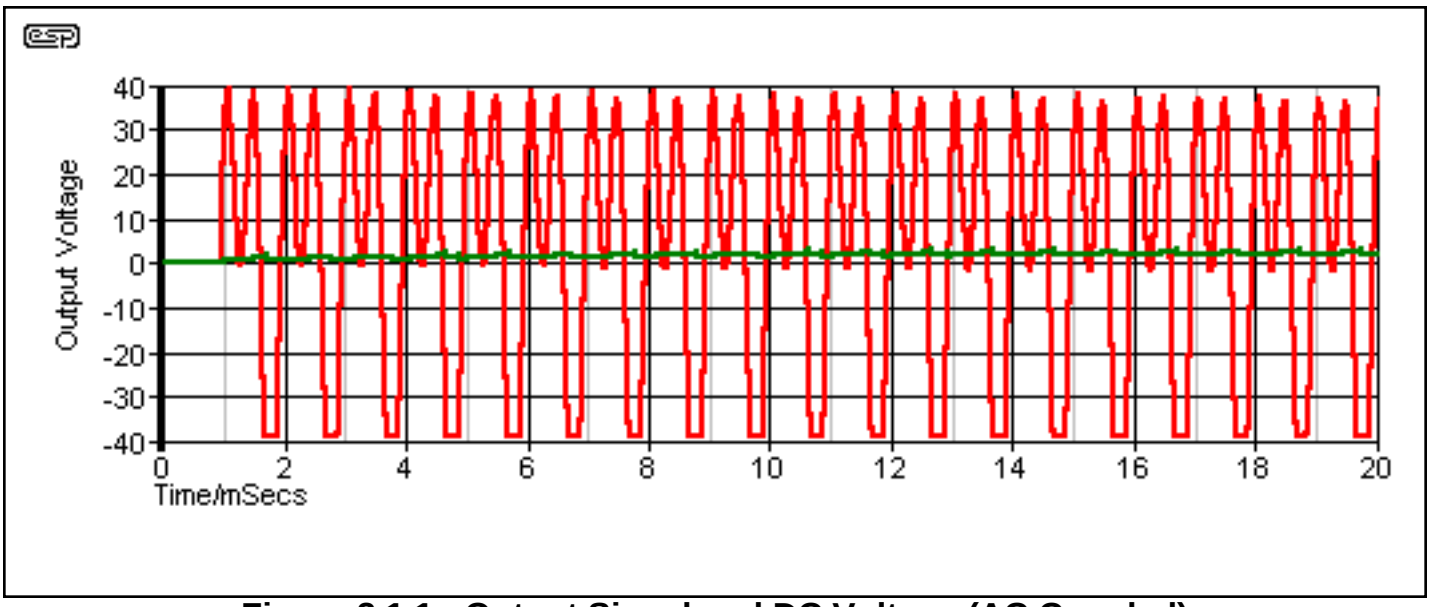


Figure 2.1.1 - Output Signal and DC Voltage (AC Coupled)

With the above, we can expect to see the cone move a little when the signal is first applied at a level sufficient to cause the degree of clipping shown. It will quickly settle back to normal, and the audible effect will be that of a slightly clipped signal only.

AC Coupled (Input Only) - Leaving the input capacitor in place and removing (shorting out) the feedback capacitor, the amplifier is now DC coupled from the input stage onwards. After 20ms, there is 2.7V DC output, rising to 3.8V at 100ms. Note that making the amplifier responsive to DC ensures that there *will* be DC applied to the loudspeaker if the amp is allowed to clip - even briefly.

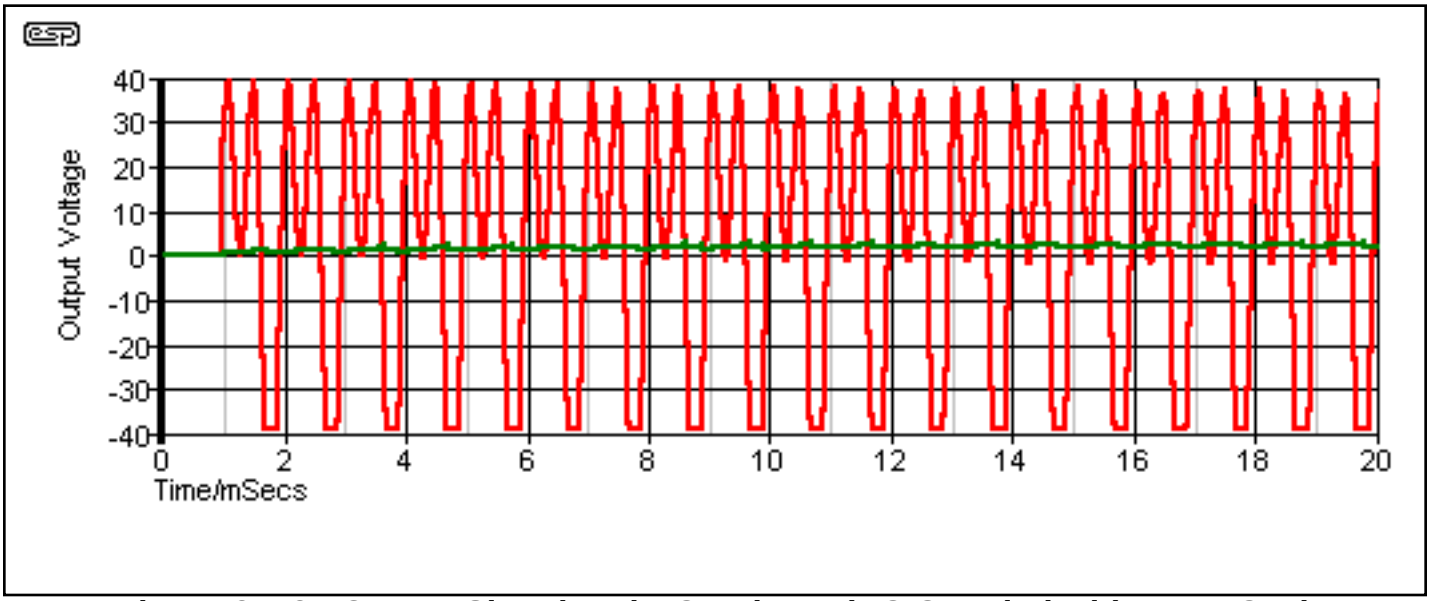


Figure 2.1.2 - Output Signal and DC Voltage (DC Coupled with Input Cap)

In this case, when the signal is applied and the amplifier clips, the cone will be seen to move (outwards if polarity is normal). While nearly 4V DC will cause a problem with a woofer, the effect with a direct coupled midrange (having relatively little cone excursion and using an active crossover) will be a disaster. You can expect the loudspeaker to distort badly - adding to the distortion already created by the amplifier. To explain the reason that using only an input cap will alleviate the problem marginally (compared to the fully DC coupled case) requires a brief explanation ...

When an amplifier clips, it is no longer a linear system for the duration of the clipping, and the feedback circuit is inoperable. The input stage is also no longer linear, and a small amount of rectification of the input signal takes place. *This changes the voltage at the base of the input transistor.* In the case shown above, the average voltage on the base of the input transistor (Q1) changed from around -32mV at idle, to -34mV after 20ms and -35mV after 100ms - nowhere near enough to correct the problem, but noticeable nonetheless. Be aware that this will not occur with all amp input stage topologies - some will never rectify the signal, regardless of input level or output distortion. As is to be expected, the average DC voltage at the base of the feedback transistor (Q2) follows that of the output, but attenuated by the ratio of the feedback resistors (voltage gain at all frequencies is 23).

DC Coupled - With a completely DC coupled amplifier (having neither the input cap nor the feedback cap), the situation is made even worse. The effective DC voltage presented to the loudspeaker load is 3.5V after 20ms, rising to 4.4V at 100ms. Although the voltage with this waveform will not get a great deal worse than measured, there are some waveforms that can easily impress up to 15V DC onto the voicecoil - this is likely to push the coil so far out of the gap that gross intermodulation distortion will occur in the

loudspeaker, adding to the harmonics generated by the amplifier clipping.

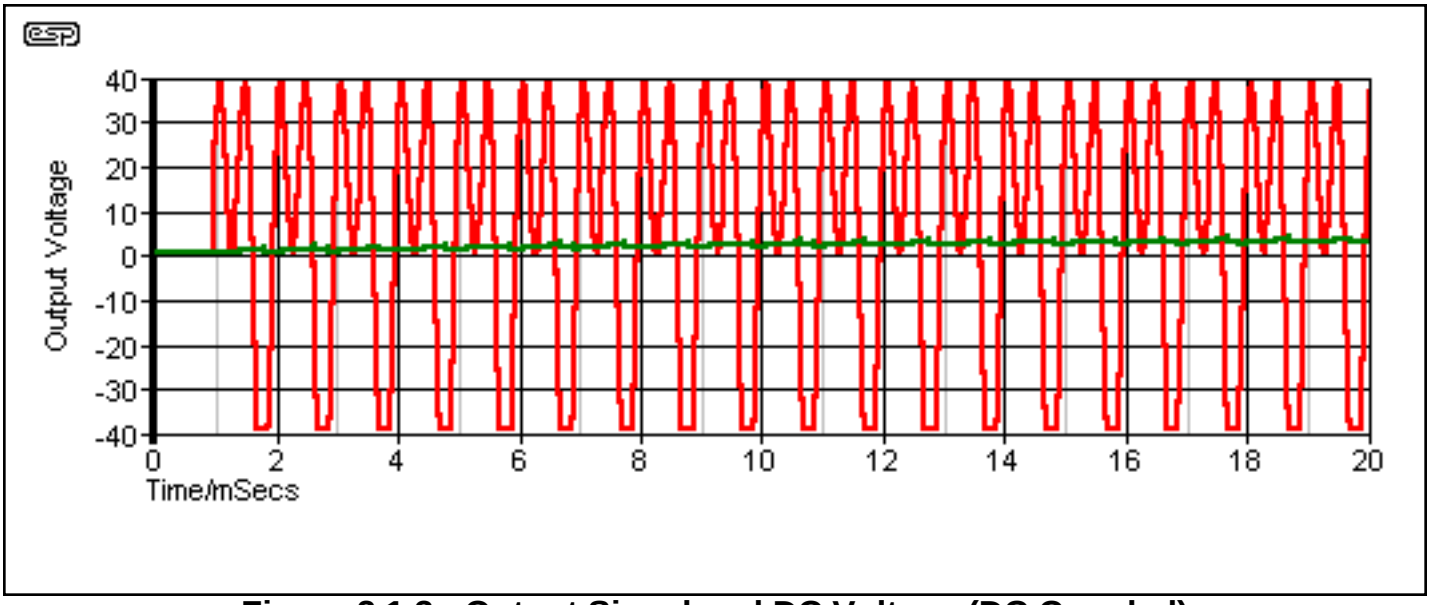


Figure 2.1.3 - Output Signal and DC Voltage (DC Coupled)

This is obviously the worst possible scenario - loudspeaker distortion will be very high indeed, and it is doubtful that even woofers (including long-throw subs) will be able to cope. This is especially true if the waveform is less 'friendly' than the one used for these simulations.

2.2 - Sources of Asymmetrical Waveforms

Most people probably tend to think that their music signal is relatively symmetrical, and this is usually true when averaged out over a long period (perhaps 30 seconds or more). However, there are a great many sources of asymmetry within the programme material itself, and these include ...

- Human voice (spoken or singing)
- Wind instruments (flute, horns, reed instruments, etc.)
- Plucked and struck stringed instruments (guitar, harp, piano, bass, etc.)

This list is not extensive, but covers the major 'culprits' - there are obviously many others such as drums of all kinds, synthesisers or other electronic sources, and not all notes or tones will be asymmetrical. In fact, any instrument may be symmetrical or asymmetrical depending on how it is played, the note being played, specific fingering techniques, etc. The averaging effects of a large orchestral ensemble will tend to create an overall symmetrical waveform in the long term, but there are many periods where a solo of any instrument will be not only highly asymmetrical, but made louder in the mix to even out

the sound level (compression).

Suffice to say that in any given piece of recorded music, it is almost a given that there will be periods of sufficient level and asymmetry to cause the problems indicated above if the amplifier is overdriven - the polarity of the asymmetry will vary as well, and often even within a relatively short period.

Any amplifier that is driven to even mild clipping will show the effects described, depending on whether it is AC or DC coupled. DC servo systems will not be fast enough to remove the DC component - these are usually made with rather long time constants to prevent the servo from interacting with the music material, and may actually make matters worse with some signals.

The signal I used for these simulations is relatively benign - there are a great many signals that could be used that would give far more dramatic results, but I elected to use a reasonably realistic signal without pushing the limits. In reality, the signal I used is reasonably typical, but of course there will be signals that are a great deal worse (either by accident or design).

2.3 - Other Issues

Apart from the possibility of the DC component causing gross loudspeaker distortion, in extreme cases it may also cause the amplifier's DC protection circuitry to operate. I have heard of some amplifiers that 'solve' this problem by setting a high detection threshold for their DC detection circuitry - it may be as high as 20V for a high powered amplifier. While this prevents 'false tripping' of the protection circuit, it also reduces the level of protection offered.

Given the right (or wrong) signal and sufficient clipping, there is the real risk that a loudspeaker driver may even be damaged. The voicecoil may be slammed into the rear polepiece of the magnetic circuit, suspensions may be stretched and the assembly's alignment compromised. This will lead to eventual driver failure.

While so far I have only mentioned conventional cone loudspeakers, the situation is made a great deal worse if there is a transformer used at the amplifier output - either to drive an Electrostatic Loudspeaker (ESL) or a 70V or 100V public address distribution system. In either of these cases, the DC that flows in the low resistance primary winding of the transformer may even cause amplifier failure, not to mention gross distortion (again) because of transformer core saturation.

3.0 - Conclusion

There is no doubt that a clipped asymmetrical waveform will generate a DC component in the output of an amplifier. There is equally no doubt (as evidenced above) that a fully DC coupled power amp is the worst possible case. Since there is no requirement whatsoever for an audio amplifier to reproduce DC, it follows that designs that are fully DC coupled are of no benefit to the listener, and indeed may cause far greater problems than they are supposed to 'solve' (according to those who insist that DC reproduction is somehow 'better').

Wherever possible, amplifiers should never be allowed to clip - this much is well known to anyone who is interested in quality reproduction. That it very likely will happen at some point is also accepted - parties in particular being the worst offenders. For those who use relatively small (i.e. low powered) amplifiers as a matter of course, the risk is greater, although the DC voltages so created are also reduced because of the lower overall supply voltage.

For those who prefer valve (tube) amplifiers, they are only partially immune from this problem. Since transformers cannot pass DC by their very nature, the DC cannot get to the voicecoil. There is still the real risk that distortion will increase dramatically if an asymmetrical waveform causes amp clipping, since it is now the output transformer that takes the burden of the DC ... this may cause core saturation. The problem will be worse with push-pull amplifiers, and their distortion could easily rise to the levels commonly found in single-ended designs.

To what extent have the effects described here influenced reviewers (who typically never use any instrumentation, and usually never know if the amp under test is clipping or not)? I have no idea, but it is not unreasonable to assume that some degree of clipping *must* be experienced from time to time, and that *will* affect the outcome of a subjective test ... but with absolutely no technical detail to indicate the actual cause of the problem should it exist. Reviewer-speak will obfuscate the real issue(s), and the lack of instrumentation leaves us in the dark.

Finally, it must be pointed out that this shows that clipping with real-world (speech or music) signals creates not only the harmonics that have been described in innumerable web pages, but also generates a subsonic signal that is potentially very damaging to drivers, but is never mentioned. This signal has the capability to cause driver damage at worst, or unwanted cone modulation and additional loudspeaker distortion at best -

neither can be considered a desirable outcome.

The moral of the story is to avoid clipping at all times - even momentary (supposedly inaudible) clipping will generate an unwanted low frequency or subsonic signal whose frequency will be completely unrelated to anything in the programme material.

4.0 - Acknowledgments & References

There is really only one primary reference that is pertinent to the facts here, and I was alerted to it by a contributor (who has also had some direct conversations with me on the subject). I considered the topic both interesting and important enough to put this information together, in the hope that disinformation and incorrect data elsewhere may be dispelled. The effect is real, is easily measured or simulated, and should be far better known than seems to be the case at present. To see the debate that started it all, go to ... alt.audio.pro.live-sound

To obtain an understanding of other effects that asymmetrical waveforms can influence (and also showing a waveform vastly more asymmetrical that the one I used here), see Allpass Networks in a speech chain by James L. Tonne.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 13 Oct 2004

The Audio Pages

Elliott Sound Products **Distortion & Feedback**

Copyright © 2006 - Rod Elliott (ESP)
Page Published 06 May 2006

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 What *Is* Distortion?
  - 2.1 How a Transistor Causes Distortion
  - 2.2 Historical Perspective
- 3.0 Principle of Negative Feedback
  - 3.1 Oh No, Not a Water Analogy!
- 4.0 Distortion Analysis
- 5.0 Examining the Feedback Loop
  - 5.1 TIM/TID - Transient Intermodulation Distortion
- 6.0 Amplifier Circuit Delay
- 7.0 Local vs. Global Feedback
- 8.0 Conclusion
- 9.0 References
- 10.0 Simulation Download

1.0 - Introduction

Claims abound regarding how bad negative feedback is, how it ruins the sound, and how zero feedback amplifiers with comparatively vast amounts of distortion sound so much better with music. Entire papers have been written on the topic, new methods described to quantify the audibility of different harmonics, and new measurement techniques are suggested and described ad nauseum.

Of those papers, articles and semi-advertisements, many make completely incorrect assumptions as to how feedback actually functions in an amplifier, and some extrapolate these false assumptions to arrive at a completely non-sensical final outcome. Before continuing, we need to clear up one very important point ...

> Feedback does not - repeat *does not*- cause the signal to travel from the output, back into the inverting input, and continue through the amplifier several (or multiple) times. By not doing so, it does not (and can not) create additional harmonics that did not exist before the feedback was applied.

Feel free to re-read the last statement as many times as you need to. This is a claim that has been made on numerous occasions, and it is simply false. The whole idea of feedback is that it is as close as possible to instantaneous - feedback is applied to the input of an amplifier in direct proportion to the signal at the output, and for all intents and purposes at *exactly* the same time. (This means that the amplifier must be fast enough to keep up with the input signal at all times.)

Once you have grasped the logic of how feedback actually works (as opposed to the way some people think it works), you are a long way towards understanding that many of the evils attributed to feedback are due to a lack of understanding, and have nothing to do with feedback itself. It has been claimed that applying feedback can actually increase the levels of higher order harmonics [1], however, this claim does not stand up to scrutiny. It is reasonable to expect that measurement errors or flawed assumptions are almost certainly the cause of this 'problem', but some parts of the industry will never let the truth get in the way of a good story.

For the majority of the tests described, the effects were simulated rather than measured. There are some very good reasons for this, with the primary reason being that the simulator has access to 'ideal' amplifiers. These have infinite bandwidth, infinite input impedance, zero distortion and zero output impedance. Being perfect, they also contribute zero noise. This enables one to perform experiments that simply cannot be done in the real world, and provide a level of accuracy that is also unattainable using real circuits. Likewise, the signal sources have zero distortion, so resolution exceeds anything attainable using actual circuitry.

---

2 - What *Is* Distortion?

It is useful to understand what distortion is, and how it is produced. The generation of harmonics is not a weird function of a valve, transistor or MOSFET, but is a physics phenomenon that occurs whenever a waveform is not a pure sinewave. A pure tone contains only one frequency - the fundamental. By definition, this pure tone is a sinewave - no other waveform satisfies the criterion for purity. As soon as a sinewave is modified, the waveform that now exists is created by adding harmonics. Likewise, anything that adds harmonics changes the waveform - the two

are inextricably intertwined. Amplifying devices do not add harmonics per se! Amplifying devices modify the waveshape, and this *requires* that harmonics are added to create the 'new' waveform. The creation of harmonics is a physics requirement, and has nothing (directly) to do with the type of device that caused the modification to the waveform. Devices with high linearity modify the sinewave less than devices with lower linearity, so fewer harmonics are created in the process.

Because the sinewave is a pure tone, it has long been used as a measure of the amount of non-linearity for amplifying devices. Even very small wave shape modifications can cause a large amount of distortion (and hence harmonic generation), and it is for this reason that sinewave THD (total harmonic distortion) tests are still used. Despite many claims to the contrary, a sinewave is not an 'easy' test - quite the reverse. Less than 1% distortion of a sinewave is easily heard (depending on the exact type of distortion), while it may be completely inaudible with some music or barely audible with others. Any device that amplifies will also distort, and the purity or otherwise of the output signal shows non-linearities very clearly. Interpretation of the test results does take some background knowledge though, and simply quoting a percentage with no qualifying parameters is completely useless.

Strictly speaking, simply turning a sinewave on or off causes distortion, because a truly pure tone is not only without harmonics, but has existed (and will continue to exist) for eternity. While this is real, no-one will ever take it to that extreme.

Because of the nature of a non-linear device which modifies the waveshape and thus causes the creation of harmonics, it should be obvious that it is not the amplifying device that generates the harmonics directly - it *only* modifies the waveshape. The harmonics are the result of the modified waveform - nothing more. To explain how a device modifies the waveform it is necessary only to look at the device's transfer function, and understand the process of amplification.

Amplification is an (almost) instantaneous process. An amplifier does not 'see' a complex waveform any more than we can experience all of last week simultaneously. As the Compact Disk medium has demonstrated, time can be separated into discrete fragments, and digital data can be derived that describes the instantaneous voltage at that point in time. This process is repeated 44,100 times each second. Compared to an analogue amplifier, this is very slow. The analogue domain does not use time fragments - all processing is done on a continuous basis - *but*, the amplifier is only capable of processing one instantaneous voltage level at any one time. The input voltage is a moving target, and the output signal follows it as closely as possible.

If an amplifying device has a gain of 10 when its (instantaneous) input voltage is 100mV, the output voltage will be 1V. If the device is non-linear, then the gain may fall to 9.5 when the input voltage is 1V, so the output will be 9.5V instead of 10V. This is distortion! That's it! The amplifying device does nothing more than change its gain slightly depending on the amplitude of voltage or current it has to deal with at any value of input voltage.

2.1 - How a Transistor Causes Distortion

Let's look at a common bipolar transistor as an example. The primary (but by no means only) form of distortion is caused by the internal emitter resistance of the transistor. Figure 1 shows a simple single transistor amplifier. A bias resistor is shown - it must be pointed out that this biasing method is never used in practice, because it is too dependent on device gain, temperature and supply voltage. Proper biasing that allows for thermal effects, device parameter spread, etc. is beyond the scope of this article.

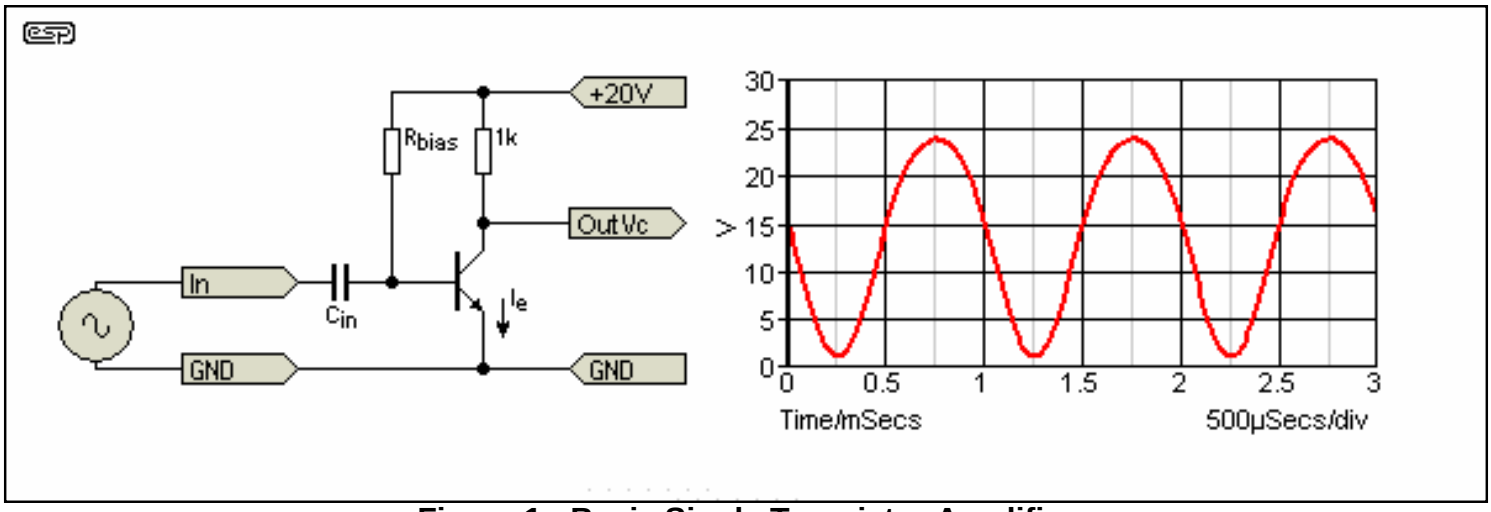


Figure 1 - Basic Single Transistor Amplifier

This is a very basic amplifier, but it embodies all the issues that face other amplifying devices as well - valves, JFETs and MOSFETs all have similar non-linearities, but for different reasons. It just happens that with a transistor it is easy to describe in simple terms. The output waveform is also shown, and distortion measures 12%, being second (-18.5dB), third (-52dB) and fourth (-56dB) harmonics. All others are over 90dB below the fundamental. It is generally taken that ...

re = 26 / Ie (mA) where re is the internal emitter resistance and Ie is the emitter current

The gain is determined by the ratio of the collector resistance to the emitter resistance, and is approximately ...

Av = Rc / (Re + re) where Av is voltage amplification, Rc is collector resistance, Re is external emitter resistance, and re as above

Re (the external emitter resistance) has not been included in the circuit of Figure 1, which has a gain of about 390. As we shall see, this varies over the output voltage range, so the measured value gives a false impression because of waveform modifications. Table 1 shows how much the circuit of Figure 1 will vary the emitter current and hence the (theoretical) gain, depending on signal level. The base current has been ignored, but this also has an influence - albeit rather small.

Vc (Volts)	Ie (mA)	re (Ohms)	Voltage Gain
29	1	26	38
25	5	5.20	192
21	9	2.89	346
17	3	2.00	500
13	17	1.53	654
9	21	1.24	807
5	25	1.04	962
1	29	0.89	1115

Table 1 - Gain Variation of Figure 1 Circuit

You can see from the table how the waveform of Figure 1 comes about. When the collector voltage is high, the gain is lower, and the waveform is flattened. When the collector voltage is low, the gain is much higher, so the waveform becomes elongated. As is obvious, the gain varies over a wide range, and any waveform applied to the base must become distorted. Transistors show a logarithmic response when the base to emitter junction is driven from a voltage source, and table 1 shows this effect quite clearly.

Because the transfer function is non-linear, it must alter the wave shape. If the wave shape is altered, harmonics are produced. To reduce distortion (of all forms), the application of negative feedback will make the amplifier more linear, and this results in fewer harmonics. There is no mystery and no magic. It doesn't matter if the feedback is global (applied around a complete circuit) or local (applied to each device individually). In general, global feedback gives better results than local feedback, but only if the amplifier has high open loop gain (i.e. gain without feedback).

Prior to adding feedback, it is advantageous to improve the circuit's linearity by other means if possible. Since the gain of a transistor varies widely with emitter current, maintaining a constant current (via the collector) will help. Since transistors are current controlled, using a variable current for the input will also help - distortion can be halved by this alone, but voltage gain is reduced. In the case of the above circuit, using a 15mA constant current source instead of the 1k resistor increases the voltage gain to 3227, and reduces distortion to 4% - using current input (via a series resistor) reduces gain, but also reduces distortion even further.

The additional gain from the use of a current source load allows us to apply feedback - if the gain is set at 400 (close enough to the 390 measured before), distortion is reduced to 0.7%. The second harmonic is now -43dB, the third is -70dB and fourth is at -95dB (all with respect to the fundamental). Compare these figures with those obtained for the circuit as shown - no comparison! This is covered in more detail in Section 6.

Alternatively, *Re* (the external emitter resistance) can be added to create 'local feedback'. By

adding an external resistor, we actually do nothing more than (partially) swamp the variation of *re* with emitter current. While this makes the circuit more linear, it is not really feedback at all - the correct term is degeneration. Gain variation (and hence distortion) is reduced because *Re* + *re* is much greater than just *re* alone and base current is also more linear, but one of the benefits that feedback (as opposed to emitter degeneration) gives is reduced output impedance. Emitter, cathode or source degeneration does not lower output impedance.

2.2 - Historical Perspective

There is a great deal of information that was compiled a long time ago that seems to have been forgotten, dismissed, or simply neglected. Of particular interest is the section on distortion in the Radiotron Designer's Handbook [2]. Since some (many) of the detractors of negative feedback advocate single ended triode operation, one would expect that they would have examined what was considered 'high fidelity' back in 1957, rather than claim that amplifiers that were considered low fidelity back then represent high fidelity today. This is not a tenable position!

Of some interest is a table of harmonics based on a fundamental of C - taken for convenience as 250Hz. The table is reproduced below. It shows the musical relationship of each harmonic up to the 25th with respect to the fundamental, based on the natural or just musical scale (as opposed to the equally tempered scale that is used for most instrument tuning).

Harmonic	Frequency	Note	Comment
1st	250	C	Fundamental
2nd	500	C^1	
3rd	750	G	
4th	1000	C^2	
5th	1250	E	
6th	1500	G	
7th	1750	-	Dissonant
8th	2000	C^3	
9th	2250	D	
10th	2500	E	
11th	2750	-	Dissonant
12th	3000	G	
13th	3250	-	Dissonant
14th	3500	-	Dissonant
15th	3750	B	
16th	4000	C^4	
17th	4250	-	Dissonant

18th	4500	D	
19th	4750	-	Dissonant
20th	5000	E	
21st	5250	-	Dissonant
22nd	5500	-	Dissonant
23rd	5750	-	Dissonant
24th	6000	G	
25th	6250	G#	Dissonant

Table 2 - Harmonic Distortion on the Musical Scale

Obviously, harmonic distortion that extends to the 7th or beyond is to be avoided. It is (or was) well known to guitar amp manufacturers that the seventh harmonic should not be reproduced if possible (even during overdrive conditions) because of just this issue - discordant (or dissonant) harmonics simply don't sound nice.

Another table shows the levels of distortion that were considered objectionable, tolerable and perceptible for various frequency limits and triode or pentode valves. This table is also reproduced, but I have only included the 15kHz results - other bandwidths were listed, but no-one would consider a bandwidth of 3750Hz acceptable these days.

Source	**Mode**	**Distortion**	**Comments**
Music	Triode	2.5%	Objectionable
	Pentode	2.0%	
Speech	Triode	4.4%	
	Pentode	3.0%	
Music	Triode	1.8%	Tolerable
	Pentode	1.35%	
Speech	Triode	2.8%	
	Pentode	1.9%	
Music	Triode	0.75%	Perceptible
	Pentode	0.7%	
Speech	Triode	0.9%	
	Pentode	0.9%	

Table 3 - Comparative Distortion Tests (Olson)

These figures are interesting compared to amplifiers of today. Both triode and pentode amplifiers used in the test had an output of 3W, and were conducted in a 'typical' listening environment. While modern (competent) transistor amps will invariably beat the distortion criteria by a wide margin (at any level or frequency), modern SET amps seem to be considerably worse than one would hope, many having distortion that rates as objectionable -

and this table was compiled was a very long time ago indeed.

For those who have access to the complete text of the Designer's Handbook (or at least Chapter 14 which concentrates on fidelity and distortion) I strongly recommend that it be read in its entirety. There is a great deal more to it than I have the space to reproduce here, and the fundamental principles have not really changed, despite the passing of the decades since it was written.

There is an informative section covering intermodulation distortion, in which it is pointed out that there is no direct correlation between THD and IMD. It is also pointed out that no actual amplifier has *only* second or third harmonic distortion - every form of distortion is accompanied by multiple harmonics, although either second or third harmonics can be the most dominant.

3.0 - Principle of Negative Feedback

Negative feedback (or just feedback) has been used for many years to linearise amplifiers. Between 1935 and 1937, Harold Black of AT&T received three U.S. patents relating to his work on the problem of reducing distortion in amplifiers by means of negative feedback. The invention caused little controversy for many years, but eventually this happy situation had to end - at least in the hi-fi industry. Feedback is used extensively in medical, military, aerospace and industrial applications and seems not to cause any problems there, despite its bad reputation amongst some audiophiles.

Although many of the early attempts were less than perfect, it must be understood that the results without the feedback would have been many times worse. Negative feedback cannot make a dreadful amplifier sound good, but may make it sound acceptable. There is no possibility that the use of feedback will make a good amplifier sound bad. Not only are distortion components reduced, but negative feedback also increases the input impedance, reduces output impedance, and linearises frequency response. It is not a panacea, but it does come very close.

So, let us examine what feedback really does. Figure 2 shows the basics of a gain block - in this case, an operational amplifier (opamp). It may be comprised of any number of devices, and the active components can be valves (tubes), transistors, FETs, MOSFETs or any combination thereof. The gain block will be assumed to have infinite gain and infinite bandwidth for the initial analysis - we all know this is not possible, but it makes understanding the principle easier.

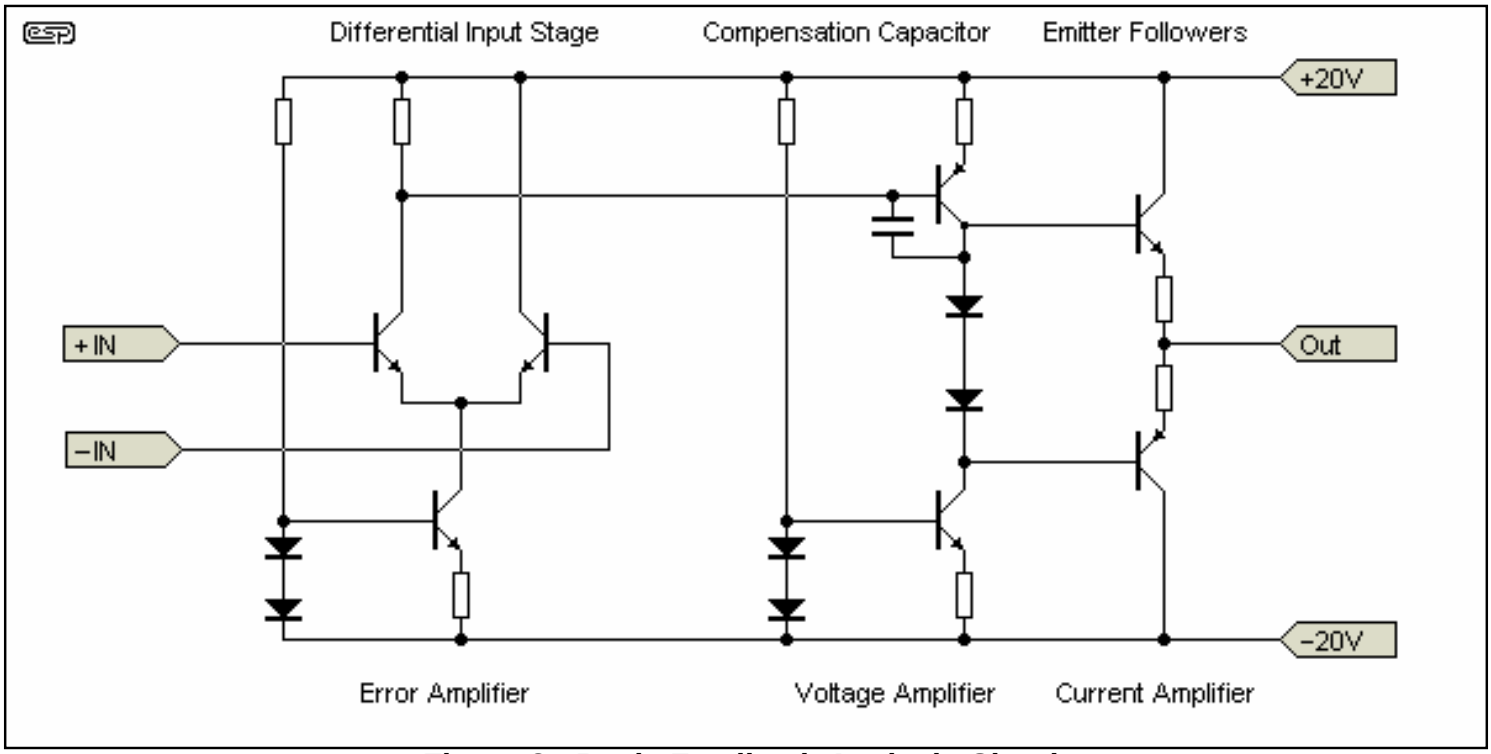


Figure 2 - Basic Feedback Analysis Circuit

An amplifier (power amplifier of conventional topology, opamp, etc), consists of three discrete stages. These are ...

1. Error amplifier (commonly referred to as the input stage)
2. Voltage amplifier stage (VAS)
3. Current amplifier (output stage)

Each of these may be as simple or complex as desired or needed, and each can use a different technology. The functions of each stage are (or will become) self explanatory, and a quick look at any of the project amplifiers (e.g. P101, P3A, etc.) will show that the same basic stages are used in most amplifiers.

If you have read the article Designing With Opamps, you will know the two rules of opamps (a typical semiconductor power amplifier may be thought of as an opamp for all intents and purposes). These rules are ...

1. An opamp will attempt to make both inputs exactly the same voltage (via the feedback path)
2. If it cannot do so, the output will assume the polarity of the most positive input

In any linear circuit, rule #2 is inapplicable unless there is a fault or overload condition, so only rule #1 needs be considered for this discussion. As shown below, a voltage of 1V is applied to the non-inverting input - the normal input for an audio amplifier. I will state at the outset that only one thing is important - the value of the voltage presented. We need not concern ourselves with frequency - indeed, time is utterly inconsequential.

Referring to the practical circuit shown in Figure 9, in order to fulfil rule #1, the amplifier's output voltage must be exactly 11V. This figure is simply derived from the voltage divider formula ...

$V_{out} = V_{in} * (R1 / R2 + 1)$ Where V_{out} is the voltage at the -ve input and V_{in} is the voltage at the amplifier output

Therefore, at the inverting input we should measure ...

$V_{-in} = 11 / (10k / 1k + 1) = 11 / 11 = 1V$

The first rule is satisfied, and the system is stable. The error amplifier is the critical element here. If the input voltage changes, the error amplifier simply detects that its two inputs are no longer the same, so commands the VAS to correct the output until equilibrium is restored. This is *not* an iterative process, which is to say that the amplifier does not keep feeding the input signal (meaning a significant part of the input waveform) into the inverting input to be re-amplified, re-distorted and re-compared. This is where some of those who criticise negative feedback have made their first error.

The output of the amplifier simply keeps changing in the appropriate direction until the error amp detects that the voltages are again identical, at which point the output of the error amp ideally just stops where it is, and so does the rest of the chain. In reality, there will be a small amount of instantaneous correction as the two voltages approach equality, but this *must* happen much faster than the input signal can change with normal programme material.

The fact that the correction is usually done well before the input voltage has even changed significantly clearly means that no part of the feedback signal is fed through the amplifier over and over again - that just doesn't happen. In our ideal device, the change is instant, in a real device it is possible to measure the time it takes for the correction to be made. For an audio amplifier, the correction must be completed faster than the highest frequency of interest can change - how much faster is open to some conjecture, and that will be looked at later in this article.

All amplifying devices have some distortion. Desirable though it may be, a distortion free amplifier doesn't exist - other than in a simulator. Some opamps come very close (with feedback), but inherent non-linearities within the amplification chain are inevitable. Without feedback, the distortion components tend to be low order (i.e. second, third, fourth, etc., with diminishing amplitudes as the order increases. The application of negative feedback reduces the amplitude of these harmonics (hence the term harmonic distortion), in direct proportion to the amount of feedback applied.

A common claim is that, because the feedback signal is re-amplified, the distortion components

are subjected to additional distortion. This supposedly creates high order harmonics that did not exist as a result of the original distortion mechanism in the amplifier. Since the feedback acts as an ultra-high-speed servo system, it is difficult to imagine why it is assumed that high-order harmonics are 'generated'. They are not generated at all, but simply become more easily measured because all the lower harmonic clutter is removed (in part at least).

Also notable is that *any* signal that is created within the feedback loop (most commonly noise) is also cancelled by global feedback. Because this generates signals that did not exist at the input, the error amplifier 'sees' any such extraneous signal as a deviation from the input signal, and cancels it to the best of its abilities. Note that input device noise is *not* cancelled, because the error amplifier cannot differentiate between noise it has created and the input signal.

That this works was amply demonstrated many years ago when the only cheap opamp was the venerable uA741 and a few others of similar noise performance. These are (still) notoriously noisy, so many designers added an external input stage using low noise transistors. This addition reduced the noise to acceptable levels, even for sensitive high-gain amplifiers as used for phono preamps and tape head amplifiers. The external transistors formed the error amplifier, and being low noise types were able to cancel out much of the opamp's internally generated noise - the additional gain also improved distortion performance.

This ability of the feedback loop to cancel internally generated signals (be it noise or distortion products) is so critical to your understanding of feedback that I have included a circuit and simulation results. These probably show more clearly than any other method how feedback works to remove anything that is not in the original input signal, by using the error amplifier to correct the output by applying an 'anti-distortion' component to the amplification stages within the feedback loop.

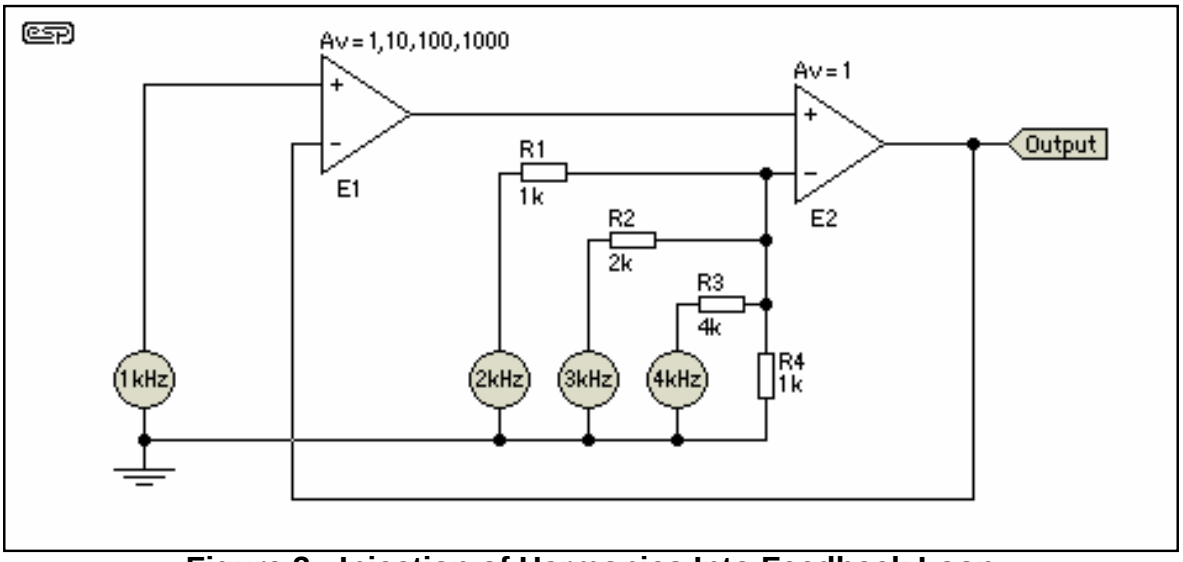


Figure 3 - Injection of Harmonics Into Feedback Loop

All signal sources have the frequency indicated, and all are set for an output of 1V peak (707mV RMS). Because of the simulator, there is no concern with frequency drift, so the distortion

waveform will remain the same - this test can be run easily with real opamps, but attempting any harmonic relationship is pointless because the frequencies will drift. If you have access to synchronised oscillators it's not a problem, but I don't, and I doubt many others will either.

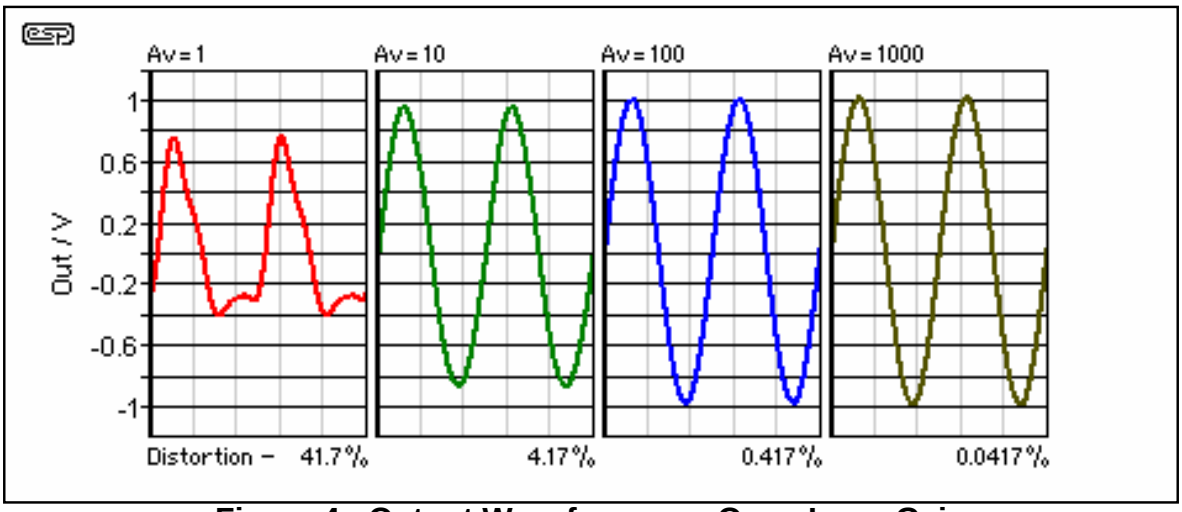


Figure 4 - Output Waveforms vs. Open Loop Gain

The first waveform is with E1 set for unity gain. There is some degeneration, but no feedback. If the feedback loop is disconnected, the waveform remains the same, but at a higher amplitude. As gain is increased (only the gain of the first stage (error amplifier) is changed), the distortion is reduced in direct proportion to the error amplifier's gain. There is no point reproducing a spectrum for this test, as the relationships are fixed by the 2, 3 and 4kHz signal sources. Only the total amplitude of the 'harmonics' is reduced with respect to the fundamental.

Although the circuit shown is configured as a unity gain buffer, adding feedback resistors to give the circuit gain makes no difference to its ability to remove the injected harmonics. To verify this, the error amp was set to a gain of 10, and the gain of the whole stage was increased to 10 by means of a 9k resistor from output to inverting input, and 1k from inverting input to ground. There was a significant gain error (Av = 5 rather than 10 as set by the resistors), but the rejection of the extraneous signals was just as effective.

Likewise when the error amp's gain was 100 (Av = 9.09) and 1000 (Av = 9.9). This is normal behaviour for an opamp - the open loop gain ideally needs to be 1,000 times greater than the required gain to achieve gain accuracy of 0.1%. While interesting and useful to know, that is not relevant to this article.

The above circuit will work with opamps too. Voltage controlled voltage sources are convenient in the simulator because their gain can be changed where one has no control over the open loop gain of an opamp, and some changes are needed to make a 'real' opamp work. However, the same distortion reduction is clearly evident.

3.1 - Oh No, Not a Water Analogy!

Sorry, but yes . A negative feedback system may be thought of as a servo, but that won't help anyone who is not familiar with servos. A toilet cistern is another matter - everyone has seen one, although not everyone has looked inside. I encourage you to do so . The cistern is a good example of a simple negative feedback system. Unlike an amplifier (which is bipolar - it can generate positive and negative output voltages), a cistern is more like a regulated power supply - these also use negative feedback to maintain a stable voltage.

When water is let out of a cistern, the water level falls, and this in turn opens a valve. The water is replaced until such time as the level is restored to its original preset level. If water is allowed to escape at a low (but variable) rate, the float valve (ball cock) will regulate the water level perfectly, maintaining the same level even as you allow more or less water to escape. This is a simple example of negative feedback at work in your bathroom. For expedience, I have neglected the uncertainties of the mechanical linkages and valves (as well as the inertia of the water itself), but you knew that already.

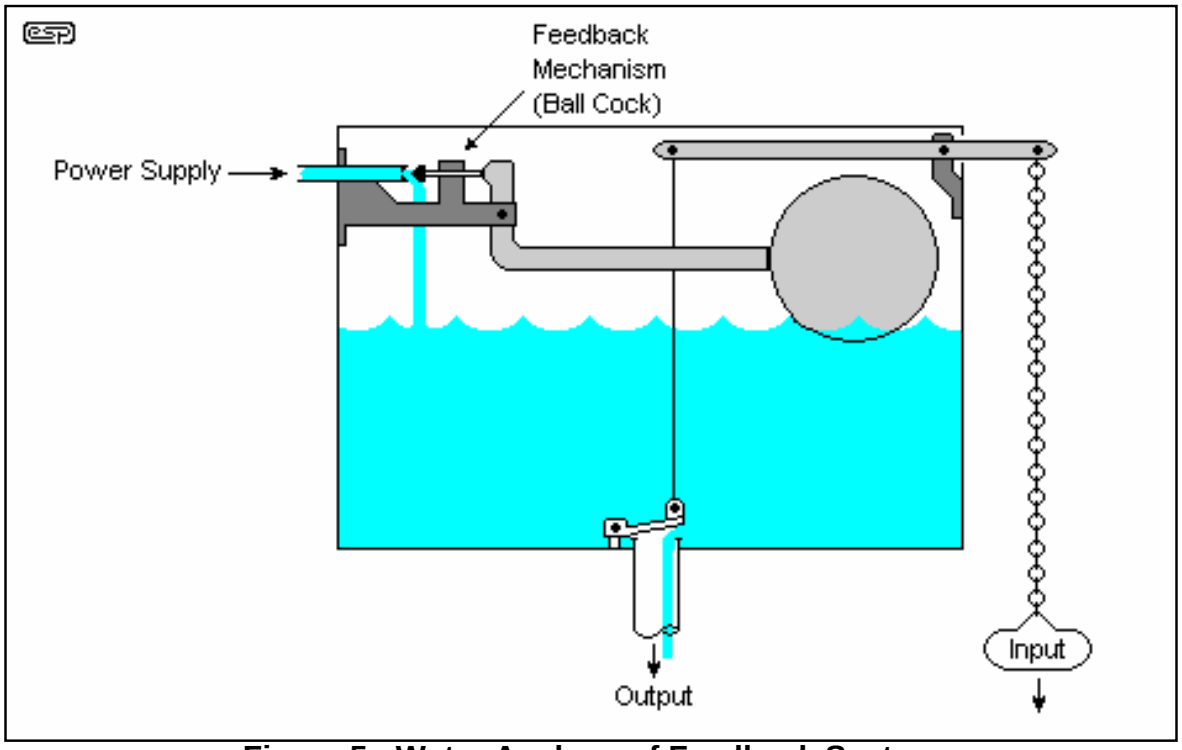


Figure 5 - Water Analogy of Feedback System

Should the water be allowed to escape faster than it can be replenished, the system is in an overload condition. This is no different from an amplifier where the input signal changes faster than the output - the system cannot keep up, so the output is 'distorted'. I am unsure if this will help, but if it does improve your understanding of negative feedback, then it was worth it.

For those in Australia, be aware that the above analogy cannot be used because our water reserves are too small to allow the luxury of playing with water. We will just have to imagine that it works .

4.0 - Distortion Analysis

So, having established that the output signal is not re-amplified over and over again instantly removes one of the criticisms of negative feedback - that it creates frequencies that didn't exist before feedback was added. Since there is no re-amplification of the signal, there can be no new frequencies created, other than the distortion of the waveform caused by device non-linearity. Figure 6 shows a simulation circuit, using a diode to create distortion [3]. The distorted signal is enclosed within the feedback loop (Feedback) of a pair of VCVS (voltage controlled voltage sources - 'perfect' amplifiers in the world of the simulator). A second circuit (Open Loop) applies the same distortion, but simply amplifies the distorted signal to obtain the same RMS voltage. C1 and C2 provide DC blocking to remove the diode's forward voltage.

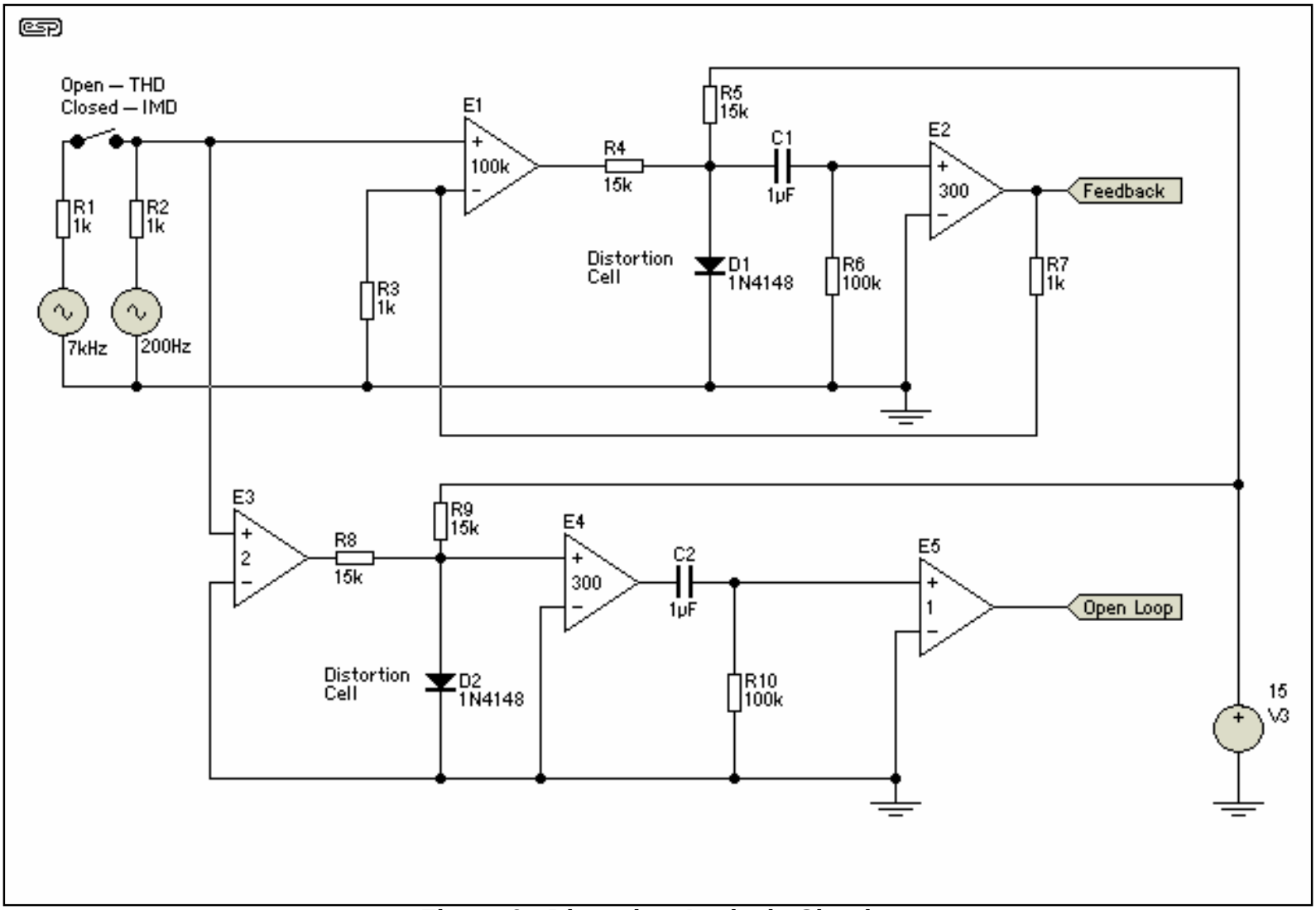


Figure 6 - Distortion Analysis Circuits

The applied input signal is 2V peak at 200Hz + 500mV peak at 7kHz, so we can see both harmonic and intermodulation products as generated by the non-linear element - a forward biased diode, passing ~15mA. This attenuates the signal greatly, and applies a controlled amount of distortion, measuring at 8.5% for a single frequency. In each case (feedback and open loop) the input voltage to the distortion cell was maintained at as close as practicable to

the same level, although quite wide variations do not cause significant changes to the distortion level.

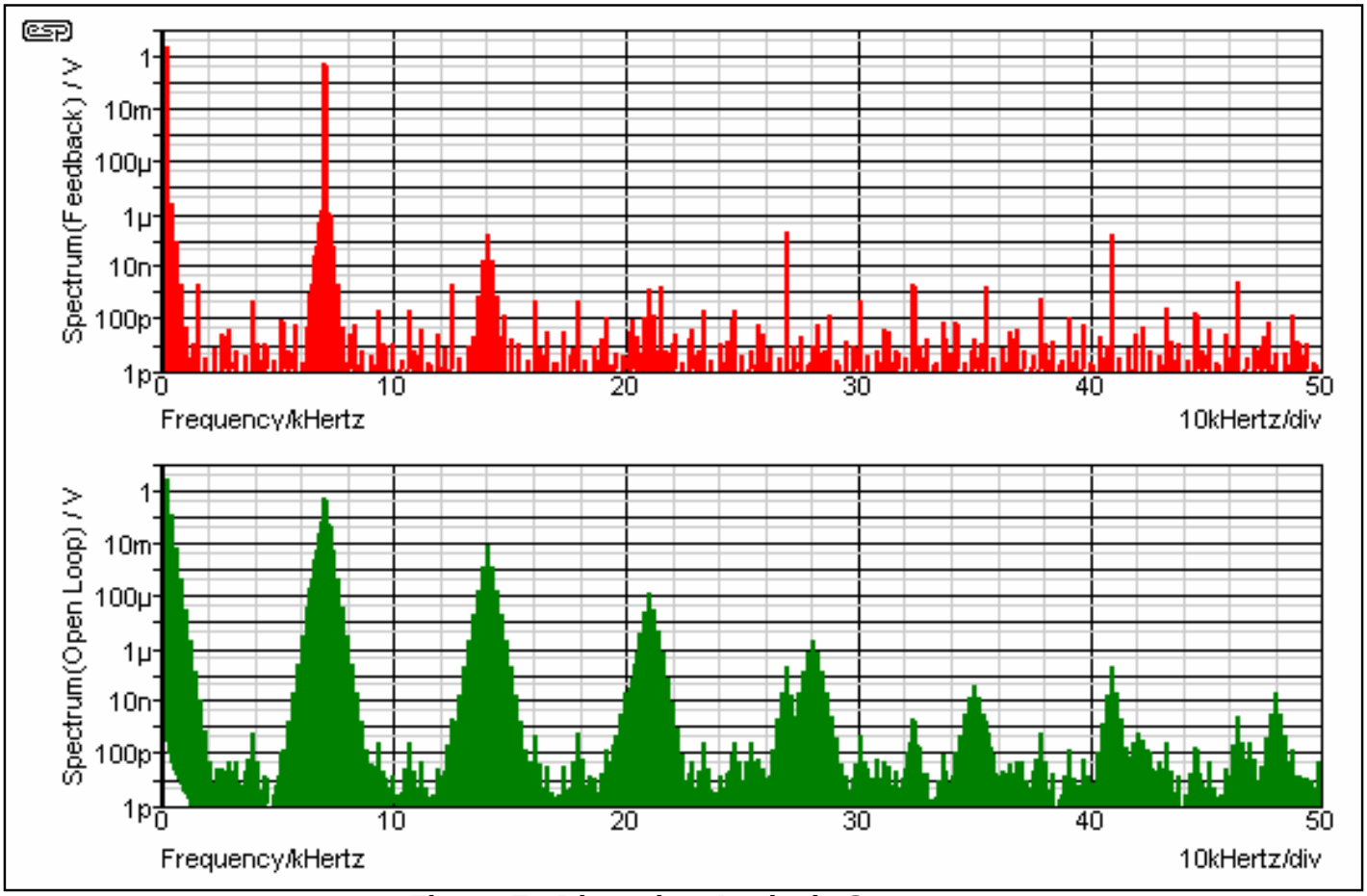


Figure 7 - Distortion Analysis Spectra

Looking closely at the FFT analysis of both the feedback and open loop circuits shows clearly that the distortion is reduced by the application of negative feedback. There is no evidence that any individual harmonic frequency is at a greater amplitude when feedback is applied, but you can see some signals that are not affected either way - these are simulation artefacts, and should be ignored. Note that the base level of 1pV is -240dBV - this can never be achieved in reality, so you can ignore any value below 1uV (-120dBV). Even this is rather adventurous, and 10uV is more realistic at -100dBV.

Note the peaks at and around 14kHz, 21kHz, 28kHz and 35kHz. These are highly affected by feedback because they are harmonics and intermodulation products of the 200Hz and 7kHz input frequencies, and are virtually eliminated by applying feedback.

The spikes at 26.92kHz and 40.92kHz are not affected, because these are artefacts of the sampling rate (a simulator works in a manner similar to any digital system, and uses sampling to convert the 'analogue' signal into digital for processing).

For reference, I have also included a spectrum analysis for a single 1kHz sinewave. This makes the picture clearer, and is the way THD is measured using spectrum analysis. The harmonics

are seen clearly, and it is notable that a circuit that one may assume would produce only even harmonics also produces odd harmonics. There is a school of 'thought' that is convinced that single-ended triode amplifiers (for example) produce only even ('nice') harmonics, while yucky push-pull amps produce only odd harmonics. This is not the case. While it is true that push-pull amps do indeed cancel the even harmonics, if the first claim were true, a push pull amp using triodes would cancel the even harmonics (which they do), leaving no distortion at all at the output (which they don't). Even-order harmonic distortion in isolation does not happen - it is invariably accompanied by odd-order harmonics, as demonstrated by the open loop response shown below. Taking the 'even order distortion only' argument to extremes, in order to obtain *only* even order harmonic distortion, the first harmonic (the fundamental) cannot be present because it is an odd number! While a bridge rectifier can achieve this, the sound is unlikely to gain wide acceptance .

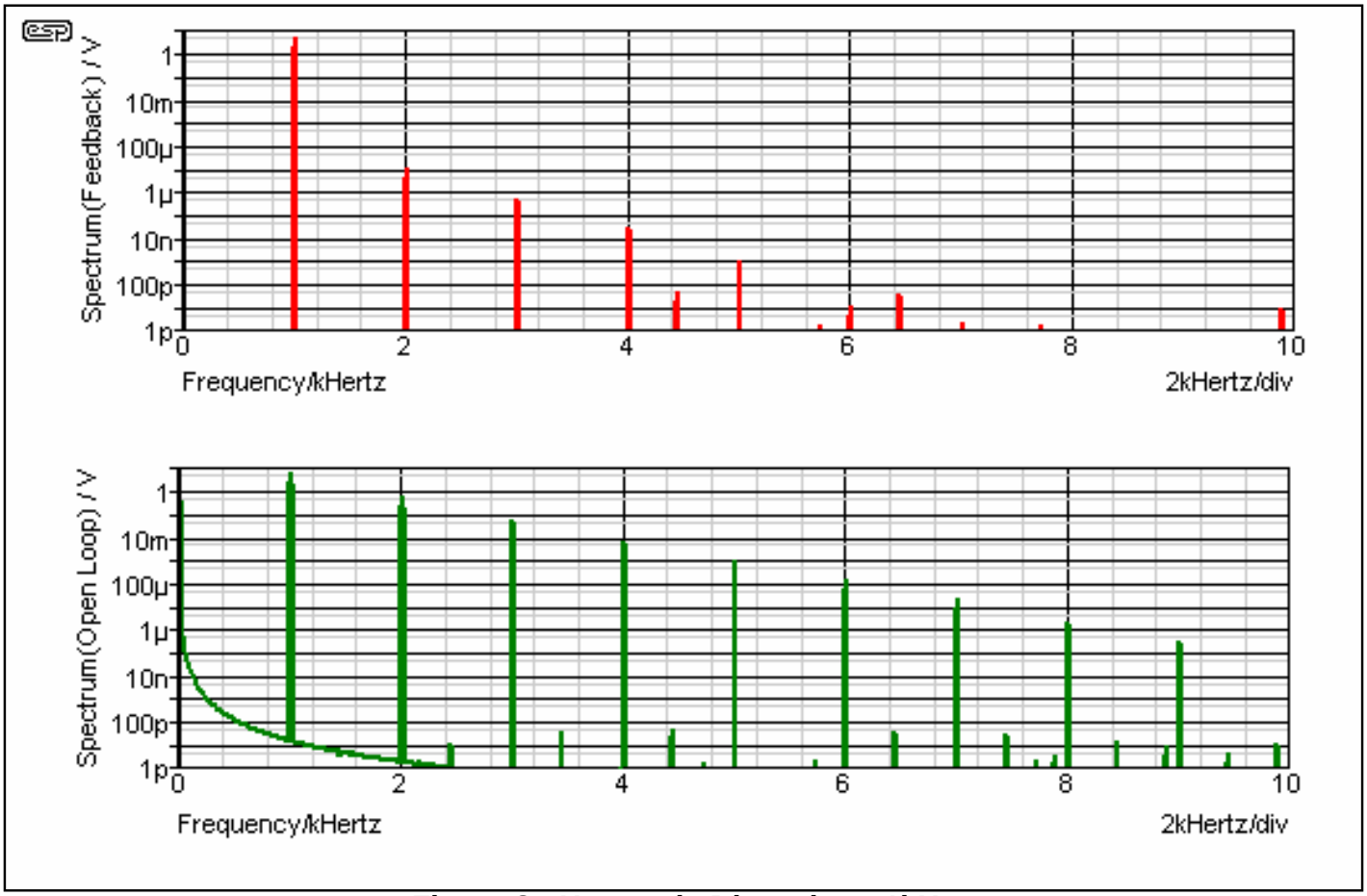


Figure 8 - Harmonic Distortion - 1kHz

Note that the open loop distortion products show diminishing amounts of both odd and even harmonics. Only those up to the seventh harmonic (7kHz) are relevant - all others are more than 100dB below the fundamental. When feedback is applied, *all* of the distortion products are greater than 114dB below the fundamental. Also, note that not *one* distortion product is at a greater level than in the open loop circuit. The spectra shown only extend to 10kHz because there are no significant harmonics above that frequency. Reducing the gain of E1 reduces the feedback ratio and increases the level of the harmonics as one would expect. Changing from 100k to 10k (20dB) increases the amplitude of the harmonics by 20dB. If E1 is reduced to a gain

of 1k, the second harmonic is increased to -74dB with respect to the fundamental. This effect is quite linear over a significant range.

As with the intermodulation test above, there are artefacts of the simulation and FFT process. The small peaks at 4.44kHz and 6.44kHz are not related to the 1kHz input signal, but are so far below the noise floor that it wouldn't matter if they were real. These signals exist in both cases (and at the same amplitude).

5.0 - Examining the Feedback Loop

Having looked at some examples using ideal amplifying devices with no real-world limitations, it is now time to examine real circuits. Unlike their simulated counterparts, real amplifiers have finite bandwidth and slew rate (maximum rate of change), finite input and output impedances, and are not free of distortion. For the audio frequency range, this makes very little difference, despite claims that these limitations lead to Transient Intermodulation Distortion or 'TIM' - now pretty much universally discredited, but still quoted by some [4].

An amplifier simply needs to be somewhat faster than needed for the highest frequency of interest. Just as in the explanation given above, real amplifiers don't care if the input is AC, DC, or a mixture of multiple frequencies. The only things of interest are the instantaneous voltage level and the highest frequency of interest (and its relative amplitude). The latter determines how quickly the output must change to prevent it from losing control.

One major limitation in any amplifier is propagation delay - how long it takes for a signal applied to the input to reach the output. Propagation delay depends on actual semiconductor delays, as well as phase shift introduced by the dominant pole capacitor. This component is almost invariably needed to maintain stability, because the amplifier must have less than unity gain when the total phase shift through the amp is 180°, otherwise it will oscillate.

Without the dominant pole compensation, propagation delays will be sufficient to cause a 180° phase shift while the amp still has significant gain. For example, if an amplifier has a propagation delay of 1us, this causes the phase to be reversed at 500kHz, so the amp will oscillate strongly unless the gain is reduced to slightly less than unity for any frequency of 500kHz or above.

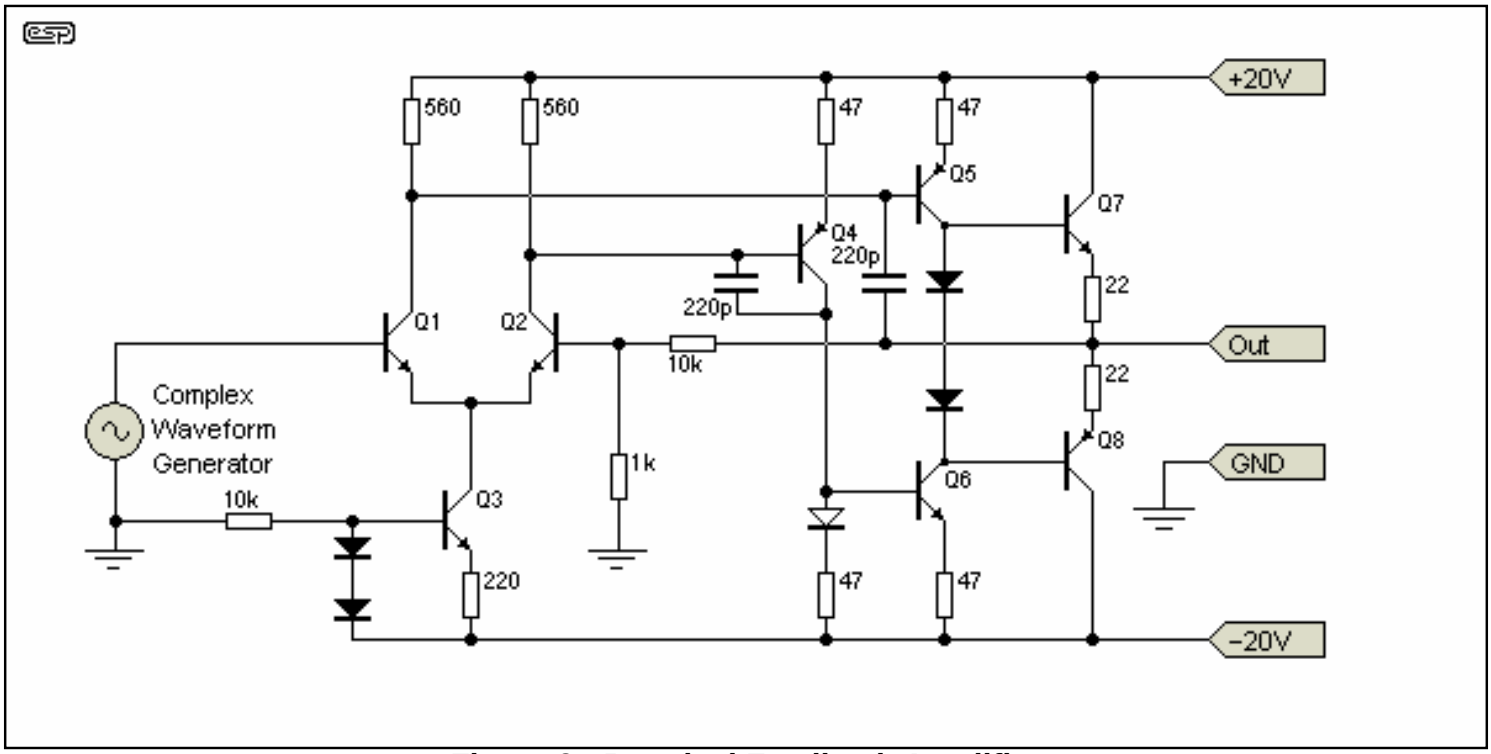


Figure 9 - Practical Feedback Amplifier

In order to obtain approximately equal slew rate for positive and negative going signals, the circuit of Figure 6 was used. Q1, Q2 and Q3 form the error amplifier, Q4, Q5 and Q6 make up the VAS, and Q7, Q8 is the current amplifier. Open loop gain is 20,000 (86dB), and the HF compensation caps (220pF) cause the open loop frequency response to be 3dB down at 2.4kHz. As is typical with such circuits, there is less feedback available at high frequencies because of the requirement for the dominant pole capacitors. They are not needed for open loop operation, but all linear (audio) applications will use the amplifier as a closed loop (feedback) circuit.

At an output voltage of 1kHz / 3.7V RMS, open loop distortion is 2.3%. Input impedance is about 7k, with output impedance at about 200 ohms. The distortion components are low order as expected, with only second and third harmonics at significant levels. The fourth harmonic is at -85dB relative to the fundamental.

Adding feedback, but maintaining the output at the same voltage, things change much as we would expect. Distortion at 1kHz now measures 0.0014%, and only the fundamental is above -98dB (the level of the second harmonic with feedback). What happened to all the high order harmonics 'generated' by the addition of feedback? As fully expected from previous tests, they simply don't appear - *all* harmonics are suppressed to much the same degree, but with some dependence on the open loop gain (and hence feedback ratio). In this case, closed loop voltage gain is 11 times.

With feedback, frequency response is -3dB at 4.3MHz (no, I don't really believe that either), input impedance a more respectable 5.8MΩ at low frequencies, falling to a bit under 1M at 20kHz. Output impedance is well under 1 ohm. Apart from the rather optimistic frequency

response reported by the simulator, the figures are pretty much what I would expect.

The slew rate is 11.5V/us positive and 18V/us negative - not exactly equal, but it will have to do. The maximum slew rate for a sinewave is determined by ...

$$\text{Slew Rate} = (2 * \pi * V_{peak} * f) / 10^6 \text{V/us}$$

So, it we want to get 10V RMS output at 100kHz, the required slew rate is ...

$$V_{peak} = V_{RMS} * 1.414 = 10 * 1.414 = 14.14\text{V}$$
$$\text{Slew Rate} = (2 * \pi * 14.14 * 100\text{k}) / 10^6 = 8.9 \text{ V/us}$$

Despite the gain rolloff after 2kHz and the relatively low slew rate for the desired frequency (it's not even double that needed for a positive going signal), the distortion measures 0.038%, and no harmonic exceeds a level of -70dB (with respect to the output of 10V RMS). The fifth harmonic is at -85dB. Remember that this is for a frequency of 100kHz.

5.1 - TIM/TID - Transient Intermodulation Distortion

The concept of TIM (Transient InterModulation distortion) was first proposed in the 1970s by Otala, and although it created a stir for a while, most designers realised fairly quickly that it does not happen in any sensibly designed amplifier. The 'dark side' of the industry seized upon TIM as their 'proof' that feedback was bad, and the debate has raged ever since. Some supposedly objective works on the topic have glaring errors, or have completely ignored other factors [4], such as amplifier output impedance and its effect on the response of a loudspeaker. It is notable that almost without exception, driving a speaker with higher than normal impedance sounds 'better'. Frequency response is less linear, damping factor is (much) lower, but somehow it sounds really good - at least in the short term. However, it is a grave error not to eliminate this variable from a test, because the sound difference is usually unmistakable.

According to the theory, when an amplifier has feedback around it, the delays between the input and output changes will be such that *huge* amounts of TIM will be produced. Naturally, a sinewave will never show the effect (at any frequency), and traditional measurement techniques will be useless for identification of this mysterious distortion mechanism. A useful test is to apply a squarewave at (say) 1kHz, with a sinewave superimposed upon it. This test will certainly let you know if there is a problem, but although I have used the test many times on amplifiers that should have vast amounts of TIM, no problems have ever been seen.

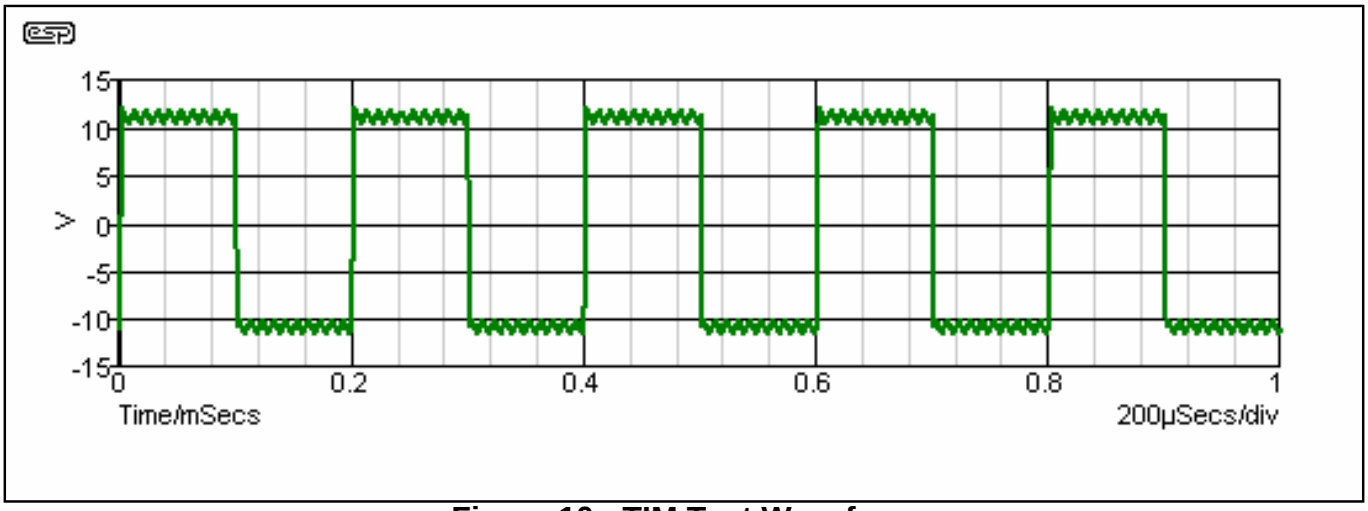


Figure 10 - TIM Test Waveform

Figure 10 shows the output waveform of the Figure 9 amplifier, which consists of a 10kHz squarewave whose slew rate is limited by the amplifier, with a 100kHz sinewave superimposed. This combined signal forces the amplifier into slew rate limiting, where the output cannot keep up with the input. The rise and fall times for the input squarewave are set at 1ns - many times faster than the amplifier can accommodate. Regardless of that, the sinewave shows very little modification - certainly there is a small section that is simply not reproduced at all, but this is with input frequencies and rise times that *do not occur in any type of music !*

Although a CD is capable of full output level at 20kHz (a slew rate of 5V/us for a 100W / 8 ohm amplifier), such a signal will *never* occur in music. This is a good thing, because tweeters cannot take that much power anyway. An examination of the maximum level of any music signal vs. frequency will show that the level at 20kHz is at least 10dB below that in the mid band - 10W for the amplifier above, or a slew rate of 1.6V/us. No sensible designer will ever limit an amplifier to that extent, but allowing 5V/us is easy, and will let the amplifier match the maximum rate of change of the CD source. In case you were wondering, vinyl can't hope to match a CD for output level at high frequencies, because at the first playing with the best cartridge and stylus available the high amplitude high frequency groves would be damaged forever. That vinyl can reach higher frequencies than CD is not disputed, but the level is very low. Fortunately, very high frequencies are never present in music at very high amplitudes.

As for claims that local feedback is 'good' and global feedback is 'bad' this is generally false. Global feedback around a competently designed amplifier will generally give much better results than multiple local feedback loops. Remember that waveform modification causes distortion, so a number of low gain stages with local feedback will generate additive distortion because *each stage applies its own amount of modification to the signal*! This is real, and the exact opposite of what may be claimed by local feedback proponents.

An amplifier with a single gain block and one global feedback loop will, provided it has reasonably good open loop linearity, simultaneously remove a significant amount of distortion from all stages at once, and there is no additive effect due to cascaded stages. This point is

rarely (if ever) mentioned.

6.0 - Amplification Circuit Delay

It is obvious that nothing in life is instantaneous. When a signal is applied to the input of an amplifier, there is a delay before the amplifier can react to the change, and this is determined by the speed of the devices used. Logic circuits typically have nanosecond delays from input to output, and this is also the order of delay one can expect before an amplifier as shown in Figure 7 will react to a change of input. According to the simulator, it takes about 5ns for the amp to respond to the fact that the input has changed - this is still using the very fast squarewave as an input. The output then swings in the appropriate direction at its maximum slew rate until the voltage at the inverting input again equals that at the non-inverting input. Once the voltages are equal, it takes about 220ns for the output to stabilise, settling so that the two input voltages are exactly the same. These times are very short - it takes the output 1.3us to change from +11V to -11V, so the 'reaction' time is close to negligible. It would be pointless to try to reproduce all the waveforms, so I suggest that you download the simulations. The files are in SIMetrix format, and are ready to run.

However, the above was all done with a signal that is much faster than the amplifier can handle, and it is more useful to examine what happens when the input slew rate is limited to something sensible. By adding a filter to the squarewave signal, the rise time can be limited to a somewhat more realistic value. A 32kHz, 24dB/octave filter was used, and this limits the output signal from the amplifier to 1.85V/us - well within its range, but still a great deal faster than any real music signal will create. Everything is now within the linear capability of the amplifier. The output is delayed by 46ns compared to the input, but this is inconsequential. Of more importance is how the amplifier reacts to the combined sine and square wave signal. It is not immediately apparent from the output, but in fact the sinewave is almost completely unaffected - the portion that would otherwise be cut off due to slew rate limiting now simply 'rides' the slope of the squarewave - if compared (after correcting for the level difference), the input and output are virtually identical - there is no evidence whatsoever of anything that could be classified as transient distortion - even with a 100kHz signal.

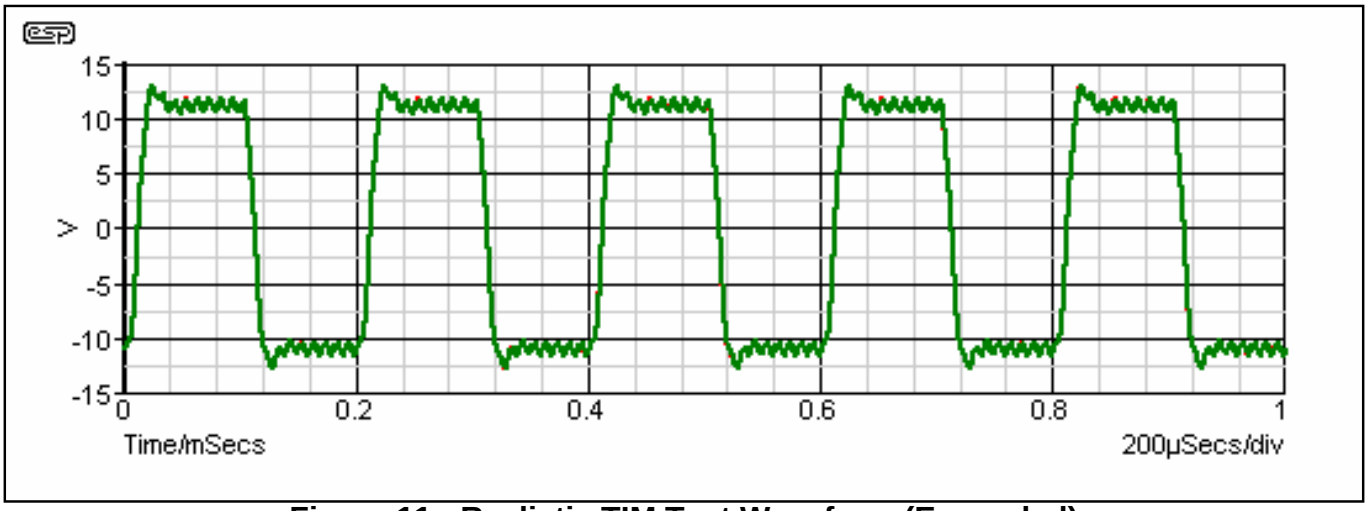


Figure 11 - Realistic TIM Test Waveform (Expanded)

There are two graphs in Figure 11 - green is the scaled input (increased in level to match the output) and red is the output signal. They are perfectly overlaid, indicating that the difference between them is very small indeed. Differences can be seen if the graph is expanded far enough, but the resolution of any oscilloscope will be such that the two waveforms will appear identical. The simulator can resolve details that are imperceptible with real test equipment. It is worth pointing out that the ESP sound impairment monitor (SIM) will detect the difference in real time using real world signals. Even the modified waveform of Figure 9 does not represent *any* signal that can be recorded or produced by any musical instrument (or combinations thereof).

Once the combined input signal is made sensible, the difference between the input and output signals can be seen, and it is primarily the result of the time delay (mainly phase shift) through the amplifier circuit. By using the SIM technique (measuring the difference voltage between the two inputs), all that remains is a residual signal that correlates with the gain of the amplifier at the frequencies used. The residual signal contains no non-linearities whatsoever, and is shown in Figure 12. The input stimulus this time is a 5kHz squarewave, filtered at 24dB/octave with a filter having a -3dB frequency of 32kHz. Superimposed on this is the same 100kHz signal used for the previous tests. The signal shown is the difference between the inverting and non-inverting inputs of the amplifier. Some of the signal shown is the result of the amplifier's error correction stage (the long-tailed pair) and VAS over-reacting slightly, and is also affected by the amplifier's total propagation delay and phase shift.

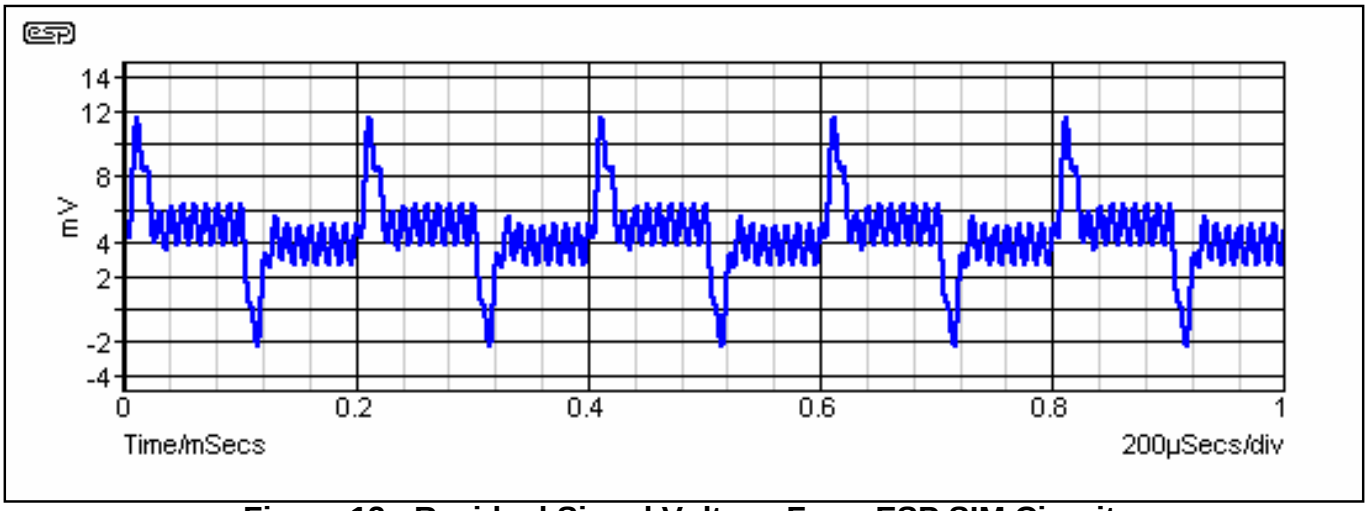


Figure 12 - Residual Signal Voltage From ESP SIM Circuit.

The important point here is that the amplifier *must* be maintained within its linear range. All amplifiers, including 'zero feedback' designs, can be forced outside their linear range. The whole idea of an amplifying circuit is that it should be linear, so no test signal should be used that dramatically exceeds the parameters of those of a normal source (such as music). To do so highlights 'problems' that do not exist in reality, so their inclusion is pointless at best, and grossly misleading at worst. The test signal used to obtain the above waveform is still a savage test - far more so than any music signal will produce, and deliberately much closer to the amplifier circuit's own limitations.

One can also measure the difference between an amplified version of the input signal, and that passing through the real circuit. In this case, the error signal is ~58dB down from the amplifier output, but is mainly the result of phase shift and very small gain errors - it is not a non-linear (distortion) component. At the upper test frequency of 100kHz, the amplifier has an open loop gain of only 470. With a design gain of 11 and an open loop gain of 470, the *actual* gain works out to be about 10.75 - this (as well as phase shift and DC offset) will always cause some error. It is important to understand that this is simply a small gain error, and does not contribute towards non-linear distortion.

These same tests have been performed (using test equipment, not the simulator) on various amplifiers shown in the project pages, with very similar results to those described above. There remains no evidence that any sensibly designed amplifier cannot keep up with recorded music, regardless of genre. The most common real amplifier fault one is likely to encounter in the listening room is clipping. Since clipping forces an amplifier out of its linear region, the main concern is how long the amp takes to recover from the overload.

This is a test I always perform, and a well behaved amp should recover almost instantly. The simulated circuit of Figure 7 recovers in less than 500ns for both positive and negative peaks, clipped with an input signal +4.5dB above the maximum level at 10kHz. Normal maximum level is 1.75V, and the input was driven with 3V (both are peak input levels). Recovery from clipping is not substantially affected by the input level. The recovery time is substantially less than the

sampling rate of a CD (44.1kHz = 22.675us), so the loss of information is only a fraction of one sample. Most amplifiers *should* recover in a few microseconds. If they do not, then there is a problem with the design.

7.0 - Local vs. Global Feedback

There is a constant argument regarding the benefits of local rather than global feedback. The following two circuits show an essentially similar design, but one uses two stages with only local feedback, while the other has been optimised for global feedback. The value of the feedback resistor was adjusted to give identical overall gain, in this case 40 (32dB). Conventional transistor current sources were used in the second circuit, the only difference being the use of a voltage source instead of a pair of diodes. The difference is minimal.

The strange resistor values in the global feedback circuit were a matter of expedience, and were used to set the gain and collector currents so that both circuits were run with the same current and collector voltage. Normally, one would not go to so much trouble, but for this experiment it was important to eliminate as many variables as possible.

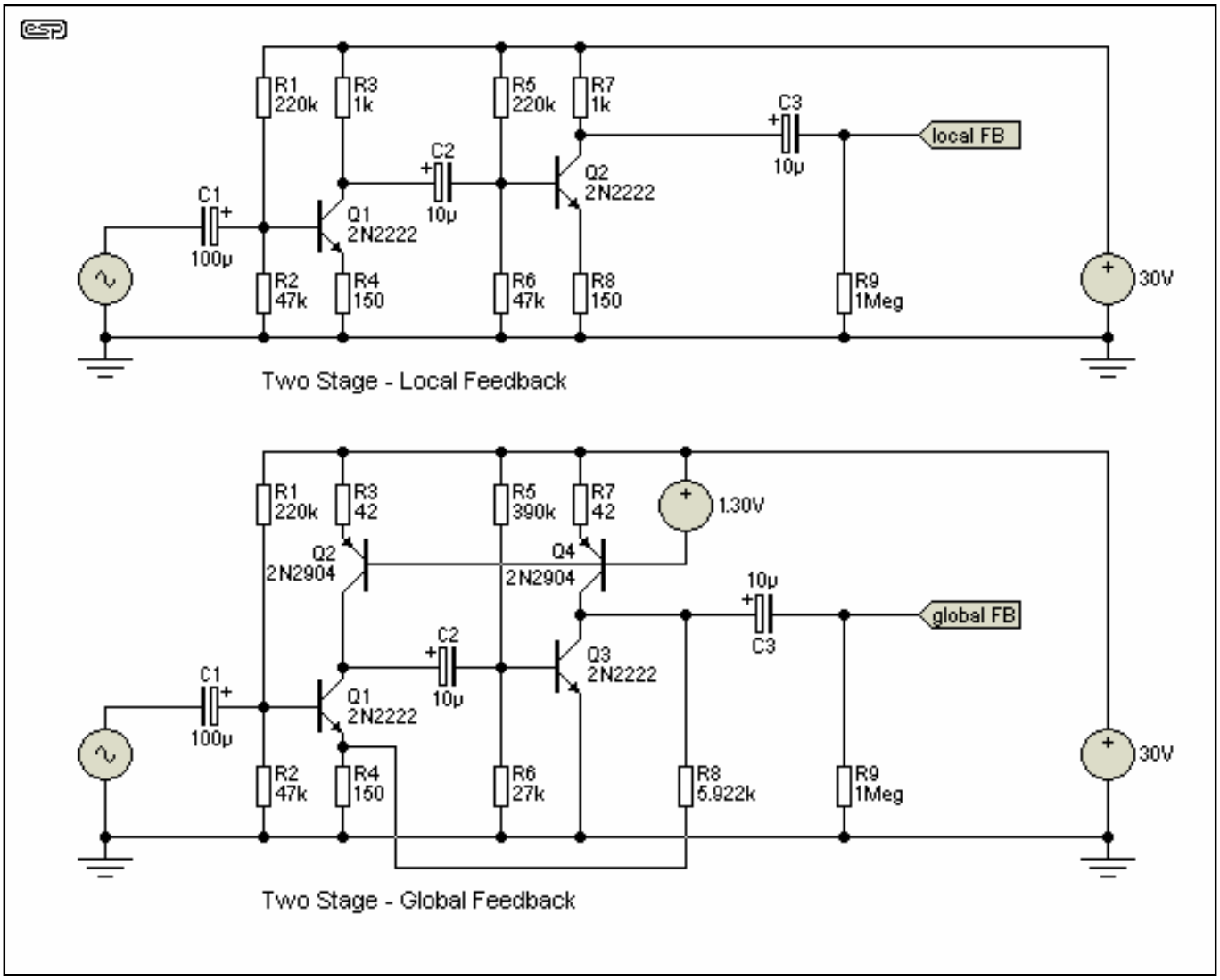


Figure 13 - Test Circuits for Local & Global Feedback

There are some quite surprising results. The global feedback circuit has less than half the distortion of the local feedback version (0.035% vs. 0.082%), but there are many other advantages as well. Input impedance is higher, output impedance lower, and global feedback makes the circuit faster and with better frequency response. The full listing is shown in Table 4, and it is obvious that global feedback is superior to local feedback in every respect.

Parameter	Local FB	Global FB
Distortion	0.082%	0.035%
Input Impedance	17kΩ	37kΩ
Output Impedance	1kΩ	<26Ω
-3dB Bandwidth	10.4MHz	24.7MHz
Open Loop Gain	40	286,000
Rise Time	28.8ns	11.9ns
Fall Time	32.3ns	10.6ns

Table 4 - Local vs. Global Feedback

One would think that there *must* be a down side. Something so simple can't possibly be that much better without a sacrifice. Can it? Yes, it can. Figure 12 shows the spectrum of the two circuits. As you can see, global feedback reduces all the harmonics, and the 'nasty' third harmonic is reduced far more effectively by global feedback than local. Not what you might expect, but there it is.

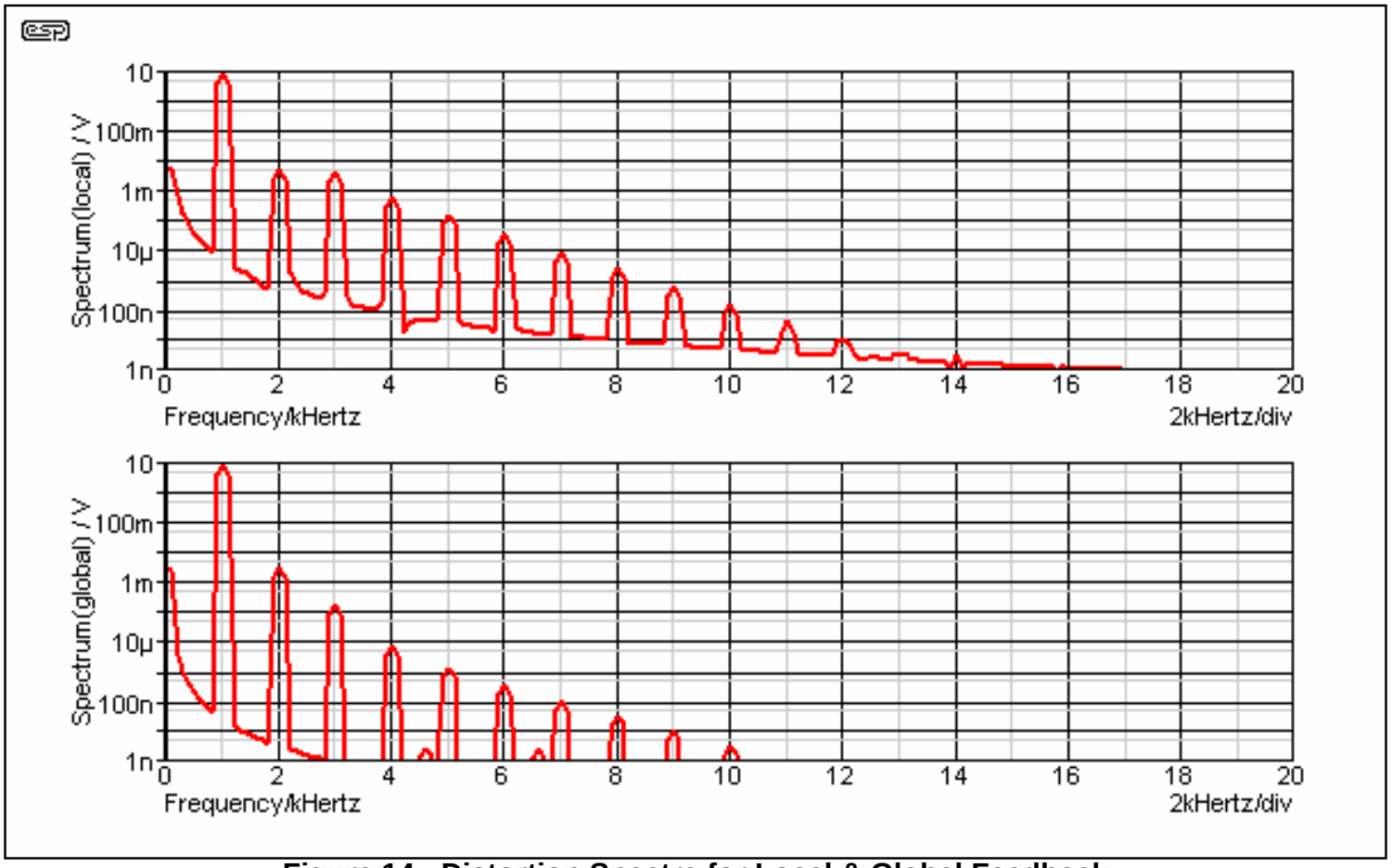


Figure 14 - Distortion Spectra for Local & Global Feedback

On the basis of this, global feedback wins on all counts. If you were to build the two circuits, you would find that the overall situation will not change, although some of the parameters will. This is due to component tolerance, but will not affect the final outcome.

It is notable that global feedback works best when there is lots of it. The claims that global feedback should be used in moderation are just silly, and have never considered the reality of good circuit design. The higher the open loop gain the better, but eventually you will run into stability issues, so some form of frequency compensation becomes essential.

Designing for stability and high open loop gain can be a challenge at times - especially for power amplifier circuits. However, when it is done (and done properly), there is no doubt that global feedback lives up to all the claims for it, with virtually no down side at all.

Well, there *is* a down side, but we have to look for it and know what we are looking for. Because nearly all opamp style amplifiers require a dominant pole capacitor to prevent oscillation, this causes a loss of open loop gain as frequency increases. Less open loop gain means less feedback, so upper harmonics are not attenuated as well as low order harmonics.

This could lead one to imagine that the application of feedback has indeed increased the level of the high order distortion components, but in fact it has done no such thing. What *has*

happened is that the feedback at higher frequencies is insufficient to reduce the upper harmonics as effectively as those at lower frequencies. Their amplitudes have been reduced, but not by as much as the low order harmonics. High order distortion products can therefore be seen extending out well past the audio band, at a similar level to the lower order components. For example, we may find that the tenth harmonic is reduced to perhaps -80dB, but the eleventh is only at -81dB, the twelfth at perhaps -81.5dB and so on.

Examining the spectrum may show that the relative levels of all subsequent harmonics remain at much the same level, well beyond the audio band. In this respect, the addition of feedback can easily be blamed for all the upper harmonics. The problem really lies with the gain of the amplifier, which rolls off the frequency response too early.

This effect can be seen plainly with crossover distortion on an unbiased P101 MOSFET power amp. At 1kHz, there is virtually no visible crossover distortion, even when the output stage has zero quiescent current. At 10kHz, the distortion is clearly visible on the oscilloscope, even though it is not audible with a single tone (the 3rd harmonic being at 30kHz). Needless to say there is zero visible (and almost zero measurable) crossover distortion at 10kHz once the amp is biased correctly, but this highlights the open loop gain issue. At 10kHz there isn't enough feedback to be able to correct the crossover distortion, but there is more than sufficient at 1kHz.

The solution is simple enough - make sure the amp is as linear as possible before feedback is added (which in the above case means setting the bias current correctly). While there is no doubt that a wider open loop bandwidth is beneficial, this must never be at the expense of amplifier stability. A small amount of distortion at the uppermost frequency range is far better than an amp with marginal stability - oscillating amplifiers definitely don't sound very nice at all.

8.0 - Conclusion

Read any articles about distortion you may come across (including this!) with care. Like death and taxes, distortion is inevitable, however it can be minimised with careful design and a proper understanding of how feedback can be used most effectively to ensure that distortion doesn't spoil your listening experience.

Loudspeakers contribute far more distortion than the vast majority of amplifiers, but some forms of distortion can be very intrusive - especially crossover distortion in transistorised amps. Fortunately, it is a simple matter to design an amp using sensible circuitry and modern transistors where crossover distortion is (for all intents and purposes) non-existent. Total harmonic distortion figures of well below 0.1% at any normal power level from a few milliwatts to several hundred Watts are easy to obtain. The distortion of most modern amps will contain only a few low order harmonics at all power levels up to the onset of clipping.

Of far more concern is the addition of distortion to the recording, either deliberately or by

accident. Nothing that you do in your home system can eliminate that - once a signal is distorted, you are basically stuck with it.

To avoid the use of global feedback based on so-called 'research' is most unwise. As demonstrated above (and by many others), correctly used, global feedback is as close to a panacea as we are ever likely to find. The idea of any hi-fi system is to reproduce the source material as faithfully as possible, and to deliberately add distortion to everything you hear (due to amplifier deficiencies) because it sounds 'nice' is simply not high fidelity. If that is what you want to hear then there is no problem with that, but by adding so much additional material (by way of harmonics and intermodulation) you have a tailored sound system, not a hi-fi.

Harmonic distortion and intermodulation are linked together, so much so that it is virtually impossible to have one without the other. By ensuring that each element in the amplification chain is as linear as possible, you minimise both THD and IMD, both of which are easily demonstrable. This is a far better option than trying to minimise TIM, the very existence of which has been called into question countless times since it was 'discovered'.

Finally, I have included a pair of simple circuits that can be used to create distortion. Testing these using my workshop speaker system, the distortion of a 400Hz sinewave was (just) audible at < 0.5%. This same level would be inaudible on most music, being primarily low order as seen on the residual of my distortion meter. It is probable that had I used headphones or a better speaker system, low order distortion would be found to be audible at lower levels, but this simple test shows just how revealing a sinewave really is. While those trying to 'prove a point' will claim that a sinewave test is too simple and reveals little, this is obviously *not* the case. As noted above in Section 2, a sinewave is not an easy test at all, and anyone who claims otherwise is seriously mistaken.

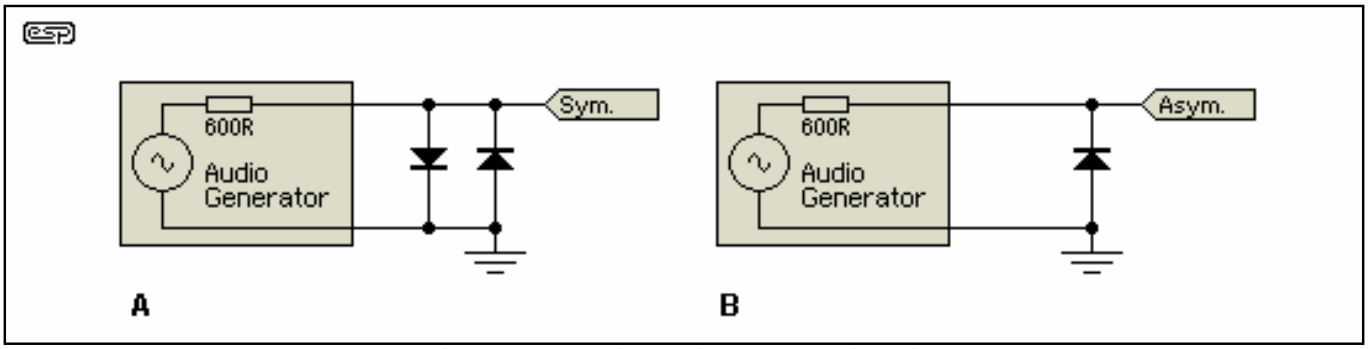


Figure 15 - Distortion Test Circuits

The circuits shown will need to be carefully tweaked to suit your test equipment and amplifier, so consider them to be more of a general idea than a definitive test circuit. The amount of distortion for both A (symmetrical) and B (Asymmetrical) is adjusted by varying the input level, and no attempt has been made to level match the distorted and undistorted signals. The distortion itself is sufficiently prominent that full blind AB testing is not needed to get a general idea, but would be essential for a scientific study. The day after I did these tests, a friend came to my place, and I repeated the test with him. The distortion meter was disabled so we had no

visual cue, and we arrived at almost exactly the same result with both test circuits.

Be aware that you may find that you can't hear any distortion until it is greater than the 0.5% I measured. Try moving around (even a few centimetres or so will be enough). Why? When listening to a steady tone, standing waves and reflections can combine to make a single frequency much louder than it should be, or almost inaudible. This effectively changes the distortion spectrum, making it sound much greater or less than the actual value. While this effect may have contributed to my hearing only 0.5% distortion on a sinewave, I did move around to make sure that the distortion was audible in more than one position. I neglected to measure the sound level when the test was done, but it would have been around 75dB SPL - any louder becomes very irritating.

For around 0.5% distortion measured and using an asymmetrical diode clipping circuit (B), the harmonic levels will be pretty close to the following (note that all harmonics are referenced to the level of the fundamental, all voltages are peak) ...

Fundamental 400Hz	448mV	0dB (reference)
2nd harmonic	1.35mV	-50dB
3rd	942uV	-53dB
4th	599uV	-57dB
5th	348uV	-62dB
6th	167uV	-68dB
7th	68uV	-76dB

It is probable that only the first few harmonics would have been audible, since those above the fifth are approaching my hearing threshold. Even this may be wishful thinking - anything above the third is well below the ambient noise floor.

Sine waves are too simple to use as a test? We think not!

9.0 - References

1. Negative feedback doesn't always decrease amplifier distortion! John Atkinson stereophile.com/news/10065/
2. Radiotron Designer's Handbook, F. Langford-Smith, Fourth Edition, 1957, pp603-616
3. Zero Distortion? Ian Hickman Electronics World, March 1999, pp224-228
4. New Methodology for Audio Frequency Power Amplifier Testing ... Daniel H. Cheever - University of New Hampshire, December 2001

10.0 - Simulation Download

SIMetrix Simulation Files Right click, and select "save link as" from the menu.

To view or run these simulations, you need the SIMetrix simulator on your PC. The freeware version of the simulator can be downloaded from Catena Software.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2006. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 04 May 2006

Elliott Sound Products | **The Whys And Wherefores of Guitar, Bass and Keyboard Amps**

By Rod Elliott

Page Last Updated 04 Feb 2001

ESP Home Page

Articles Index

Contact The Author

Index

- Guitar Amplifiers
- Powering Up
- Bass Amplifiers
- Keyboard Amplifiers
- Odds and Ends

Guitar Amplifiers

IN THE STUDIO and on stage, instrument amplifiers are vitally important to the musician. Specialised equipment is a necessity, and the instruments themselves are becoming more and more demanding of the amplifier-speaker combination with which they are used,. The size and power ratings of such systems has grown considerably in recent years; it is not ,uncommon to find instrument amps today, which are more complex and larger than a band's entire PA system of a few years ago. As one of the most controversial pieces of amplification ever produced, we shall first take a look at the guitar amplifier, which has evolved from a general purpose "as long as it works" device, to a sophisticated piece of equipment - at least in the professional sector, where the so-called "universal" amplifier is no longer relevant.

Most guitarists will say that, with a given amplifier, a certain volume level is required to obtain the right sound. While this is to a point psychological, there are very good technical reasons to support this statement. The "right sound' usually means a fairly high level of harmonic distortion, both from amplifiers and speakers - distortion which gives much needed assistance to an electric guitar by providing the missing harmonics and increased sustain. A certain amount of sustain is obtained by acoustic feedback, particularly with semi-acoustic guitars, however the majority is the result of the amp being driven into "clipping", or distortion. This has the effect of maintaining the level at a (more or less) constant level for much longer than would normally be the case (Figure 1).

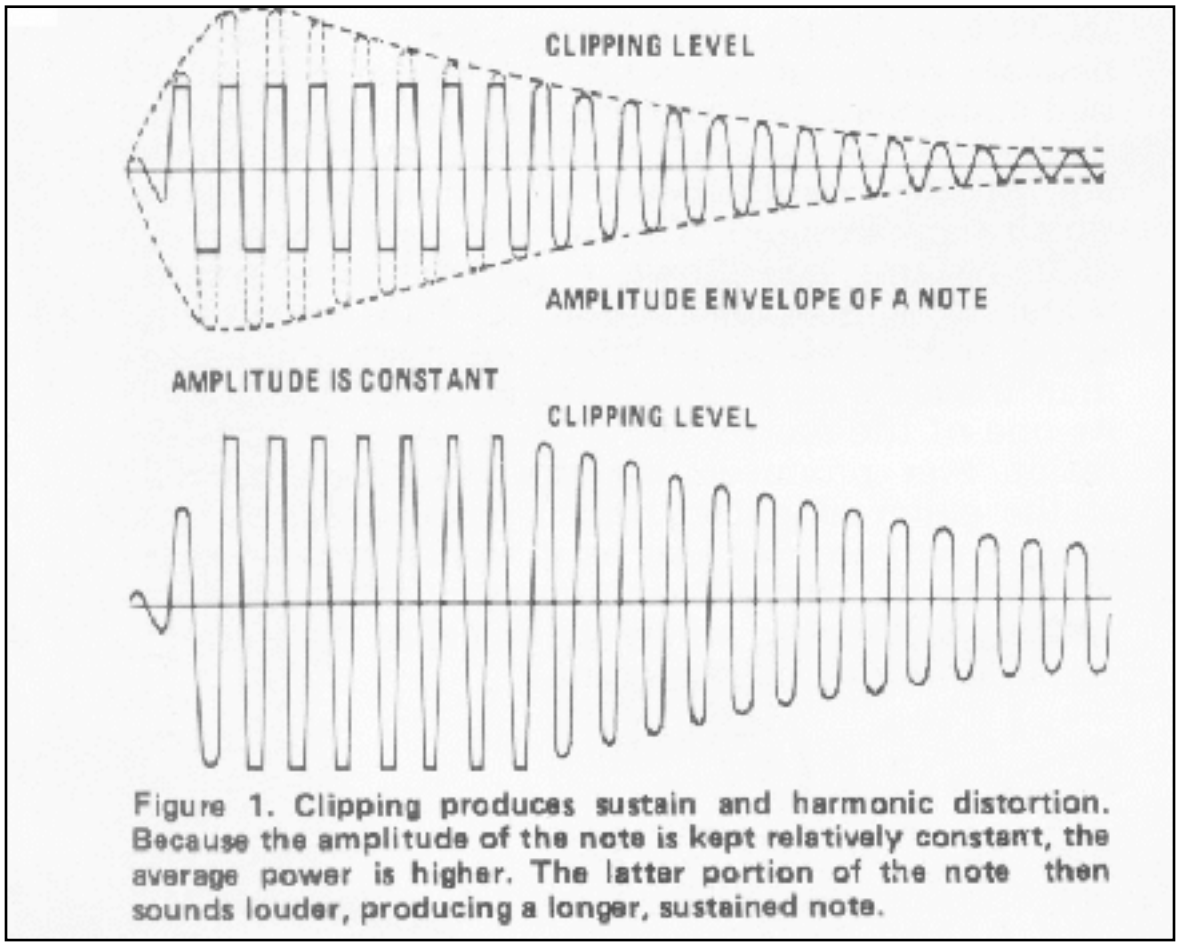


Figure 1. Clipping produces sustain and harmonic distortion. Because the amplitude of the note is kept relatively constant, the average power is higher. The latter portion of the note then sounds louder, producing a longer, sustained note.

If an amplifier is overdriven, the sound produced is usually considered objectionable, so special measures must be taken to obtain the desired characteristics.

Firstly, there is the bandwidth restriction of the speaker cabinet. For the low-end this may be achieved by using an open-back box, to allow bass cancellation, or to contain the speakers in an enclosure of limited internal capacity. High-end roll-off is usually a characteristic of the speakers themselves: speakers - without "whizzer" cones or aluminium domes will generally sound smoother because of their limited high frequency response. A broad peak in the speaker's response at 2kHz to 4kHz adds "presence" and bite, and the response should roll off beyond that point. It is preferable that frequencies above about 7kHz are not reproduced at all,

as this will exclude the 7th harmonic (which is discordant) of all strings of the guitar. Damping or venting of guitar speaker enclosures is usually avoided like the plague because damping will lower colouration, a desirable feature in musical instrument amplification. Without colouration all plucked string instruments, for example, would sound much the same.

- **Vented boxes are avoided because they emphasise (or allow to be reproduced) the lower frequencies - these are usually not desired, since they tend to muddy the sound. This is especially true for "heavy metal" playing styles, since severe amp overdrive is commonly used to achieve the sound, and even small amounts of low frequency material (such as string handling) will cause large amounts of bass "waffle".**
- **The majority of guitar speaker boxes (as noted above) are either open backed, or are relatively small sealed enclosures. There are naturally exceptions, but even 20 years after this article was first written, are still uncommon.**

Additional bandwidth restrictions are usually designed into the output transformer of valve amplifiers. Transistor amplifiers usually do not have an output transformer, so this method of filtering is precluded. The effect of all this manipulation of response is to "clean up" the distorted output waveform of the overdriven amplifier - the sound will become less harsh, and the transition from overdriven to "clean" becomes less apparent.

In fact in a valve amp the clean or "no distortion" state exists only when there is no input. Distortion is present even at low levels, and increases with the input signal.

The above situation is in contrast with the majority of transistor amps, which generally have very low distortion up to the point of clipping, after which distortion rises rapidly. Maximum distortion is the same as for a valve amp - i.e. a square wave, but the transition is such as to be very noticeable; this can be objectionable when the guitarist is playing clean but for the odd note or chord which overdrives the amp, causing a marked change in tonality (Figure 2).

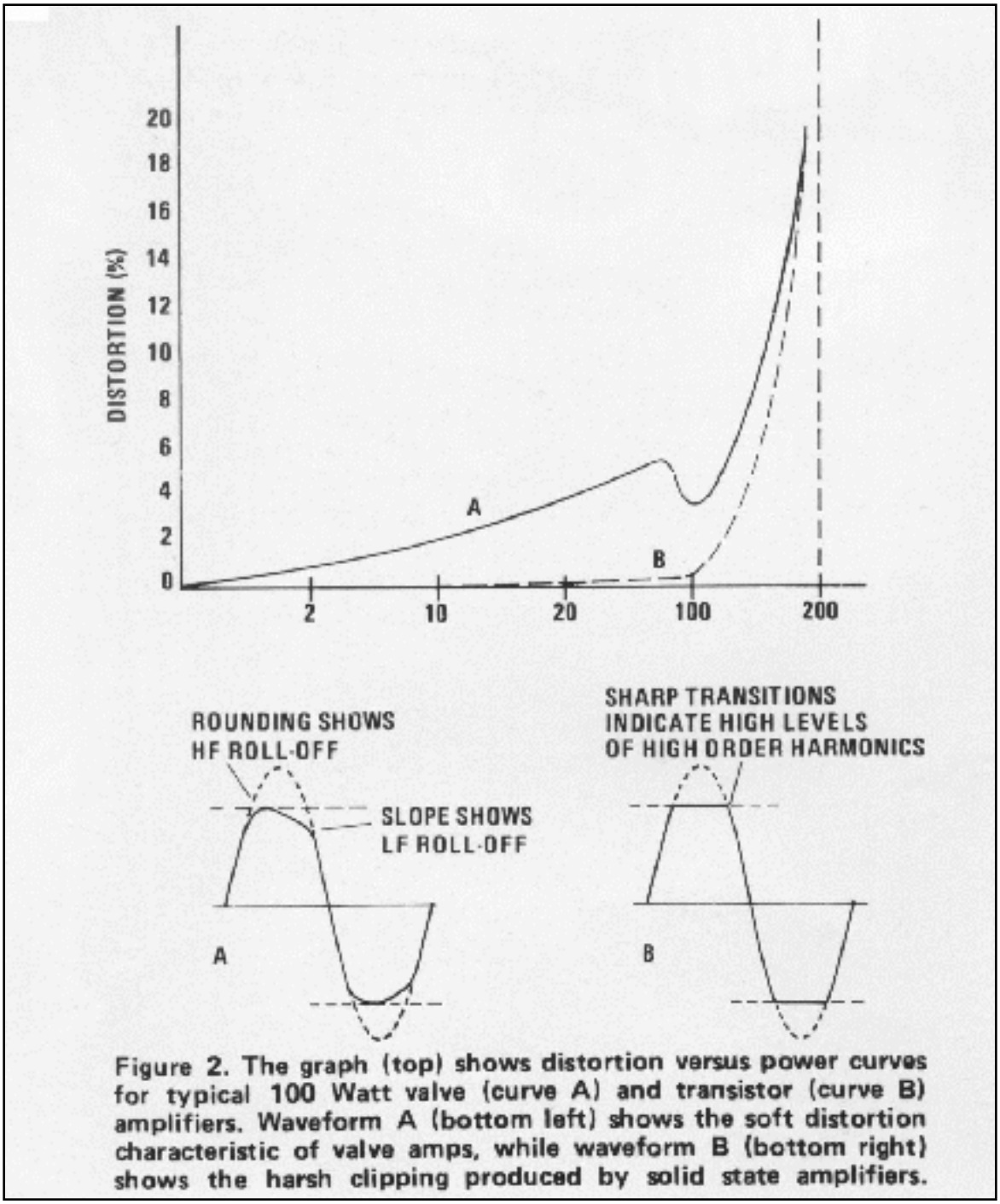


Figure 2. The graph (top) shows distortion versus power curves for typical 100 Watt valve (curve A) and transistor (curve B) amplifiers. Waveform A (bottom left) shows the soft distortion characteristic of valve amps, while waveform B (bottom right) shows the harsh clipping produced by solid state amplifiers.

Output impedance and "dynamic output" are other major contributors to the sound of a guitar amp. Output impedance, in this instance, has nothing to do with speaker "ohmage", but is the actual source impedance of the amplifier, as seen by the speaker. In a valve amp, the source impedance may be as high as 200 ohms for a nominal 8 ohm output. This allows the speaker freedom to exhibit its own natural resonances, plus the colouration provided by the enclosure itself. A low output impedance on the other hand, will repress colouration by damping the speaker, in much the same way as would fibreglass in the cabinet. This tends to make the amp sound flat and lifeless, lacking the subtle colouration and tonality which makes a fine musical instrument - and after all, a guitar amplifier is as much a musical instrument as the guitar itself. Transistor amps usually have a low source impedance, typically less that 1 Ohm.

- **It should be noted that the power rating for guitar amps is (very) commonly quoted at 10% distortion. This means the amp is typically about 80W at the onset of clipping (usually taken as the "real" maximum power level). This is justified on the basis that few guitarists will operate their amps with no distortion at all. This assumption is not necessarily valid, but has prevailed for a very long time, so is unlikely to change now.**
- **The source impedance of valve amplifiers is generally lower than that quoted above, and in fact it will generally match the speaker impedance (assuming no feedback is applied). I appear to have goofed in the original article.**
- **From the very first guitar amplifiers I ever built using transistor output stages, I used current feedback to increase the output impedance of the amp. This restores much of the "valve sound", and provides a simple but effective method of providing short-term short circuit protection (for example while plugging 6.5mm phone jacks into their sockets while the amp is on). This technique has now been adopted by nearly all guitar amp manufacturers. Regrettably, this technique does little to improve dynamic output. Output impedance of my amps was typically 200 Ohms.**

Dynamic output in this context means the ability of an amplifier to deliver power to the load (the loudspeaker). Because of the high source impedance and low efficiency of valves, output transformer and the power supply, a valve amp is capable of developing approximately 75% of its rate power into double its rated impedance, and up to 50% of rated power into four times rated load. This means the amp can provide relatively constant power regardless of the variations of impedance with frequency which are a characteristic of speaker loads.

In contrast, a transistor amp can usually deliver only 55% or so of its rated power into double rated load, and about 30% into four times rated load. These variations may not appear to amount to very much, but they do make a big difference in terms of spectral balance and dynamic performance. Such is the difference in fact (and this also applies to valve hi-fi amps, bass amps, etc) that a 100 watt valve amp will generally sound as loud as a 150 watt transistor amp.

The technical reader with an understanding of dB and the logarithmic response of the human ear may be sceptical about the foregoing paragraph, but differences of 1dB are audible when one has a suitable point of reference; in musical groups, where the balance of instruments is critical, such differences are very noticeable.

Powering Up

The level of sound produced by a guitar amp is usually way out of proportion to the quoted power rating. For example, a 100 watt guitar amplifier played normally (for a rock guitarist) will produce an average sound pressure level of about 120 dB at 1 metre. Try that with your hi-fi and you will most likely have speaker cones (and/or your neighbour's projectiles) all over the listening room floor. The reason is quite simple. Because the amp is being overdriven, peak

limiting will occur - clipping of the output waveform - which reduces the peak to-average power ratio of the program material. The peak to average ratio of a guitar signal is about 20 dB; that is, the peak level as a note or chord is struck will be 100 times higher than the overall average level. An overdriven amplifier may reduce this ratio to 2:1 (3 dB) or less, and as a result the average power is higher.

A further power increase results from the fact that a severely clipped waveform will approach a square wave, which has twice the power of a sinewave of the same amplitude. The result is a 100 watt amplifier delivering an average power of perhaps 170 watts short term (i.e. maximum level indicated on a VU meter), and up to 130 watts averaged over a 5 to 10 second period.

Compare this with an undistorted amp handling program with a peak-to-average ratio of 100:1 (20dB). The long term average power will be in the order of 2 to 5 watts with peaks of 100 watts. With mild distortion, say less than 15%, the average power climbs rapidly to around 25 watts; increase the distortion further and the power climbs again until, with maximum overdrive, a 100 watt amp will be delivering 200 watts!

- **An important point was raised by a reader, who said that the above explanation only applies if the power supply remains at a constant voltage with the increased load. This is very true, and in reality, most don't, so the maximum power is limited by the capabilities of the supply. Valve amps are usually worse in this respect than transistor amps, and the maximum power at full clipping may only be 150% of the rated power.**
- **Despite this, the average SPL (Sound Pressure Level) is maintained at a high level, and the effect is that it sounds louder than one might imagine it should. This is similar to using compression (as done with ads on TV). The maximum SPL is not increased, but because everything is at maximum, it appears to be louder than it really is.**

So loud, in fact, are many popular guitar amps, that guitarists are now experimenting with methods of obtaining "the sound" at lower power levels. There has been a trend for some time now towards smaller amps and/or some provision for simulating amplifier overload. The latter is best achieved by using a master volume control and some form of clipping circuit, which should ideally be placed (electrically) between the tone controls and the power stage. A clipping circuit placed before the tone controls will not sound like amp overload because of the large amount of treble boost built into guitar amps. A clipping circuit so placed will produce only the dreaded fuzz box sound, once much loved but now generally disdained because of its harsh unnatural sound.

- **Having said that, there seems to be a resurgence of old analogue technologies in all areas of electronic music, and this includes the lowly fuzz box. There have also been moves to try to emulate the asymmetrical clipping characteristics of valves (some dating back many years).**
- **One of the most innovative (and complex) attempts to recreate the valve sound with solid state electronics is the Vox AC30 Simulator - a contributed article on the ESP Project Pages.**

The addition of a graphic equaliser set, now incorporated in many brands such as the Zoom (above), greatly expands the tonal variation possibilities of an amplifier.

Bass Amplifiers

Many of the requirements for guitar amps also apply to bass amplifiers, in particular those factors relating to colouration and harmonic distortion provided the latter is not obtained from amp overload. A high output impedance and high dynamic output are also desirable features.

One of the major differences is the power requirements. Whereas a guitar amp will usually be overdriven, a bass amp should not be; the sound will become muddy and indistinct, lacking punch and definition. Bass is, by its very nature, difficult to reproduce. The open E string for example, has a fundamental frequency of 41.2 Hz, a difficult frequency to reproduce even given ideal circumstances, let alone at high power and with enclosures which must be able to be moved around! As a result, few bass systems have appreciable output below about 80Hz - and this includes the large "W" bins used for PA systems. Up to four times as much power may be needed at 40Hz to match the acoustic output at 400Hz (Figure 3).

Because it is undesirable to drive a bass amp into distortion, and the bass should be able to equal the guitar in volume, much more power is required for bass than for guitar. In fact if we assume a guitar amp rated at 100 watts and operating at an average power of 130 watts, and assume a bare minimum bass peak-to-average signal ratio of 10:1 (10 dB), our bass amp

would need to be at least 1000 watts to match the guitar amp without producing audible distortion! A certain amount of clipping of transients is permissible, and in a well designed system will not be audible because of the masking effect of the rest of the band. It will, however, be very audible on a bass solo if the same volume is used.

Harmonic distortion in the bass region is usually very high, especially in the speakers. Some are capable of a total harmonic distortion approaching 100%! This "impossible" distortion level is caused by speaker "doubling" - the result of a combination of speaker-cone breakup, voice coils which actually leave the magnetic gap and poorly designed enclosures. The main faults in enclosures are truncated horns, excessive horn flare rates, badly vented cabinets, and poor internal bracing of cabinets which are often simply too small for the speaker. Intermodulation distortion arises from the large excursions of a speaker cone at bass frequencies which cause the voice coil to partially leave the magnetic gap. When this occurs, efficiency is lost and distortion appears at the peaks of the waveform. Any superimposed high frequencies (harmonics) will also be distorted.

- **It must be noted that many of these problems have been reduced with modern loudspeaker drivers, whose excursion is often prodigious. Linear cone travel of +/-10mm or more is not uncommon now, and there are many very efficient speakers that are specifically designed for use in small enclosures. Coupled with the very high powers that are common today, it is possible to make a bass amp that will faithfully reproduce down to 30Hz in a cabinet that is still portable.**

It should now be apparent that a 200 watt bass amp, typical of many systems, is quite inadequate for a reasonably loud rock band, and in fact is often barely adequate for a club band; it will usually be pushed so hard that it will be clipping for up to 60% of the time. If we now connect the above amplifier to a speaker system whose efficiency, especially at the lower frequencies, is not very high, add equalisation to provide a measure of bass boost, then the amp overload problem becomes even worse.

The obvious solution is to use more speakers in larger enclosures - and more of them - driven by even more powerful amplifiers. An effective alternative is bi-amping; an electronic cross-over separates amp-speaker combinations which may be more easily optimised for the frequency range being covered. This approach has been used with some success, and although it is more costly it will eventually become a standard technique. The advantages of a bi-amped system are lower intermodulation distortion, and greater acoustic output for the same total power as a conventional amp/speaker combination.

Unfortunately there are many trade-offs which must be made, both for financial and physical reasons, and many of the better combinations are precluded by prejudice. For example, the average. bass players' opinion of 300mm (12") speakers is not printable.

A variation on bi-amping is to use separate amps and speakers for a stereo bass guitar. In this

arrangement the bass and treble pick-ups on the instrument are brought out separately and fed to separate amps. This technique has been around for many years but as few instruments are wired for stereo, most bass players have not had the opportunity to experiment with it. Any two pick-up bass can be easily modified however, and it is a trick well worth trying.

Keyboard Amps

A modem keyboard setup is probably the most complex and expensive item in any band, with the possible exception of the main PA system.

The basic requirements, at least insofar as the bass and midrange are concerned, are much the same as for bass amplifiers. Power requirements are of the same order, and similar problems involving loss of definition may occur due to insufficient power, poor enclosure design and a lack of understanding of the demands which are placed on a keyboard system. These include the extreme dynamic range (if pianos, the vast pitch range of synthesisers, and the special difficulties presented by such instruments as string ensembles, Mellotrons and Leslie type organ speaker cabinets (which are usually not loud enough unless miked).

- **Mellotron?? I'm not going to even try to explain this one (you wouldn't believe me unless you were there). Try a web search, you might well be surprised.**

The majority of intelligence and energy of keyboards is in the mid-range, the band of frequencies from about 180Hz to 1200Hz (Roughly F3 to D6) that cover the fundamental frequencies of the most used section of any keyboard instrument. If this section is not treated with the respect it deserves, it will be difficult to obtain a good, clean sound from the majority of instruments.

Obviously, the power required will depend on the volume at which the musician intends to play, but 100 watts would have been a minimum for all but the quietest bands. In a rock band this may have to be increased to around 1000 watts, with 450 watts on bass, 450 watts on midrange and 100 watts for treble.

An exclusively mid-range enclosure should not use speakers larger than 300mm (12') if a good transient response is to be obtained, and a flat frequency response is desirable.

Bass Boost	**Power**
Ref. 440 Hz Level	Ref. 10OW @ 440 H z

0dB	100 W
1dB	126 W
2dB	158 W
3dB	200 W
4dB	250 W
5dB	316 W
6dB	400 W
12dB	1600 W

Figure 3. The right column shows the power equivalent to the bass boost in the left column, referenced to 100 watts at 440 Hz (A above middle C).

Horn loading is an excellent way of achieving high efficiency, and if properly designed will provide low distortion, good frequency response and excellent transient response. The bass enclosure should not be horn loaded, because a horn which will work properly down to the lowest frequencies will be too large. Instead a reasonable sized vented enclosure should be used.

- **A better alternative today would be to use a sealed and equalised enclosure (see the EAS subwoofer project on the ESP Project Pages). This will give the best transient response, but requires much more power. On the positive side, the box will be very much smaller that would otherwise be possible.**

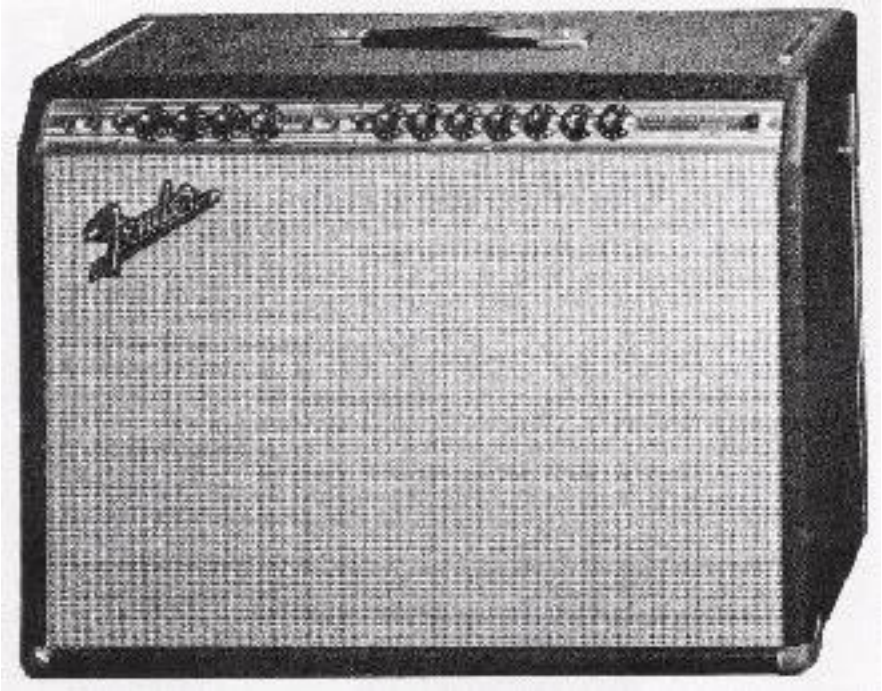
The high frequency end of the spectrum needs careful design, and must be protected from high power at high frequencies which the synthesiser, in particular, is capable of producing. (A synthesiser is the only instrument able to develop the same power at 20kHz as at 40Hz). Some form of protection for the high frequency drivers is therefore desirable, and becomes essential if H.F. horns are used. Such protection will also assist the musician in retaining his/her hearing (although from personal experience I can assure the reader that hearing retention is unlikely).

Passive crossovers are not recommended for high power keyboard systems except at the top end, from horns to "ring-radiators", where they will provide the most cost-effective solution without loss of performance. In general, though, their excessive power loss and the difficulty of optimising the crossover points makes them undesirable. A large keyboard system, then, is basically a very large hi-fi, with colouration being something to be avoided. It is not, or should not, be an extension of the instrument because of the number of different instruments which must be accommodated. Any colouration which is introduced should therefore be controlled by the musician.

Odds and Ends

Equalisation circuits are used in nearly all forms of amplification, and some have become very sophisticated. The old faithful bass and treble controls still exist, but the majority of mixers and amplifiers now have additional controls. Midrange and "bright" have been with us for some time, as has the presence control and graphic equalisers and semi parametric EQ are now being more widely used.

The range of effects available is enormous and growing; the echo, reverb and tremolo of yesteryear have been joined by esoteric devices like automatic double tracking ("ADT") digital delays, phasing, flanging, chorus, true vibrato (frequency modulation), peak limiters, compressors, harmonisers etc and etc.



The simplicity of an amp like the classic Fender Twin Reverb (above) still appeals to many musicians. The Twin Reverb is used by both rock and jazz guitarists, and is also popular with electric piano players.

More important now than ever before are leads and terminations. Cannon type connectors are replacing 6.5mm jacks on most equipment, and guitarists are changing over to Cannons because of their greater reliability. Heavy duty, low loss cables are being widely used, providing less noise and lower failure rate than standard cable, which responds poorly to having a 200kg rack of power amps wheeled over it.

- **Although there was (at the time of writing) a tendency for guitarists to use cannon ("XLR") connectors, this movement has seemingly died. This is a shame, since they are far more robust and reliable than 6.5mm jacks. There are, however, a number of professional locking 6.5mm phone plugs and jacks available, that did not exist when the article was written.**

Well, there you have it. If some of the points made here prove contentious, so much the better, for it will stimulate discussion and debate, both of which go towards a greater understanding of the problems of instrument and amplification. With a better knowledge of the needs of musicians, the technical persons amongst us will be better able to analyse the difficulties encountered by players, be able to pinpoint and maybe even eliminate them. On the other hand, a little technical knowledge will assist the musician in describing his problems, and will be very helpful to the technician or engineer who is expected to solve them, often without even knowing what they may be.

Rod Elliott has many years experience in audio electronics, especially in the pro-audio field. He was a partner in Bumett-Elliott Audio Developments, and has undertaken original design and development work on many different types of equipment. He has written reviews for international publications, as well as technical reports for many instrument-amplifier manufacturers. Rod's current activities (apart from writing for us) include running a four-track recording studio called 'Fly-By-Night, a part-time business, SKAManufacturing, and a full-time business, Elliott Sound Products. Rod also lectures in electronics for the School of Electronics and the Academy of Sound Recording Engineers.

Reprinted from the SONICS 1980 Yearbook

Additional notes and comments have been added where appropriate, but for the most part the article is as originally written. The byline above was true at the time, but is now way past it's "use by" date - except for Elliott Sound Products, which is very much alive (although it is part-time at the moment). Such is life. The scariest part is that this was over 20 Years ago!

PLEASE NOTE: The inclusion of two photographs of guitar amplifiers shall not be taken as an endorsement of the brands featured, nor shall this statement imply any criticism of same. They were included in the original article, and are thus reproduced here. No more, no less.

ESP Home Page

Articles Index

Contact The Author

Updated 04 Feb 2001 - Added extra info on guitar amp overload, and equalised boxes for bass and keyboard use.

Elliott Sound Products **VI Limiters in Amplifiers**

Phil Allison
Copyright © 2000

Return to Articles Index

Back to Home Page

Introduction

This is another of Phil Allison's contributions, and discusses specific experience that Phil has had with short circuit protection circuits in power amplifiers. These are commonly called VI Limiters, since they limit current (I) based on the voltage (V) across the power transistors.

As Phil will explain, not all manufacturers get these circuits right, and the sonic artefacts can be extremely unpleasant. In extreme cases, the very circuit designed (?) to protect the output stage can do exactly the opposite, and cause the amplifier to fail, and / or damage tweeters into the bargain. Have I ever mentioned in these pages that I don't like protection circuits? Hmmm, I thought so :-)

Testing Amplifiers To Their Limits (By Phil Allison)

Virtually any power amplifier would be quickly destroyed by a short on the output if there were no internal limit on the output current. This is because the power supply can deliver a short term power level much greater than the capacity of the transistors used in the output stage. The answer is some form of output current limiting to prevent damage under short circuit or near short condition on the speaker terminals. Many schemes have been used to protect the output devices in an amplifier, from fuses and current trips in the power supply to elaborate electronic circuits.

The most common technique used is known as "VI limiting". A pair of NPN and PNP transistors is connected across the drive signal to the output devices and fed from a resistor

network that senses both the voltage across and the current through them. When the combination of V and I, which is the power dissipation, is more than what can be handled safely then the protection transistors conduct the drive signal away just enough to prevent output transistor failure (see Figure 1).

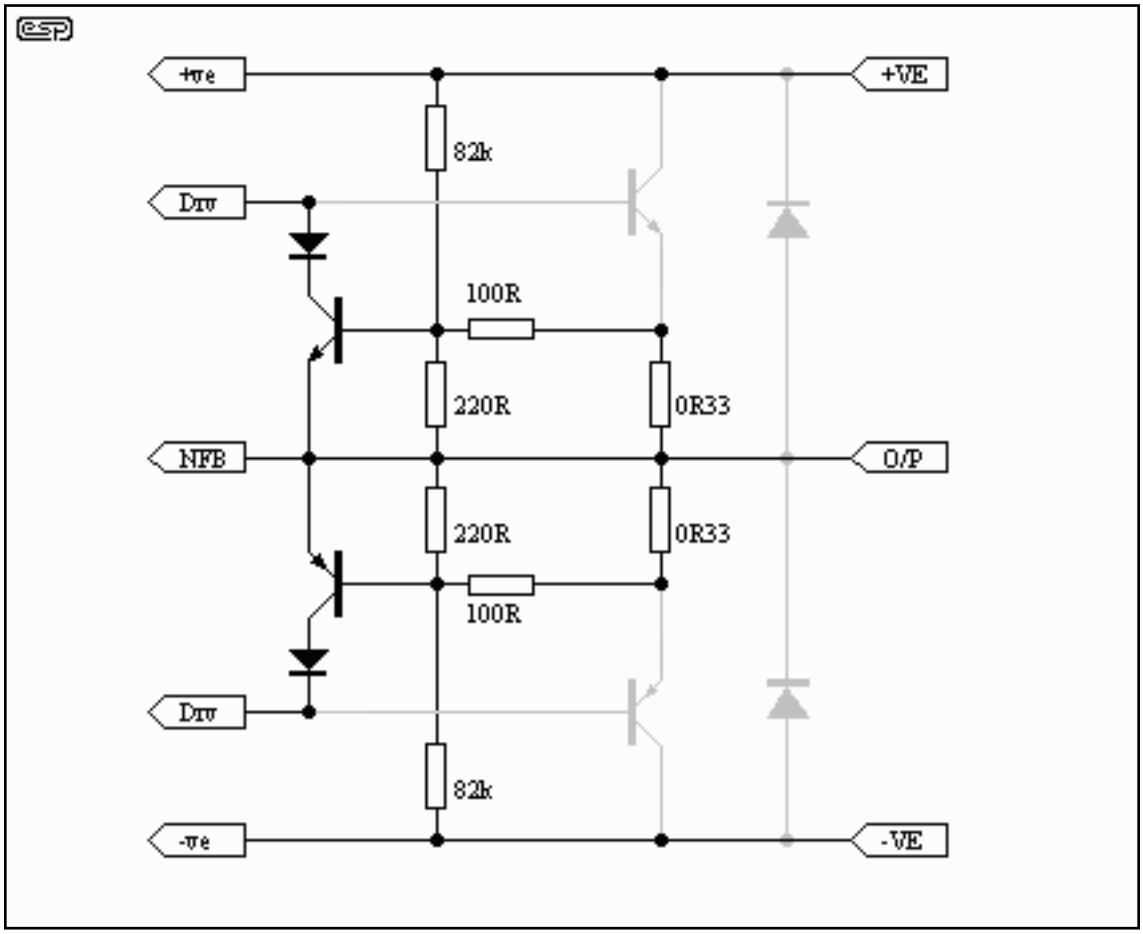


Figure 1 - Basic VI Limiter Circuit

The output transistors are shown in representative form, coloured grey, and may be darlington, compound pairs, or any combination thereof. The remainder of the amplifier is not shown, since this is largely irrelevant to the operation of the protection circuit. The diodes shown in grey are sometimes included (see Editorial Comment).

You may wonder why amplifiers need to deliver any current into a short circuit, after all surely this is a fault situation. True enough, provided all the amplifier ever had to do was drive pure resistors. The problem becomes more obvious when you consider that amplifiers have to drive loudspeakers. Real loudspeakers are a "reactive load", that is their impedance changes with frequency.

For example, a nominal 8 ohm 12 inch driver in a sealed enclosure will have a bass resonance at some low frequency and around that same frequency an impedance peak of 20 ohms or more. At frequencies above 1 to 2 kHz the impedance rises doubling each octave due to the inductance of the voice coil. Systems having tuned ports result in double impedance peaks in

the low frequency range only dropping to the "nominal" value at two points, in the middle of the peaks and at 250 Hz where the figure is often measured. An impedance that rises with increasing frequency is inductive and the opposite capacitive. While a loudspeaker's low frequency impedance peaks are caused mechanically, by the back EMF generated by the moving voice coil, they look to the amplifier exactly as if tuned circuits had been connected. An equivalent electrical circuit can be derived and contains large inductances in parallel with large capacitances. This is in addition to any passive crossover network that may be installed. The amplifier is required to drive this complex load at any frequency and level without distortion. Figure 2 shows the equivalent circuit of a single driver in a sealed enclosure.

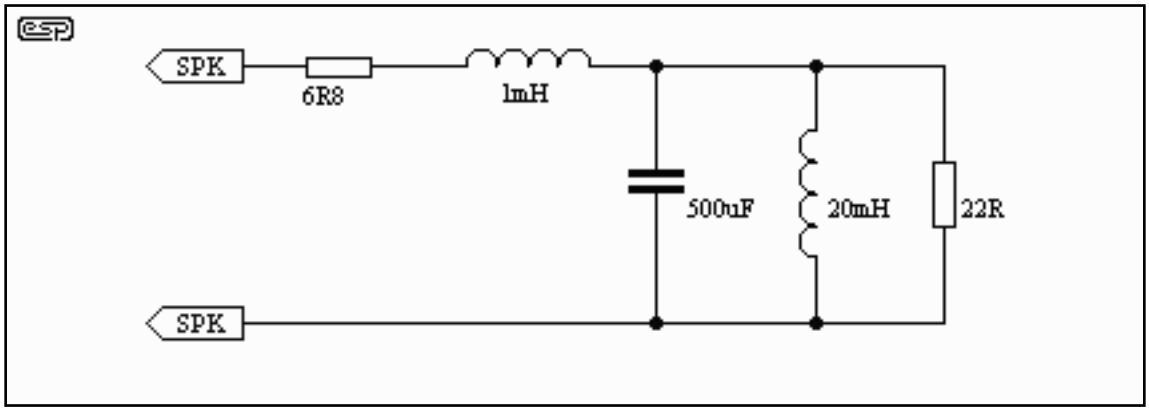


Figure 2 - Equivalent Circuit Of A Typical Loudspeaker Driver

Figure 3 shows the impedance versus frequency for the simulated speaker, and as can be seen, it looks very similar to the impedance response curve you see with many real loudspeakers. This (of course) is the whole point of the exercise.

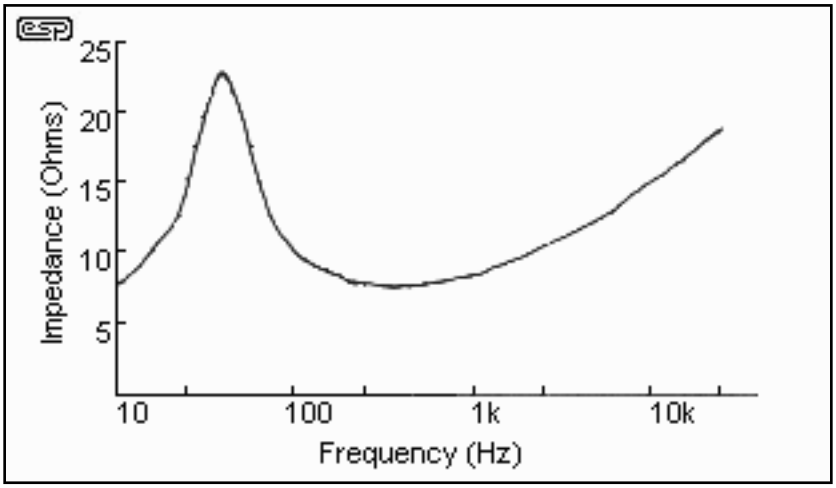


Figure 3 - Impedance Curve Of Simulated Loudspeaker

This can be built using the component values specified in Figure 2, or you may want to experiment further. Attempting to make the simulation "look" like a real loudspeaker system, including crossover networks and additional drivers is possible, but becomes quite complex. Since all speakers are different, this approach is neither suggested nor recommended, due to the cost of trying to re-create simulated versions of even a small number of systems.

Out of Phase Current

A load impedance that is inductive or capacitive has one important characteristic, the current and voltage are "out of phase". Having a current and voltage phase shift means zero volts and zero amps no longer coincide. The amplifier has to produce high current when its output voltage is zero. Also, it will have to produce negative current when the output voltage is positive.

An amplifier which limits heavily into a short circuit can do neither of these things. The amount of out of phase current a good amplifier should be able to deliver is a moot point. Amplifier makers want speaker makers to tame their designs to make life easier for them. Speaker makers reject any restrictions and often produce systems that only the best amplifiers can drive without VI limiting. The market has to sort out which combinations don't work together. This is not a happy situation but it is reality. Amplifier makers take for granted a few facts about speakers. Systems are mostly nominal 8 ohm or 4 ohm and their impedance does not go much below this value as it is close to the DC resistance of the driver(s) involved.

The assumption based on experience and used by many amplifier designers is that the phase angle will at worst be no more than 45 degrees and then be at an impedance that is about 40% higher than the nominal value. For most loudspeaker systems this assumption would be true. The above situation can be simulated by a load impedance formed by a power resistor of 4 or 8 ohms in series with an air cored inductor of about 5 to 10 millihenrys. At the frequency where the phase angle is 45 degrees the load will be either 5.6 ohms or 11.2 ohms (see note 2). Any good amplifier should be able to drive this combination at full level and any frequency without problems. However, even some well known ones don't.

What's the Problem?

If the VI limiting circuit prevents current from reaching the value needed to drive a load at a particular frequency then an unexpected thing happens. The stored energy takes over from the amplifier and dumps a large spike of voltage and current back into the output stage, twice each cycle. Each spike is equal to or more than the amplifier supply voltage and starts around the zero crossing of the wave form lasting about half a millisecond (Figure 4). Such spikes of possibly over 100 volts sound very nasty. They usually don't damage the bass driver but other speaker components (like piezo horns) and even an amplifier has been known to quickly fail because of them.

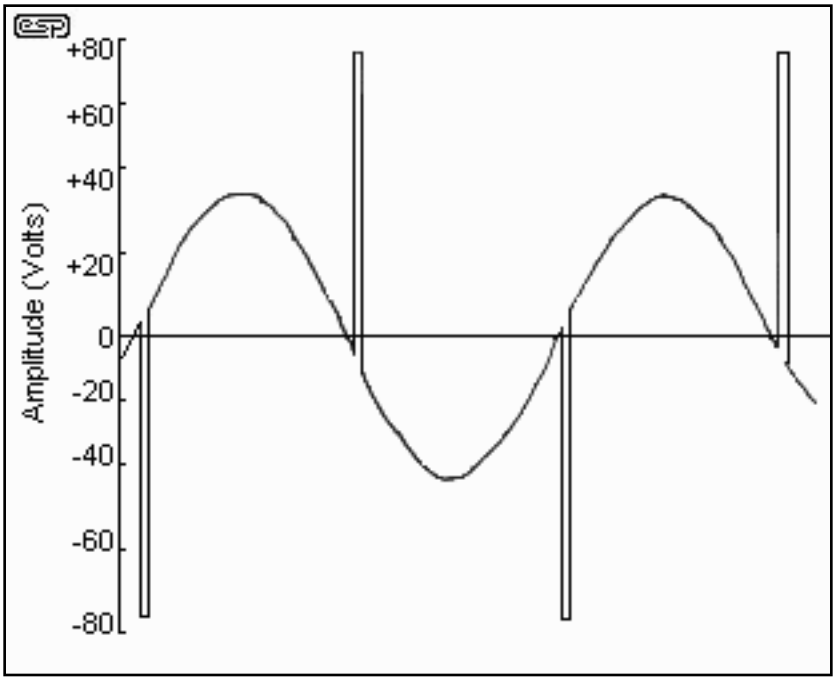


Figure 4 - Voltage Spikes Caused By VI Limiter

Why don't amplifier makers get it right?

Mostly they do. The release in the late 1970s of Hitachi's Power Mosfets eliminated the need for VI limiting allowing most MOSFET amplifiers to have only a simple peak current limit set by zener diodes across the MOSFET gate drive and rely on supply rail fuses to open when the amp was shorted. These amps can drive any known loudspeaker (See note 1). Amplifiers using bi-polar output transistors, even the likes of Motorola's best, are at higher risk of failure and fuses don't blow fast enough. Cost restrictions set a limit to how many output devices can be used, and sometimes the need to prevent output stage overheating and device failures when shorted produces a design which is unable to drive nominal 4 ohm loudspeakers. Another reason for not getting it right is that there is no industry standard for reactive load testing. Resistive load tests are all that are normally done or specified. Similarly speaker makers don't fully specify their products either.

How much out of phase current do amplifiers need?

It depends on the loudspeaker, but the more the better. A first class amplifier is one that can deliver the same current at zero volts as it does at maximum output reducing only when the output swings to the opposite polarity. Such an amplifier rated at 800 watts into 4 ohms would have a short circuit current of at least +/- 20 amps. A lower current capability is often considered acceptable as it will satisfy the majority of loudspeakers, a level of 50% of the maximum output current available at zero volts reducing to zero at full opposite supply. This takes into account the 40% higher impedance expected at the frequency of a 45 degree phase angle and the fact that current is 71% of maximum at zero volts.

Testing the Amplifier

To find what a particular amplifier can deliver in the way of out of phase current requires a couple of bench tests. Neither is complicated or out of the reach of the average technician.

Only a few low cost items are needed - a high power air cored 5 mH inductor (see Note 3) and a 0.1 ohm 10 watt resistor as well as the usual 4 and 8 ohm dummy load resistors.

Test 1
Connect the 5 mH inductor in series with the dummy load and the amplifier with the oscilloscope connected across the amplifier output as usual. Set the sine wave generator at about 250 Hz and using the 8 ohm load gradually increase the level until either spikes are seen or clipping occurs. If there is no problem switch the load to 4 ohms and the generator to 127 Hz and repeat the above. The generator can be swept from 50 Hz to 500 Hz to investigate further for any spiking though if there is no problem at either of the test frequencies then there will usually be no problem at any other. Note the power level at which any spiking occurs.

Test 2
Connect the 0.1 ohm resistor to the amplifier and the oscilloscope leads directly to the resistor. Set the generator to 250 Hz and switch to square wave. Couple the signal into the amplifier input through a capacitor of about 2.2 nF (0.0022 uF). This creates a series of spikes of low average power that does not stress the amplifier. Increase the input level until clipping of the spikes is visible and note the oscilloscope reading in each polarity. Multiply by 10 to get the answer in amps. This test reveals the current available at zero volts and should be compared with the rated output current into 4 or 8 ohms derived from the power output specification. If the positive and negative current readings are asymmetrical or very high there may be a fault with the VI limiting circuit itself or the amplifier is NOT short circuit proof.

Conclusion

A load using a resistor equal to the rated impedance and an inductor in series driven at 45% phase angle allows a simple pass or fail test for an amplifier depending on whether any spikes on the sinewave are heard or seen on a oscilloscope screen. The amount of current available into a short circuit is also a useful guide. It should be 50% of the peak current output into the rated load. Amplifiers that fail these tests may be put on "restricted duties" and used only with 8 ohm nominal loads or speakers that are known as easy to drive. Modification of the VI circuit is always possible but you have to know exactly what you are doing as the output stage may be rendered vulnerable.

Notes:

1. The Perreaux 8000B MOSFET, rated at 800 watts 4 ohms, has a maximum peak current of 60 amps at any output voltage. There are 5 power mosfets in parallel with

zener diode limiting to about 12 amps each.

2. The formula for the frequency for 45 degree phase angle in an RL network is the same as for the -3dB frequency point ... namely $f_{45°} = R / 2 * \pi * L$ For example, for 4 ohms and 5 mH, the frequency is 127 Hz.
3. Any inductor used for testing should be air cored to prevent saturation distortion and should have low resistance. If the resistance is other than negligible, adjust the resistor value to compensate.

Editorial Comment (By Rod Elliott)

Some amplifier circuits show diodes connected between the output and the +ve and -ve power rails - these are shown in grey in Figure 1. The purpose of these should now be obvious - they will "catch" the spikes generated under the conditions Phil has explained. These will (hopefully) prevent the destruction of output transistors by shunting the excess voltage and current back into the power supply which simply absorbs the energy. The diodes prevent the amplifier's output voltage from ever exceeding the supply by more than 0.65 volt or so, and therefore prevent reverse biasing of the transistors under fault conditions.

Needless to say, this provides zero sonic benefit, since the spikes will still occur, but they will now be limited to a little over the amplifier's supply voltage. The amp will probably not self destruct, but it will sound just as horrible when driven beyond its current limits.

High power amplifiers (i.e. anything over about 400W) nearly always use MOSFETs rather than bipolar transistors, and this is especially true of professional units for sound reinforcement. There are very good reasons for this, as Phil has explained.

The ability to drive virtually any load without the need for VI limiter protection is one of the all time great advantages of power MOSFETs. Naturally, the requirement that there are sufficient output devices to handle the load still exists, but MOSFETs are far more tolerant of fault conditions than bipolar transistors, and are not subject to a most undesirable problem known as "second breakdown".

This occurs in bipolar transistors when a small section of the silicon die becomes hotter than the rest due to an overload condition, or where the devices are too small for the job. This increases the gain in that small section, and it therefore does even more of the work, causing it to get hotter still. Thus a repeating the cycle is started, which results in device destruction. This happens very fast (in as little as a few tens of milliseconds), and it is virtually impossible to protect the device once second breakdown has started. The VI limiters are designed to prevent the device(s) from ever exceeding the manufacturer limits where second breakdown is known to occur.

As Phil has shown, some attempts at achieving that goal are somewhat less than a complete success.

Return to Articles Index

Back to Home Page

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Phil Allison (plus editorial material by Rod Elliott), and is Copyright © 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Phil Allison) and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Phil Allison and Rod Elliott.

Page created and copyright © 19 Aug 2000

The Audio Pages

Elliott Sound Products **Soft Clipping**

Copyright © 2006 - Rod Elliott (ESP)
Page Created 15 April 2006

Articles Index

Main Index

Contents

- 1 Introduction
- 2 Soft Clip Circuit
- 2 Amp Power Reduction
- 4 Conclusion

---

1 - Introduction

There seems to be something really nice about the idea of soft clipping. Rather than have the amplifier clip any wayward peaks with the traditional square-edged characteristics typical of solid state amplifiers, wouldn't it be nice if they did much the same as a valve amp? A comparison of the two forms of clipping is shown in Figure 1, and you can see that the valve amp's clipping behaviour is more 'civilised' (if any form of clipping can be considered civilised, that is).

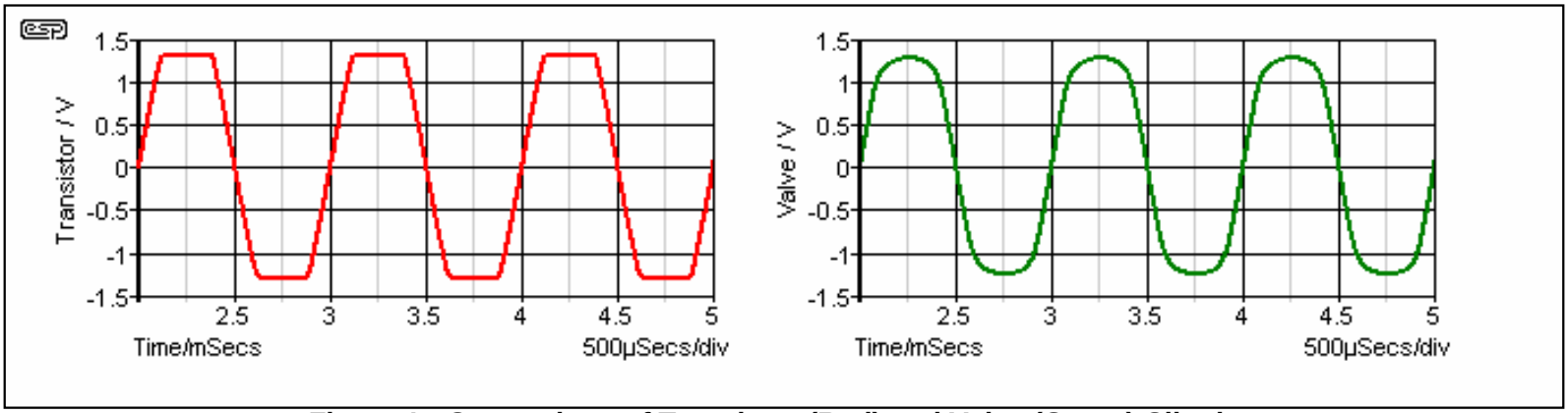


Figure 1 - Comparison of Transistor (Red) and Valve (Green) Clipping

The 'soft' characteristic (Green) shown in Figure 1 has few high order harmonics. The harmonic content is predominantly third harmonic, with a smaller amount of fifth, and lesser amounts of each additional

higher odd-order harmonic. Because the waveform is symmetrical, even order harmonics are typically at vanishingly small amplitudes. Figure 2 shows the harmonic structure of each waveform. Note that the hard clipping produces high levels of eleventh, fifteenth and nineteenth harmonics compared to the soft clip circuit. However, both signals will sound objectionable with the amount of clipping shown.

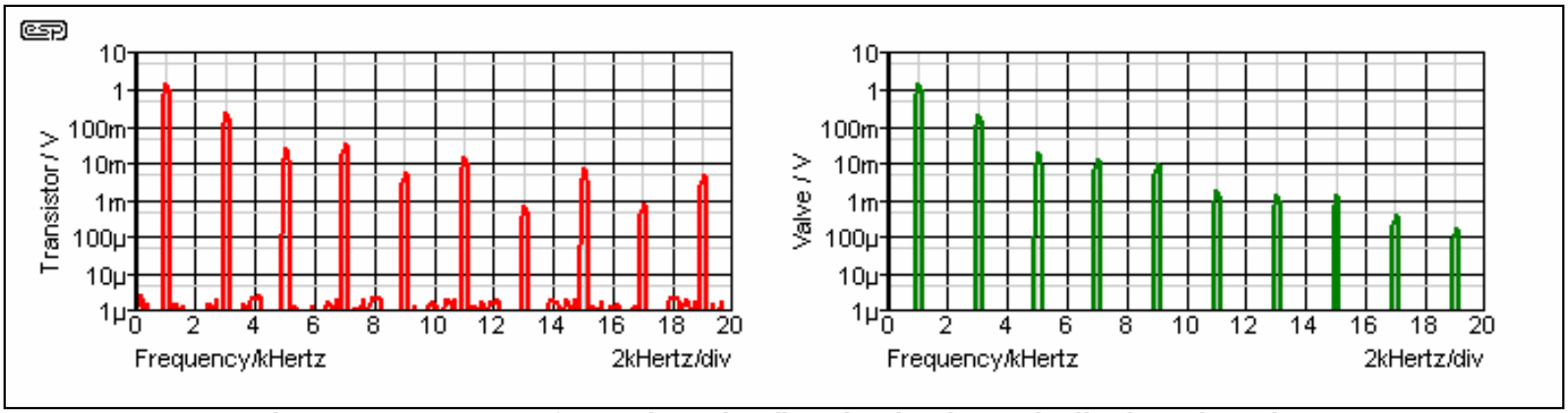


Figure 2 - Spectrum of Transistor (Red) and Valve (Green) Clipping Distortion

Note that the ragged looking transistor spectrum at low levels may appear to be 'bad', but all such signal levels are at about 1uV, so are around 120dB below the amplitude of the fundamental. The spectrum tells part of the story, but the harmonic distortion also needs to be examined. The soft clipped valve-like distortion measures 14.6%, while the transistor circuit gives 17.6% by comparison. In each case, the amplitude of the original (unclipped) waveform was identical, at 2V peak (1.414V RMS). The primary reason for the higher measure distortion of the hard clipped waveform is that the harmonics extend to well over 100kHz at levels exceeding -80dB, while the soft clipped harmonics are below that level by 23kHz.

---

2 - Soft Clip Circuit

Apart from the obvious solution of using a valve output stage (hardly a simple modification to an existing circuit), the easiest way to make a circuit that clips 'softly' is to use diodes. The above simulations were done using diodes, and reality is very close indeed to what is observed in a simulation. There is nothing new about diode clipping circuits - they have been the mainstay of guitar distortion pedals (fuzz boxes) for many years. The deliberate use of this technique for a hi-fi amplifier is less common, and as will be shown below, this is as it should be.

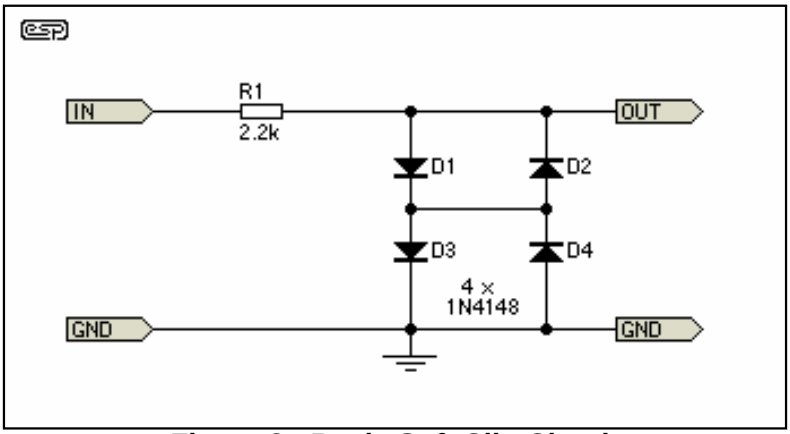


Figure 3 - Basic Soft Clip Circuit

Figure 3 shows the basic schematic of a soft clip circuit. By using diodes, the relatively soft knee of the diode conduction curve provides exactly the waveform that we need. There is an inevitable cost though, and to understand why, we need to examine the conduction characteristics of a diode. It is commonly taken that conventional silicon diodes conduct at 0.65V, although the actual figure varies depending on the type of diode and the current.

The value of R1 is surprisingly very important. If it is too high, there will be considerable distortion at even relatively low levels because of the inherently non-linear resistance of the diodes. If R1 is too small, performance at lower levels is improved, but the source amplifier (preamp, CD player, etc.) may be forced into premature clipping because of the loading. As shown, 2.2k is a reasonable compromise, but if you are willing to include an opamp that can drive low impedances (such as the NE5532 or OPA2134 dual devices), you can reduce R1 to about 680 ohms without having to worry about hard clipping from the preamp.

In Figure 4, it is obvious that not only do the diodes start to conduct at well below the nominal voltage, but they have significant internal resistance as well. It is these very characteristics that give us a soft clipping waveform, and also give us greatly increased distortion as we approach the clipping voltage. The sharpness (or otherwise) of the clipped waveform depends on the signal source impedance and the diode characteristics, and manipulation of the impedance (but maintaining the same diodes) has a very significant effect on the final waveform.

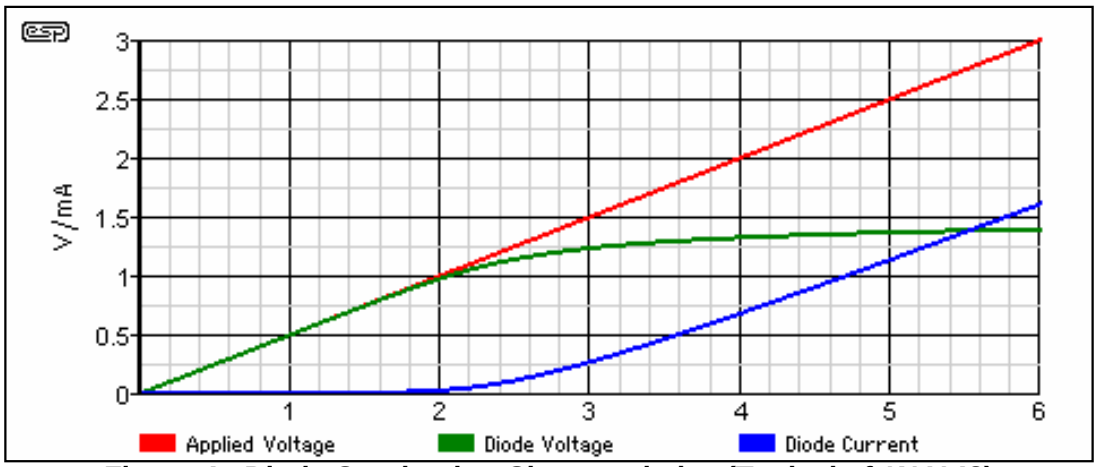


Figure 4 - Diode Conduction Characteristics (Typical of 1N4148)

The supply voltage is applied using a ramp waveform, and the diodes (two in series) are fed using a 1k resistor. As the voltage increases, there is virtually no diode current until the voltage has reached about 1V. At this voltage, diode current starts to flow, and the voltage across the diodes deviates from the applied voltage. As the supply voltage increases further, diode current also increases, and the voltage across the diodes starts to flatten out. You can see that it is not completely flat even at the extremity. The voltage continues to increase at a rate determined by the diode's internal (dynamic) resistance - in the case of the simulated 1N4148 pair used, this can be calculated to be (based on the variation across a defined area of the curve where the curve has flattened out) ...

Voltage change = 3.5mV
Current Change = 21.3uA
R = V / I = 3.5mV / 21.3uA = 112Ω

This is for the pair, so each diode has a dynamic resistance of 56 ohms. Note that you cannot simply use the voltage across the diodes and the current through them to obtain this figure, because of the diode's internal voltage drop. You will get an incorrect figure (that is much too high) if you do that. The point is that the dynamic resistance changes, depending on current. To see this effect properly, Table 1 shows the dynamic resistance (impedance) at each numbered point along the horizontal (X) axis of Figure 4.

Point	ΔVoltage	ΔCurrent	ΔResistance
1	50mV	171nA	292k
2	34mV	7.7uA	4.4k
3	11.4mV	18uA	633
4	5.55mV	21uA	264
5	3.50mV	21.3uA	164
6	2.44mV	21.6uA	112

Table 1 - Variable Dynamic Resistance of 2 x 1N4148

Much as simulations allow the easy determination of things that are very hard to measure, there is nothing like measurement to demonstrate the effects in the real world (as opposed to the cyber-world of the simulator). Accordingly, Table 2 shows a comparison between measured and simulated results, with each using the same basic parameters.

V_{in}	V_{out}	Measured THD	Simulated THD
1	830m	8.9%	7.1%
800m	740m	4.5%	3.0%
700m	670m	2.2%	1.46%
600m	587m	0.81%	0.57%
500m	500m	0.28%	0.19%
400m	400m	0.076%	0.06%
300m	300m	0.036%	0.018%

Table 2 - Measured vs. Simulated Distortion

The simulator under-estimates the distortion, and although it is (probably) more accurate, the measured results are ultimately what really counts. My audio oscillator has a residual distortion of 0.015% at 1kHz, and this needs to be factored in at low measured figures. One thing that is inescapable is that the distortion is increased at all levels. Even reducing the input to 50mV, the simulator still shows a measurable distortion ... although few amplifiers could even hope to get down to the 0.0002% level the simulator indicated.

The value of R1 (as noted above) has a great influence on the circuit's performance. Table 3 shows the simulated distortion levels for a range of voltages and series resistance. Be aware that very few opamps can drive significant levels into low impedances, so this limits the minimum value to around 680 ohms with high drive opamps, and about 1.5k or so with most others.

V_{in}	680 Ohms	1.2k	2.2k
1	4.90%	5.92%	7.08%

800m	1.44%	2.11%	3.00%
700m	0.57%	0.92%	1.46%
600m	0.19%	0.33%	0.57%
500m	0.06%	0.11%	0.19%
400m	0.019%	0.033%	0.060%
300m	0.006%	0.010%	0.018%

Table 3 - Simulated Distortion vs Source Resistance and Voltage

From the above, it is readily apparent that low values for R1 will reduce the distortion below the theoretical diode conduction voltage, but this approach loads the source amplifier. In addition, the dynamic resistance of the diodes becomes more invasive, allowing the clipped voltage to reach a higher than expected peak amplitude. For example, with an input of 2V RMS, 2.2k will cause the peak to limit at 1.32V, 1.2k gives 1.37V, and 680 ohms allows the peak to reach 1.41V. I shall leave this decision to anyone who may be interested in experimenting with the technique.

So, if you want to add a soft clipping circuit, you will benefit from lower levels of harsh high-order harmonics at clipping, but at the expense of effective power output and (much) higher than normal distortion at all levels from as low as about 20% of the maximum output voltage.

---

3 - Amp Power Reduction

If a soft clip circuit is incorporated into an amplifier, it must prevent the power amp from clipping at any sensible input level. Figure 5 shows how the level changes (this is due to the diodes' dynamic resistance) as the input voltage is changed. The red trace is with an input level of 1V RMS, and the green trace is at 5V RMS. Nearly all preamps will allow at least that much drive if the volume control is set to maximum, and this is what tends to happen at parties (and often with professional equipment as well).

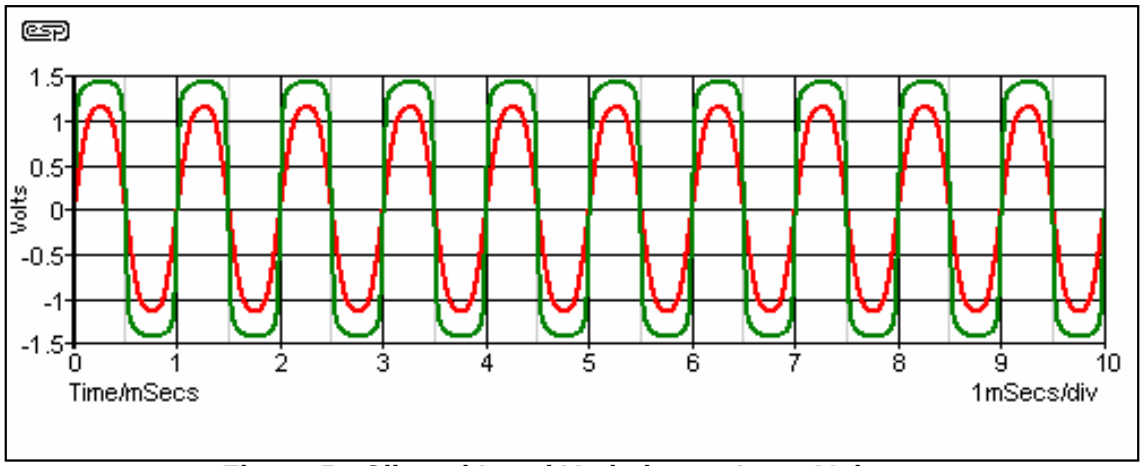


Figure 5 - Clipped Level Variation vs Input Voltage

This is essentially exactly the same information as shown in Figure 4, but uses the signal waveform. For the exercise, let's use a 100W (8 ohm) amplifier as our model, and we will apply a soft clipping circuit to it. Using the 4 diode clipper shown in Figure 3, we can do some basic analysis.

Allowing up to 5V input is probably silly, so we'll make that 2V, after which traditional hard clipping will

occur. A 100W/8 ohm amplifier will give about 28V RMS at the speaker terminals (40V peak). We have allowed for 2V input, so the gain needed (after the clipping circuit) is ...

Peak Input Voltage = 2 * 1.414 = 2.83
Peak Clipped Input Voltage = 1.32 (from Figure 4)
Gain = V_{OUT}/ V_{IN}= 28 / 1.32 = 21

The gain needed is achieved using a 20k resistor from output to inverting input, and a 1k resistor to ground. So far so good. Now, we need to decide on the maximum distortion figure we are willing to accept at a 'typical' peak output power. 1% is not unreasonable - certainly higher than we would expect from a transistor amp, but not unreasonable.

From Table 2 we see that with an input level of 700mV (RMS), distortion is 2.2% (measured) or 1.46% (simulated). The input level obviously needs to be a bit less than this, so 650mV is probably about right. The simulator says that this gives 0.94% which is close enough. The actual amplifier power at 1% distortion is determined by ...

V_{OUT}= V_{IN}* Gain = 650mV * 21 = 13.65V
Power = V^2/ R = 186 / 8 = 23W

Oh dear! A perfectly good 100W amp is now downgraded to 23W at 1% distortion, with the distortion rising quite rapidly above that. Sure, the distortion components will be low order odd harmonics, but intermodulation distortion is increased proportionally as well, and the otherwise distortion free (relatively speaking) amp output cannot be used to its full advantage. Use of a lower value feed resistance (R1) will improve matters, but as noted above may create other problems.

Will it sound like a valve amp though? Not really. While the distortion characteristics will be similar to a reasonable push-pull valve amp of roughly similar power, the higher than normal output impedance of a valve amp is another factor that gives the 'valve sound'. While this can be included as well, the clipping circuit cannot compensate for the gain variations in an amp with high output impedance.

---

4 - Conclusion

The effects of the soft clipping circuit can be modified over a wide range, by varying the feed resistance, type of diodes used and the peak levels expected. In general, the result is unlikely to be as hoped for anyway, because occasional transient clipping is usually inaudible. The harmonic and intermodulation distortion that has been added may be very audible with some material, and is unlikely to improve your enjoyment of the music.

Of course, you may find that you do like the effect (it is very common in guitar amps), so if you think that you'd like to experiment then you have enough information to make an informed decision as to how to go about it.

There is nothing that looks complicated at all in this technique, but as should now be evident, there are actually quite a few things that must be considered. As is so often the case, an apparently simple circuit can be far more complex than anticipated, and it is a matter of juggling the compromises to obtain the

results you want - within the limits of the technique.

While it would be relatively easy to incorporate the soft clip function into a DSP that may be providing other functionality (equalisation, crossovers, etc.) and avoid some of the limitations, the definition of a soft clip circuit requires that it should start to distort the signal earlier than expected. In the end, it is probably far easier to incorporate a clipping indicator so that clipping can be avoided altogether.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2006. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 15 April 2006

Elliott Sound Products **High Speed Amplifiers in Audio**

Copyright © 2001 - Rod Elliott (ESP)
Page Published 05 Jul 2001
Updated 18 Aug 01

Articles Index

Main Index

Contents

- Introduction
- 1. The Need for Speed
- 2. High Speed Opamps
  - 2.1 Cable Impedance
- 3. THS6012 Current Feedback Opamp
  - 3.1 CFB Opamp Headphone Amplifier
  - 3.2 THS6012 CFB Pre-Amplifier
  - 3.3 Zobel Network
- 4. IC Mounting Details
- Conclusion
- Terminology

Most terms that require a description use a popup window for explanation. Use the last link if the Javascript popup help screens don't work on your browser.

Introduction

Over the past few decades, there has been much musing, discussion and outright argument about the need for wide bandwidth in audio systems. Unlike most other discussions, I am not referring to extension to 30 or even 50 kHz, but well into the RF spectrum. There is apocryphal "evidence" that high speed amplifiers sound better, although after years of such debate and argument, no-one has been able to shed any light as to the possible reasons why this may be the case.

After an e-mail from Texas Instruments and a package of evaluation boards, documentation and a selection of high speed opamps, I started thinking about this issue seriously. The original aim was simply to have a look at an xDSL line driver, which was thought to have great potential as a headphone amp, and this is most certainly true - initial tests are described below, and a project describing its use is in production.

The real thing that got me to thinking was the seemingly impossible claim that different interconnects sounded

different. The tests I have run indicate that the differences between materials is so slight that it is all but immeasurable, and this is backed up by various others who have taken a similarly scientific approach, although in many cases to a far greater degree of refinement than my own testing. Despite all the tests, there still seem to be situations where different cables sound, well, *different*.

The following is somewhat speculative, since I don't have a radio frequency interference (RFI) problem where I live, and I am understandably reluctant to create some so I can test my theory properly. Lacking a fully shielded room (Faraday cage), and a large amount of very expensive test gear, it is very difficult to prove anything one way or another.

Nonetheless, I think that the discussion to follow may prove to have some merit - I am very interested to hear any feedback from anyone who is in a position to test the ideas that follow. The theory is quite solid, but the final proof is naturally in the listening and evaluation - provided it is done in a scientific and controlled manner. Any single non-blind (as opposed to a properly conducted double-blind) evaluation is a waste of everyone's time, as the results are purely subjective.

---

1. The Need for Speed

Although you will not see it written very often, the venerable uA741 opamp is quite fast enough for audio, despite the fact that it has a slew rate of only 0.5 V/us. Contrary to popular belief, audio does not have super fast transients, since the sounds we hear are predominantly created by mechanical or biological means. No mass (including air) can change its direction instantaneously unless an infinite force is used, and although I have come across a few pretty strong drummers (for example), there are none I would describe as "infinitely powerful". These simple mechanical laws (and the filtering used on CD), prevent the generation of very high order harmonics that are needed to produce extremely fast risetimes.

There is considerable evidence to show that many instruments (as well as speech) create harmonics in excess of 50 kHz, but they are low level, and do not (generally) contribute to fast rise times on transients. Having said that, it is universally agreed that the 741 is a sonic disaster - they sound bloody awful. To understand some of the reasons, we need to look at the frequency response, and its effect on the overall behaviour of the opamp.

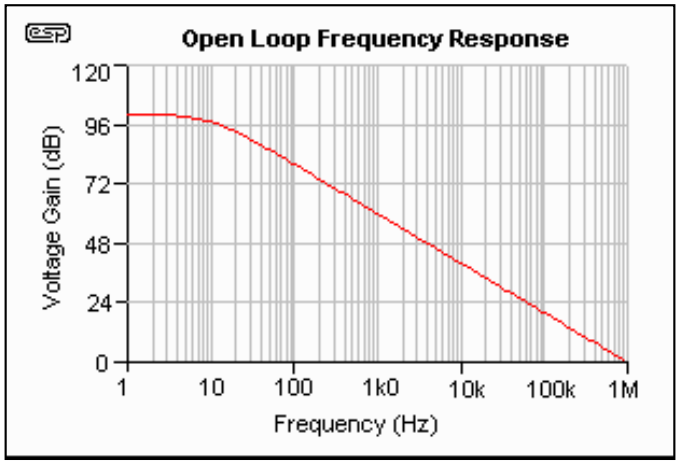


Figure 1.1 - uA741 Frequency Response

As you can see, there is a vast gain, but only at low frequencies. The internal compensation capacitor rolls off the response above 10 Hz, and this means that as the frequency increases, there is less and less feedback to correct the internal distortion. For the purposes of this discussion, there is a more insidious (and less well publicised) effect - the output impedance rises with increasing frequency. For reference, the closed loop

response is shown, where the gain has been set at 10 (20 dB). As you can see, the opamp is struggling to maintain output at 20 kHz, and at slightly above this frequency, *it has no feedback at all*!

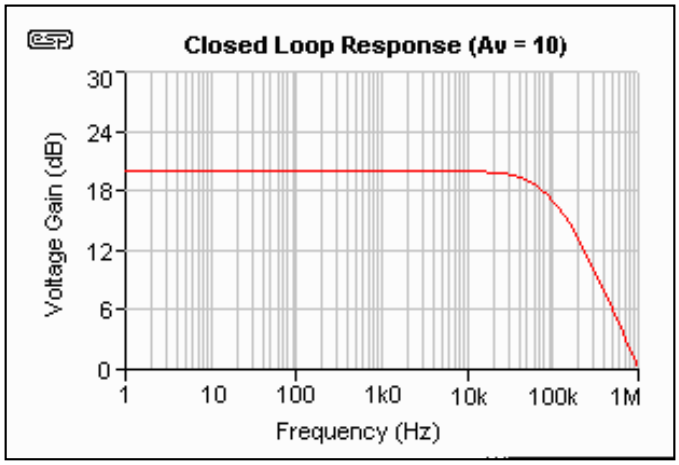


Figure 1.2 - uA741 Closed Loop Response

Even at 10 kHz, the gain has dropped sufficiently that there is little feedback available. It must be said that even gross distortion at 10 kHz is inaudible with a single tone, but intermodulation products are generated as a result of the multiple frequencies that are present with music. The effects of such distortion do not seem to have been covered in any great detail in most journals and articles, but tests do exist. They are never applied in reviews, nor are the results quoted in manufacturers' data.

The high frequency intermodulation products caused by the distortion at high frequencies may well be one of the reasons that wide band opamps are said to sound so much better than the likes of the uA741. Even so, the best of the best VFB opamps will still have the same problem - only the degree is different. There is no VFB opamp that has an open loop bandwidth that extends beyond about 1 kHz, so they will all suffer from a frequency dependent reduction in feedback factor, a consequent rise in distortion, and increase in output impedance.

This is where the use of very fast opamps becomes potentially important. A cable terminated by a zero ohm source will pick up very little external noise, and cable noise (microphony) is all but eliminated because it is the result of an extremely high impedance "generator" - the conductors and insulation material of the cable. Zero ohms is an impossible figure at best, and nearly all opamp circuits require a resistor in the output to prevent oscillation when a cable is connected. The "cable oscillation" topic is covered further a little later.

Even with a resistance of 100 ohms in series, the impedance is sufficiently low as to be quite effective in preventing noise pickup (including RF), and will still virtually eliminate microphony. Except ... as the frequency increases, so too does the opamp's output impedance. The 100 ohms at 1 kHz (for example), becomes greater even as low as 10kHz, reducing effectiveness further and further as frequency increases. By 10 MHz, the opamp's output is effectively an open circuit, having an impedance many times the characteristic impedance of the cable. At lower frequencies where opamp impedance is low, the 100 ohm resistor is a passable match to the cable, but only at one end. The cable *will* become resonant at some frequency, determined by the length of the cable, its effective inductance and capacitance, and its velocity factor. This happens with all cables, regardless of price, silver, wire size, or soaking in snake oil !

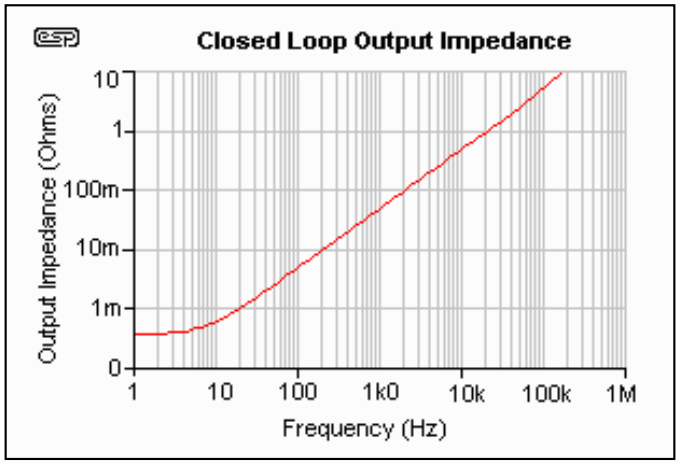


Figure 1.3 - uA741 Output Impedance Vs. Frequency

For effective shielding and transmission of RF, a cable is normally terminated at each end with its characteristic impedance. This impedance matching (source to cable to receiver) is essential to prevent reflections and standing waves at the operating frequency. Audio cables are not so terminated, and although this has no effect at audio frequencies, it potentially reduces the effectiveness of the cable at *rejecting* interfering RF signals. (It is actually undesirable to match the impedances in audio, as there is an inevitable 6 dB increase in noise, due to the reduction of signal level.)

The interfering RF affects not only the receiving circuitry, but the transmission end as well. RF signals that find their way into the output of an opamp that has a limited bandwidth (and high output impedance at the interfering frequency) can have a very profound effect on the sound quality. Non-linearities will always exist in any electronic equipment, but are normally reduced to very low levels (but never eliminated) by feedback. When the opamp (or any other low impedance output amplifying device) has a signal fed into its output that cannot be reduced to near zero by the low output impedance, the signal is fed back into the input, where rectification (AM detection) can take place. This phenomenon is often observed when equipment picks up AM radio stations, but with a high resolution system, there will be audio degradation well before the interfering signal is discernible as a radio station.

The problem may be exacerbated by the use of a feedback capacitor (as shown in Figure 1.4), which provides a low impedance path from the output back to the -ve (inverting) input. In theory, this also reduces the gain (and therefore the available feedback), but as shown, a 741 type opamp has no gain at all above 1 MHz, so any feedback path above this frequency will simply cause problems. This technique is common with fast VFB opamps, and "seems" not to cause any problems. This is a reasonable assumption in the case of an internal amplifier stage, but needs further investigation when output stages driving cables are concerned.

It is worth noting that this form of feedback will simply cause a current feedback opamp (such as the THS6012 described here) to oscillate, and must not be used under any circumstances.

Another technique I have not seen used in any commercial equipment except for power amplifiers (where it is almost mandatory), is a Zobel network at the output of preamps and other low level signal sources. This will ensure that the cable is terminated by a low impedance at nearly all frequencies, but this is difficult with opamps having limited high frequency response since the additional loading will only make distortion worse. An explanation of how (and when) this is useful will be covered later in this article.

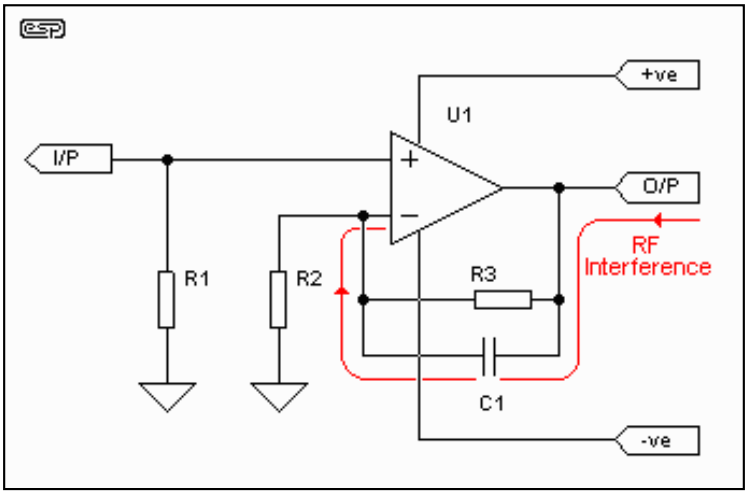


Figure 1.4 - Feedback Cap Provides RF Path to Input

The sources of interference are wide and varied. Some interfering signals are within the audio range, but the majority are radio frequencies, starting from the AM broadcast band at about 600 kHz, extending to the CB band (27 MHz), then through to amateur, commercial, TV bands and upwards. Other commercial sources of narrow band interference include RF welding equipment (used for welding plastics), microwave ovens, mobile phones and the like (although digital mobile phones are hardly narrow band, since the signal is pulsed).

In addition there are residential and commercial sources of broad band interference - brush type electric motors, arc welders, switching systems, arcing insulators on the power distribution grid - the list is almost endless.

Although it may seem that so far I have simply beaten the poor uA741 to death, it must be said that there are many other opamps that are very much faster, and also provide lower distortion, higher open loop gain and far lower noise. Even the very best of these still suffer similar problems at high frequencies (above 100 kHz), since they are designed for relatively conventional audio frequency applications. This does not necessarily mean hi-fi - there are a great many applications for the audio frequency range that have nothing to do with "audio" per se.

In case you were wondering, the noise pickup effects described above do not mean that you should rush out and get new cables. This has nothing to do with cables themselves, but has everything to do with the amplifier driving the cable, and/or receiving the signal at the other end. With the proper techniques in place, a coathanger will outperform the most expensive cable you can buy (assuming of course that shielded coathangers are available where you live :-)

A conversation with a friend recently uncovered the fact that he had a customer with an RF interference problem. Substituting the original opamp for one with a much wider bandwidth (in this case he used an OPA2134) solved the problem. A Zobel network was also added, and this managed to help the original opamp cope a little better, but could not eliminate the interference. I don't know what the original opamp was, but in any consumer equipment we can safely assume that it was cheap and nasty.

---

2. High Speed Opamps

Enter the (relatively) new devices from Texas Instruments. These were designed for xDSL (e.g. ADSL, or Asynchronous Digital Subscriber Line) applications, which use a large number of high frequency carriers to convey the data via multiple simultaneous channels (255 in the case of ADSL) . The carrier frequencies are

closely spaced, so very low distortion is essential to minimise cross-modulation of the carriers (basically intermodulation distortion). At the far end of an analogue telephone line the signal may be greatly attenuated, so low noise is also important.

The above specifications already sound very satisfactory for audio - low noise and distortion are requirements in nearly all hi-fi applications, but there's more. Unlike traditional voltage feedback (VFB) opamps, these new devices are current feedback (CFB), and have an extraordinarily wide bandwidth, so the degradation at the very highest audio frequencies is negligible, and there is no significant rise in distortion or output impedance until above at least 10 MHz. Indeed, the performance of the THS6012 in this regard is as good or better at 100 MHz than the uA741 at less than 100 kHz.

The fundamental difference between VFB and CFB opamps is quite simple ... well, it *is* when you simplify it to the level I am about to

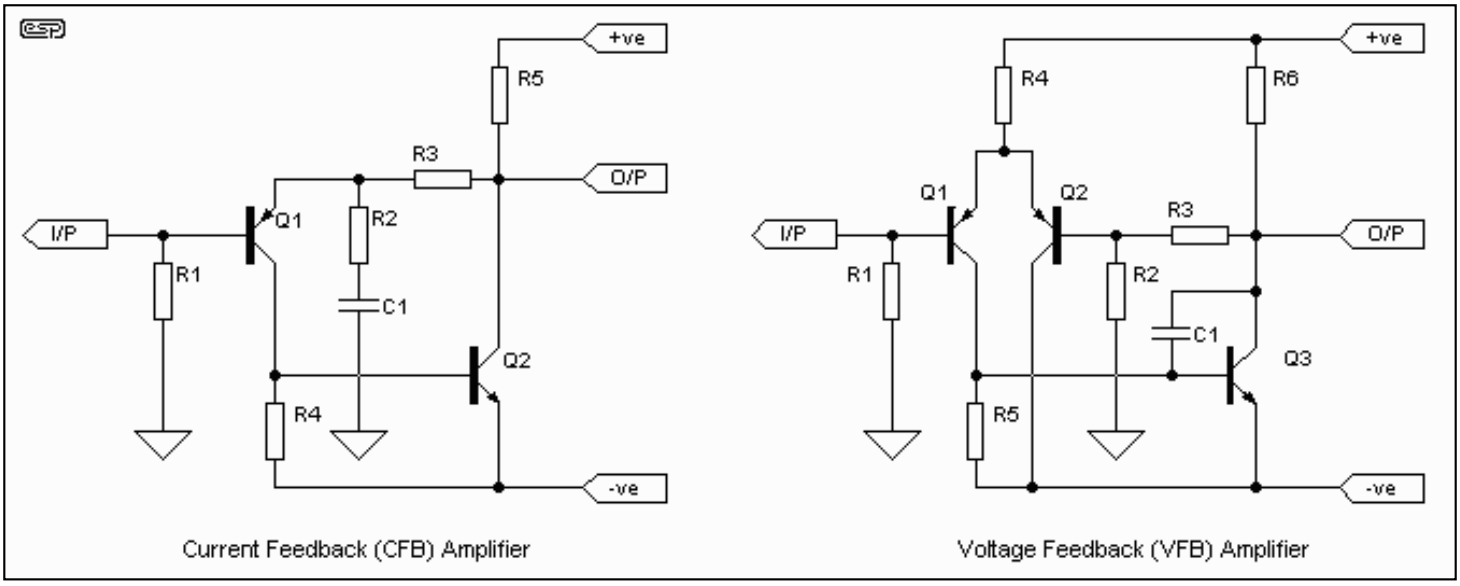


Figure 2.1 - Simple CFB and VFB Opamps

As well as those shown in Figure 2.1, other perfect examples of VFB versus CFB are some of the ESP published projects - the "El Cheapo" power amplifier is CFB (as is DoZ and the minimalist preamp), while the 60-100W and P101 MOSFET Hi-Fi amplifiers are VFB. The first thing one should notice that is different, is that the CFB amplifier has no Miller capacitor (also called the Dominant Pole). This is normally connected between the collector and base of the Class-A driver transistor. This is an absolute requirement for stability in a VFB amplifier, but with care, can be eliminated entirely (or at least reduced to a very much smaller value) in a CFB amp.

The result is a much wider open loop bandwidth for CFB amplifiers, but there is no longer any simple way to ensure DC levels through the amplifier are not shifted. In the case of the DoZ amp, minimalist preamp and El Cheapo, the DC shift problem is avoided by using capacitor coupling, but this is not always convenient or desirable. The issue is solved (at least to an acceptable degree) in the TI devices I have been testing, so DC offsets are not as big a problem as with relatively simple discrete designs.

Most commonly used opamps are VFB, and as such have an inverting and non-inverting input with approximately equal impedances. They may happily be used as either inverting or non-inverting amplifiers, and they are quite predictable in normal use. CFB opamps (to some degree) sacrifice DC accuracy for bandwidth, and this is done by eliminating the differential pair normally used as the input stage.

The gain in both the amps shown in Figure 2.1 is set by the ratio of (R2+R3) / R2, as one would expect, although it is not very accurate with the CFB amplifier (another of their little foibles). In the CFB amp above, C1 is needed because DC operation is not possible with this simple configuration. In a simulation done previously (see Amplifier Design), the VFB amp provides an open loop voltage gain of 1,640 (64dB). By comparison, the open-loop gain of the CFB amplifier is 2,190 (67dB) - considering that all other things were maintained equal, the open loop gain is somewhat better.

The real difference is the bandwidth, which was not shown in the original article. The CFB amp can provide an open loop -3dB frequency of over 1MHz, while the VFB amp (using a 33pF Miller cap) only manages 30kHz. These figures are much better than any normal opamp, but it must be pointed out that the noise and distortion figures for these simple circuits will be somewhat lacking compared to a premium opamp.

When the gain is reduced to 20dB (10 times), the -3dB frequency of the VFB amp is 5.6MHz, which is not too shabby at all. At the same gain, the CFB amp really shows its true colours, with a -3dB frequency of 67MHz - better than 10 times the bandwidth! In addition, the CFB amp has a much flatter phase curve, indicating that stability is potentially much better. The 33pF capacitor used in the VFB amp is likely to be marginal in real life, and depending on the transistors used, may need to be increased to prevent oscillation.

As one might expect, there are disadvantages to the CFB configuration as well. As well as DC offset, there is also a lower input impedance and higher bias current. This means that for optimum results, the source impedance should be very low - this is not an issue with video distribution amplifiers or ADSL equipment, since the source impedance will typically be 50 or 75 ohms. It can be an issue in audio, and in many cases a unity gain buffer will be needed in front of the CFB amplifier.

In addition, CFB opamps have different impedances for the +ve and -ve inputs. Conventional opamps use the same topology for each input - almost invariably a long-tailed pair (a.k.a. differential amplifier). As discussed in the Amplifier Design article, using this input stage means that a Miller capacitor is essential to maintain stability. The CFB opamp has a non-inverting (+ve) input that is moderately high impedance, and an inverting (-ve) input that has an extremely low impedance. As discussed above, the gain accuracy as set by the gain and feedback resistors (R2 and R3 respectively) is not as good as a VFB amp, and some tweaking of the gain resistor will be needed to obtain a precise gain.

The common unity gain buffer (where the output is tied directly to the inverting input) is not possible with a CFB amplifier, and a resistance must always be used. In addition, the resistance value is critical to stability and bandwidth - if it is too high, bandwidth suffers, and if too low the amplifier will peak at some high frequency, and may easily become unstable. At very low values (or zero ohms), instability is guaranteed, and the amplifier *will* oscillate.

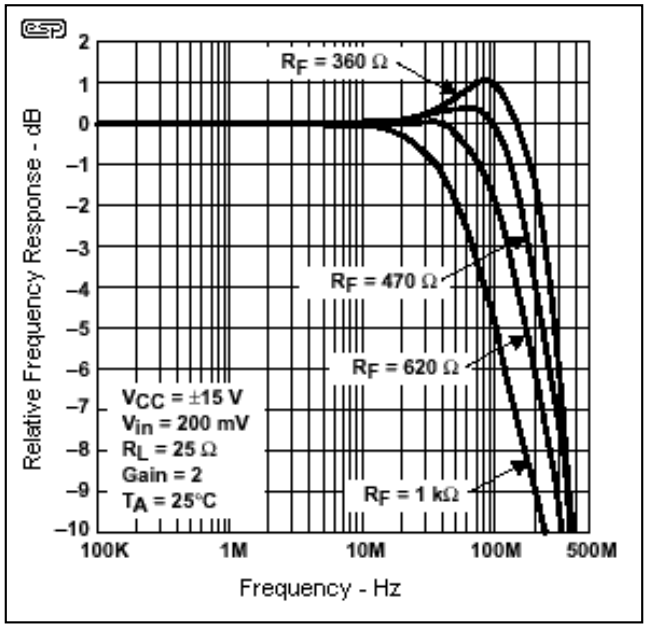


Figure 2.2 - Relative Frequency Response Vs. Frequency

Figure 2.2 shows the relative gain for various values of feedback resistor (adapted from the TI data sheet). The voltage gain is set at 2 (6dB), and as you can see, when using a 1k feedback resistance (the recommended compromise value), the amp is 3dB down at about 70MHz. Reducing the value will extend the -3dB frequency to around 250MHz, but at the expense of some peaking.

Even when set for a gain of 1000 (60dB), the THS6012's response is almost flat to 1MHz - this is more than an order of magnitude better than any compensated VFB opamp I know of - regardless of manufacturer or price. As an example of the difference, the NE5532 (dual internally compensated) will barely make it to 20kHz at the same gain. The NE5532 has a maximum open loop gain of 100dB, but only up to 1kHz - and these are still one of the fastest VFB opamps available. It is worth noting that VFB opamps must always be compensated (most commonly using the Miller or dominant pole capacitor) whenever feedback is applied. If not, they will oscillate.

For audio purposes, we don't need anything above 100kHz, but an extended response ensures that the interconnect will be terminated with a low impedance at nearly all potential interfering frequencies. This can be extended easily by using a simple RC filter at the output of the preamp. More on this later.

2.1 Cable Impedance

The impedance of cables receives little comment in the audio industry (other that for telephone circuits), and within the audio frequency band, this is perfectly reasonable, since it is unimportant. Where it is important is in RF work, but it is becoming more and more common that the two are combined - not because audio has become faster :-) but because the sheer amount of RF pollution is increasing every day.

The sources of RF interference (RFI) are discussed above, but the concept of cable impedance is such that it is worth covering here - albeit briefly. There are some cables that are quoted as being a specific impedance - 50 or 75 ohm coax, 300 ohm TV twinlead and 120 ohm unshielded twisted pair (UTP) for data connections are some examples.

The characteristic impedance of a cable is influenced by a number of factors, and it is easily calculated - although I shall spare you the gory details here. Where cables become sneaky, is when they are a specific

length compared to the wavelength of the signal. This is unimportant at audio, since the cable is always very short compared to wavelength, and the telephone system is the only audio application where impedance *is* important. This is due to the very long cable runs, up to 4km or more in some cases.

The wavelength is calculated by the following formula ...

Wavelength = C / f where C is velocity and f is frequency

For sound in air, C is 345 m/s, but in the electrical domain (in free air or a vacuum), C is 3 x 10^8 m/s. The velocity is reduced in any cable, and in a typical coaxial cable, the velocity is typically about 2.4 x 10^8 m/s. This means that for a 70MHz interfering signal, the wavelength is 3.42 metres.

RF is sneaky and cunning, and does not behave in what we might think is a sensible manner. A length of coax 857mm long is 1/4 wavelength at 70MHz. If one end is short circuit, the other end "looks" like an open circuit. An open circuit at one end appears to be a dead short from the other! If an audio cable is unterminated at each end, the final effect is completely unpredictable if RF fields are present, and it is just as likely to act as an antenna. The shield will not necessarily protect the inner conductor, but is quite capable of inducing a signal into it. Being unterminated (or marginally terminated at one end only), the interconnect is now capable of injecting RF into the amplifiers' inputs and outputs - with unpredictable effects on sound quality.

As far as I know, there are few tests to determine the susceptibility of most audio equipment's interconnects to the effects of RF over a wide range, but there is considerable evidence to show that many amplifiers are afflicted. Make a call on a digital mobile phone near most amps, and the characteristic noise is unmistakable. What more subtle effects do lower RF levels have? I don't know, but I doubt that they will enhance the listening experience.

So much for the technobabble. (For the time being, at least.)

---

3. THS6012 Dual Line Driver

There are two devices that I shall concentrate on, being those for which samples were supplied by TI. These are the THS6012 (500 mA dual differential line driver) and the THS1431 (High speed, low noise, fully differential I/O amplifier). Highly suitable audio uses are presented for each device, but it must be said that implementation is not trivial. Because of the wide bandwidth, these devices can (and do, I can assure you!) oscillate unless proper precautions are taken. Bypassing is critical, as is the PCB capacitance - especially at the inverting input. Even a few pF to earth from the inverting input will cause frequency peaking and possible oscillation, and when one of these little guys oscillates at 100 MHz or more, the current consumption (and subsequent device heating) can cause major problems.

A very brief specification of the THS6012 ...

- ADSL Differential Line Driver
- 400 mA Minimum Output Current Into 25 Ohm load
- High Speed
  - 140 MHz Bandwidth (–3 dB) With 25 Ohm load
  - 315 MHz Bandwidth (–3 dB) With 100 Ohm load
  - 1300 V/us Slew Rate, Gain = 5
- Low Distortion
  - –72 dB 3rd Order Harmonic Distortion at f = 1 MHz, 25 ohm load, and 20 Vpp output
- Wide Supply Range ±4.5 V to ±16 V

- Thermal Shutdown and Short Circuit Protection

The THS6012 is probably of greatest interest to the audio fraternity, simply because it is a dual amplifier, and is suitable for use in preamps or as a headphone amplifier.

A preamp with such a wide bandwidth is not possible with any conventional VFB opamp, but the input impedance of the THS6012 is somewhat lower than an OPA2134 or other premium opamp, and the maximum gain will be somewhat lower before oscillation becomes an issue. Remember, we are discussing an opamp with a bandwidth of up to 315 MHz - this is in the UHF (Ultra High Frequency) RF band, a mysterious place where circuit design is as much a black art as a science.

Note the slew rate, which is quite exceptional, and the distortion is quoted at 1MHz, and at 20V p-p into 25 ohms! Lower frequency or voltage, and higher impedance loads reduce it even further, and at audio frequencies is extremely low. So low in fact that I cannot measure it with my equipment (grumble).

3.1 THS6012 CFB Opamp Headphone Amplifier

Because of its high drive current, the THS6012 seemed imminently suited to a headphone amplifier, and indeed, this was the suggestion originally made by TI. Before I describe the technicalities, I will answer the question "So how does it sound?"

Quite frankly, I was disappointed. There was absolutely nothing that sounded of "amplifier" - this is probably as close to a straight wire with gain as you are likely to find. There is nothing I can say about it, other than it is as clean, tight and transparent as anyone could hope for. Bass response extends to DC, and even after the addition of a capacitor across the input to limit the HF response, it was only 0.1 dB down at 100 kHz. In short - the performance is exemplary in all respects.

In the best traditions of the subjectivists, it would be nice to be able to say "Oh, yes - the imaging is to die for, and the highs are sooo transparent". The foregoing may well be true, but me, I listen for noise and other artefacts, measure distortion (if possible), verify frequency response, and generally judge an amplifier on its accuracy. The THS6012 is exemplary in all respects, but (there has to be a down side) the input impedance is somewhat lower than one might expect, and because of the extremely wide bandwidth, instability can be a real problem if the proper precautions are not taken.

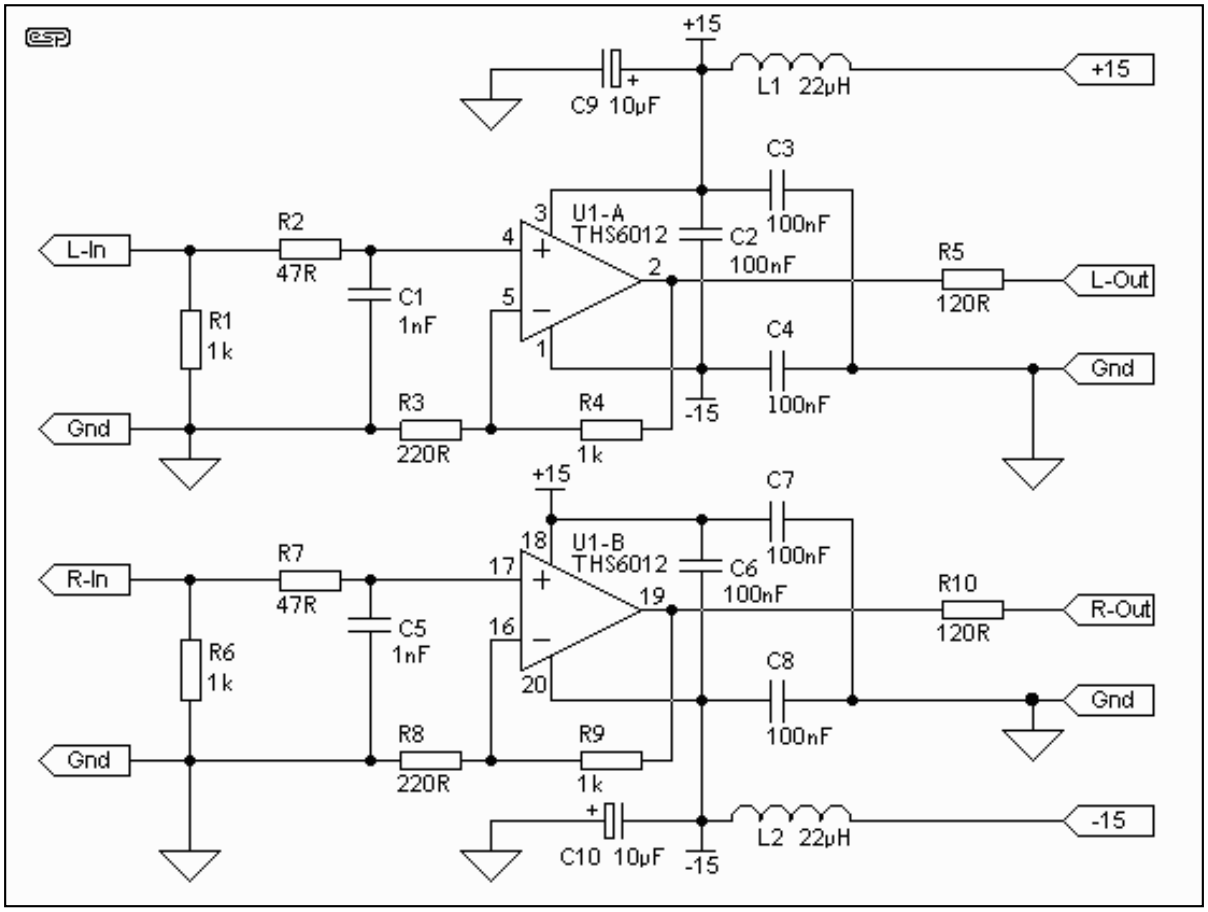


Figure 3.1 - Headphone Amplifier Test Circuit

As discussed above, this device uses current feedback (rather than voltage feedback as used in the more traditional audio opamps). In this case, the current feedback requirement means that the feedback resistors will be of a much lower value than "normal". I configured the amp with a gain of 12 (21.7 dB), and the output level was more than sufficient to cause hearing damage (if you like that sort of thing). Ideally though, the THS6012 should be operated at a lower gain, as this helps to reduce the chance of oscillation. Accordingly, most of the gain will be in a prior stage, and assuming that a high quality opamp is used, the overall performance is astonishing. A simple gain stage based on an NE5532 or OPA2134 dual opamp will be more than sufficient. If direct coupled from an opamp stage, R1 and R6 should be omitted to prevent excessive loading.

Even with the THS6012 configured for nearly 22dB, I could not measure the distortion, as it was virtually the same as the residual from my oscillator, and this applies to any frequency within the audio band. This is in contrast to the majority of small power amps that are used for headphones, whose distortion increases with frequency - especially above 10 kHz. Noise was completely inaudible. Another of my tests is for clipping performance, as there are many amplifiers that are wonderful below clipping, but fall apart when a transient clips. There is more information on this topic in the Amplifier Sound article, for those who may be interested.

Clipping performance was exemplary, showing only the very slightest rail sticking (or "overhang") at the highest frequencies. At normal audio frequencies, the THS6012 simply clipped, and resumed normal operation virtually the instant the signal was below the supply voltage again. This is (of course) what amplifiers are supposed to do, but there are a great many that don't ! Very slight clipping overhang was observed at above 80 kHz, but there is no musical instrument on the planet that will ever get any amp to clip at such a frequency, and even if it did, we certainly couldn't hear it.

The overhang when an amp clips is caused by the output devices saturating (turning on as hard as they

possibly can). When a transistor saturates, it takes a finite time for the device to recover (when the base "charge" is depleted). Some of you will remember ECL (Emitter Coupled Logic) - these logic ICs are extraordinarily fast because the transistors never turn fully on or off, thus preventing the slow-down as the base charge is dissipated. Slightly off the topic, but interesting :-)

3.2 THS6012 CFB Pre-Amplifier

The other primary uses for the THS6012 amplifier IC are as a preamp, or as a line driver feeding the signal through the interconnects from preamp to power amp. The huge bandwidth and current capacity of this device make it ideal for a buffer/ line driver amp at the output of any preamp circuit. The series resistance can be reduced much further than with VFB opamps, and a value of 12 ohms is quite satisfactory to prevent the cable from causing oscillation

This does not mean that the cheapest and nastiest cable should be used (although it will not affect the sound). A well made and sturdy cable is always the ideal, but these can be home-made for a minimal cost, and when properly driven, will outperform anything on the market.

As a preamp, the CFB opamp has one drawback, and that is its input impedance. It is too low to interface with normal signal sources without an input buffer. The optimum configuration for a preamp is to use a conventional VFB opamp as the input stage, with the THS6012 as the final gain and line driver. This combines the best features of both types of opamp, and it is possible to make a preamp that will be better in all respects than any currently available (including passive versions). I know that's quite a claim, but having experimented at some length with the device, it is well justified.

A two stage preamp means that the VFB opamp will not need to contribute very much gain, and indeed, with two stages each having a gain of 2, the overall gain is adequate at 12 dB, and the bandwidth of such a combination can easily be made to exceed 200 kHz. With a medium input impedance (47k typical) and an output impedance of (say) 12 ohms, there is the potential to build the finest preamp in the world, limited only by your imagination and assembly skills.

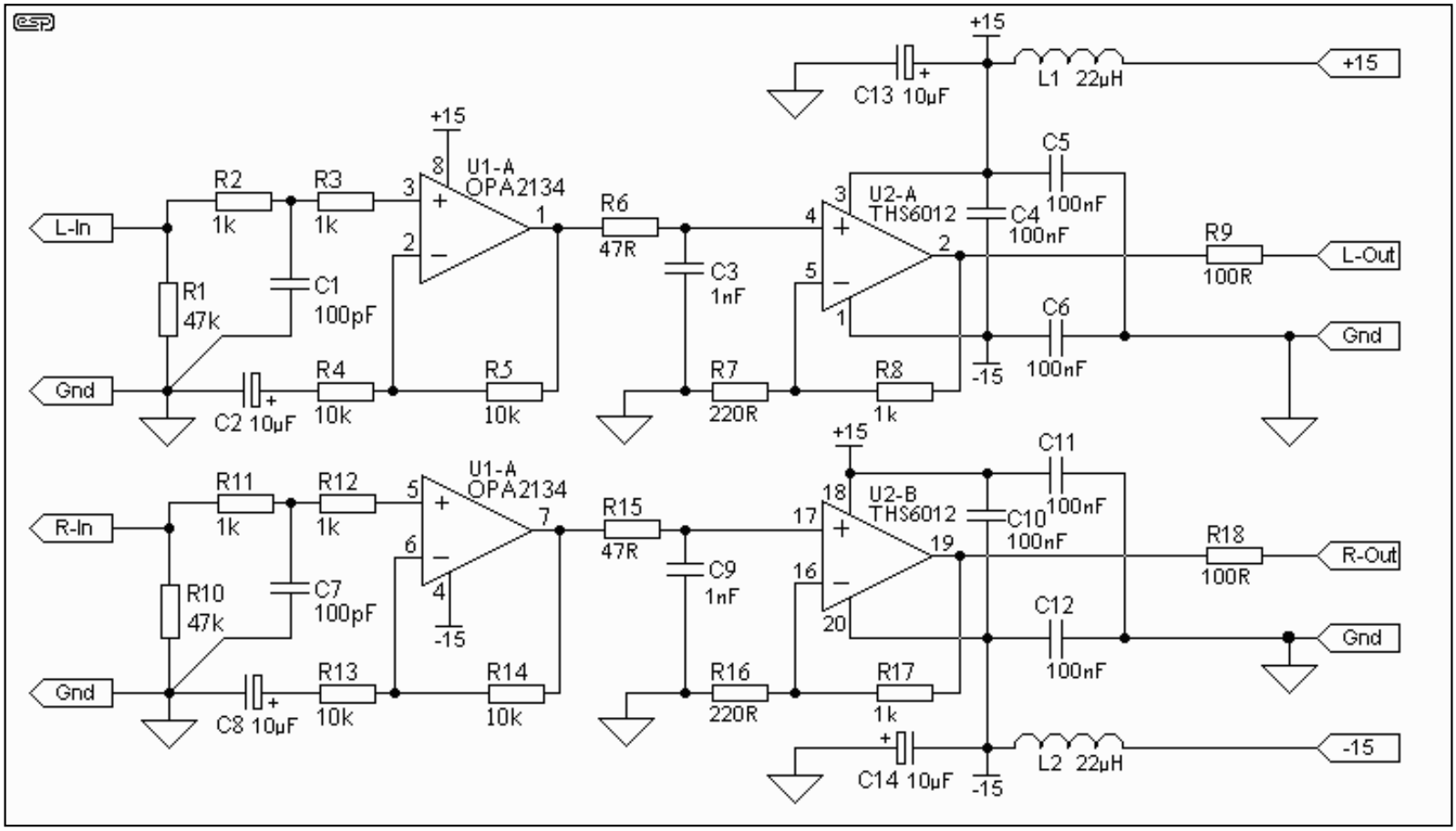


Figure 3.2 - Basic Concept of Preamp

Again, the circuit *looks* just like any other, and the output section is in fact almost identical to the headphone amp. This preamp has the ability to drive and terminate the output cable to well in excess of 500MHz. To my knowledge, there is no other hi-fi preamp available anywhere that has that ability. A Zobel network at the output is recommended - not for stability, but to ensure that the output impedance remains constant up to at least 500MHz. The inductance of a ceramic capacitor's leads (although very tiny indeed) will be sufficient to cause problems above this frequency, but this is still several orders of magnitude better than any other preamp available - regardless of price.

The OPA2134 input stage provides the necessary impedance conversion from the source to the input of the THS6012. This is a fast VFB opamp, and as shown has a gain of 2, or 6dB. The combined gain is about 20dB, which is more than enough, and may need to be reduced in many applications. Needless to say, a volume control is required, and with this arrangement you have little choice but to place it at the input. The same basic circuit can be used for the headphone amp as well - in effect they are identical applications, except for the higher current drain when driving headphones.

3.3 Zobel Network

The use of a Zobel network is well known in power amplifiers, and the most common is 10 ohms in series with 100nF, connected from the amplifier output to earth. This has a 3dB frequency of 159kHz, so above this frequency, where speaker leads are likely to present a high Q resonant circuit, the cable is effectively terminated by a low impedance that will effectively damp the cable resonance - at least from the amplifier's perspective.

That this has not been used in preamps before is somewhat surprising, but the most common (and generally

quite successful) method of preventing opamp oscillation is to use a 100 ohm resistor in series with the output. With lower resistances, this works with power amps too, but tends to reduce the available power and increases output impedance (thus lowering damping factor).

While the 100 ohm resistor (or other value, depending on the designer and the opamp) certainly works, it leaves the cable unterminated at RF, and increases the likelihood of interference pickup. The addition of a Zobel network to ensure that the cable remains terminated is risky with most opamps, since the effective impedance of the network will be very low, and because of the limited bandwidth, will need to become effective too close to the opamp's upper frequency limit and the upper end of the audio band.

This extra loading will stress the opamp - already straining to retain a flat response and with feedback falling at 6dB/octave. There is the very real likelihood that there will be an increase in distortion at high frequencies. This is not a problem in itself, but the intermodulation products will become an issue, and may be one of the reasons that opamps are often described in such derogatory terms by so many reviewers. Unfortunately (of course) most reviewers have no concept that there are different opamps with very different characteristics, so they gleefully lump them all into the same category.

A Zobel network intended to maintain an output impedance of 100 ohms at all frequencies will need to become effective at no less than about 150kHz for most conventional opamps, so the values would be 100 ohms in series with about 6.8nF, connected across the output as shown below.

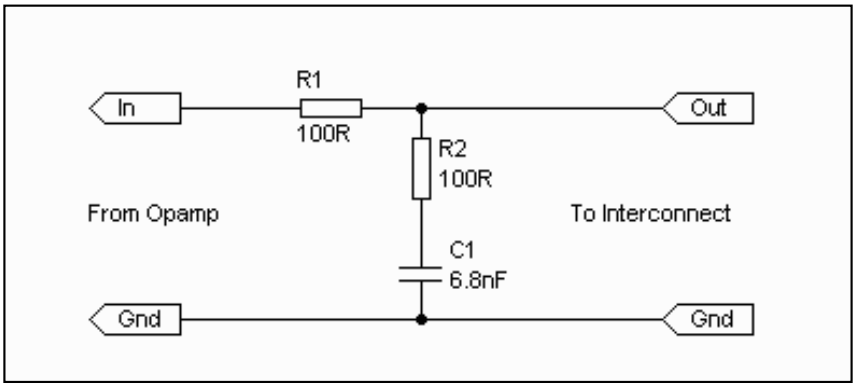


Figure 3.3 - Use of a Zobel Network for Preamps

With R1 = R2 = 100 ohms, and C1 = 6.8nF ...

This would be almost perfect for audio preamps, but the frequency is too low and the opamp loading too high! The -3dB frequency is about 160kHz, so the opamp must be able to retain its output impedance at a low value (less than 10 ohms) up to this frequency. Beyond that, the Zobel network will come into play, and will maintain the 100 ohm (or lower) termination up to that frequency where the capacitor's internal inductance is high enough to reduce the effectiveness of the circuit. Ceramic capacitors are absolutely essential for this, as they have the necessary bandwidth and low inductance that is needed.

The effect on the sound may not be minimal as we would hope, since although the capacitor is isolated from the signal path by the resistor, its reactance is significant at normal audio frequencies. At 20kHz, the cap has a reactance of 1.2k, so the opamp is quite heavily loaded at this frequency, increasing distortion.

The need for speed is again obvious - We need to use a faster opamp, and a much smaller capacitor. If the far end termination is also used, the loading (and distortion) will be much worse, and frequency response will suffer *within the audio band*.

With an ultra fast device such as the THS6012, the same circuit can be modified so the cable is matched better, and the Zobel network's influence will be so far outside the audio spectrum that it will have no audible effect whatsoever (other than reducing the cable's ability to act as a combined antenna and resonant circuit).

Substitute R1 = R2 = 68 ohms, and C1 = 1nF ...

Now, the -3dB frequency at the output is 1.6MHz, and the reactance of a 1nF capacitor is about 8k ohms at 20kHz. This is an insignificant load for any opamp, and the cable will be terminated by an impedance of not less than 68 ohms at all frequencies from DC up to several hundred megahertz. This will greatly reduce the tendency of the cable to act as an antenna, feeding RF interference into the output of the source amplifier, and into the input at the far end.

Another (identical) Zobel network across the input of the far end will maintain a reasonably good RF termination at both ends, but will have no deleterious effect for audio.

This should never be done unless you are certain that the preamp is capable of driving the impedance presented. A typical valve preamp with an output impedance of perhaps 10k, will be 3dB down at only 14kHz, and that is not allowing for the capacitance of the cable! The source impedance should ideally be less than 100 ohms.

The "far end" termination technique is intended only where wide bandwidth opamps having a very low output impedance and high drive capability are used as the line driver - it *will* cause problems with any source with an output impedance greater than about 200 ohms. With a 1k source impedance, an input circuit terminated as shown will be about 0.1dB down at 20kHz, with a 3dB frequency of about 150kHz. I do not believe that this would be good, and strongly recommend the use of a very fast, high drive capability, line driver IC.

Use of this technique is recommended, but only as part of a complete design, where the preamp and power amp (or electronic crossover) are designed to work together.

---

4. IC Mounting Details

The circuit details of both the headphone amp and preamp/ line driver are quite straightforward, but the device and its mounting are not. These ICs are only available as surface mount, which means that they are a little less friendly than "normal" ICs for the average constructor. The advantages are such that I believe the extra effort is warranted. There are considerable constraints on the PCB design, because of the extraordinarily wide bandwidth of the IC. The use of surface mount resistors is recommended for full bandwidth operation, but this further limits the design from an audio perspective. Conventional resistors can be used, but this must be done with care, or the amp will oscillate.

The THS6012 opamp IC uses TI's PowerPAD™ technology - there is a heatsink pad on the *underside* of the device, and a suitable thermal connection is important. Suffice to say at this time that using the method suggested by Texas Instruments is not recommended for home construction.

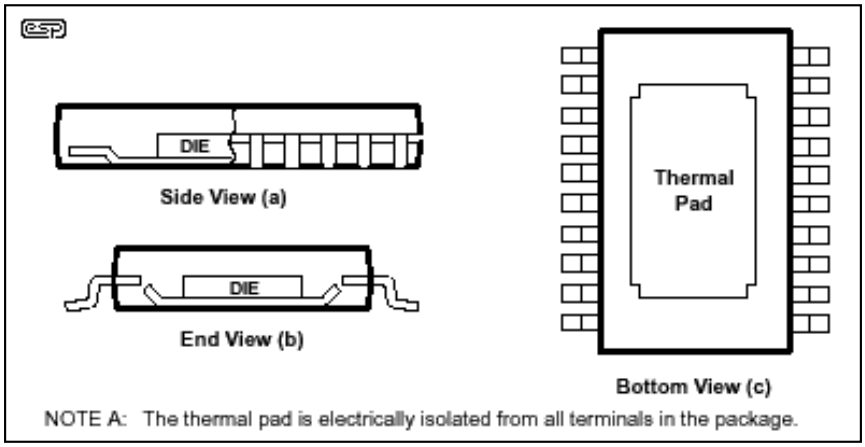


Figure 4.1 - PowerPAD Package Details

The thermal pad sits at the surface of the PCB, and with normal surface mounting and reflow soldering, it is easy enough to ensure that the pad is soldered to the board. A ground plane on the underside of the (double sided) PCB thus acts as a heatsink. This method of thermal management is ideal for automated assembly, but is very difficult for mere mortals to accomplish at home. I have devised a way of providing even better heatsinking that will be quite simple to do at home, and this will be described fully in the project article.

Proper power supply bypassing is essential, and the supply impedance must be as low as possible. This means ceramic bypass caps, as these are the only devices with the necessary size and bandwidth requirements. The size is important, because the bypass caps must be as close to the supply pins as possible - 2.5mm is the suggested maximum distance! The capacitors themselves must also have extremely low inductance, and ceramic devices are the only ones that will satisfy these requirements.

Otherwise, the circuits are unremarkable. They look just like any other opamp circuit, but with few slightly different component values. Well, that is until you listen to the results - there is nothing unremarkable about the sound, other than that it is as clean and un-coloured as if the amps were not there at all.

Conclusion

Does the foregoing mean that premium (conventional) opamps are a waste of time, and that the THS6012 should be the only amplifying device you will ever need? No. What I have attempted to convey here is that there is a place for very fast opamps, especially in areas where electromagnetic interference is a problem. These devices are also outstanding in all normal respects, and are (or should be) very worthy of consideration for your next preamp.

Simple preamps like the DoZ are not rendered obsolete either - the THS6012 just widens the selection, and provides some excellent benefits over the alternatives.

All this comes at a price however. The devices themselves are not cheap, and are much harder to mount on a PCB than the alternatives. Availability may also be an issue, and we can be certain that the normal retail electronics outlets will not stock such esoteric devices unless there is a considerable demand.

As a headphone amplifier, these opamps are superb, with not a single negative aspect. I really doubt that there is a better headphone amp on the planet - and yes, I really do mean that! Is it all worth the effort? That, I shall leave up to you.

I would like to thank Texas Instruments for taking the time to contact me, and for providing test devices and documentation for their range of high speed opamps and balanced line drivers and receivers.

Terminology

Slew Rate - the maximum rate of change of voltage for an active device. Slew rate is measured in Volts per microsecond, so an amplifier that can swing its output from +20V to -20V in one microsecond has a slew rate of 40 V/us

Characteristic Impedance - of a cable, that impedance determined by the diameter of the inner and outer conductors, and their relative spacing. Typical examples are 300 ohm TV ribbon cable, and 75 ohm coax

Velocity Factor - Electrical signals travel at the speed of light in a vacuum, but when trapped in a cable, their speed decreases. The decrease depends on the cable construction, and as an example, the velocity factor of 75 Ohm coax is generally about 0.8, meaning that the signal travels at only 0.8 of the speed of light (0.8 * $3*10^8$ = $2.4*10^8$ metres / second)

Narrow Band - RF signals that are confined to a relatively narrow frequency range (e.g. 10kHz either side of the carrier signal for AM radio).

Broad Band - Interference signals that spread over a wide frequency range, sometimes from the mains frequency all the way into the upper RF bands. Anything that creates arcs (such as most switches, welders, etc.) generates broad band interference.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Sections of the text may also be (or describe) the intellectual property of Texas Instruments. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 06 April 2001

The Audio Pages

Elliott Sound Products **Class-D Amplifiers**

Class D Audio Amplifiers - Theory and Design

© June 2005, Sergio Sánchez Moreno (ColdAmp)

Edited & Additional Text by Rod Elliott (ESP)

Page Created 04 June 2005

Articles Index

Main Index

Contents

- 1 - Introduction
- 2 - How Class-D Works
- 3 - Topologies
- 4 - Gate Driving
- 5 - Level Shifting
- 6 - Output Filter Design
- 7 - Feedback
- 8 - Other topologies
- 9 - Some Final Notes

---

1 - Introduction

A completely new technology for audio amplification has been evolving during the last 15-20 years that has a clear benefit over current widespread Class-A, and AB topologies. We are talking about the so-called “Class-D”. This benefit is mainly its high

power efficiency. Figure 1 shows typical efficiency curves vs. Output power for Class-B and Class-D designs.

The theoretical maximum efficiency of Class-D designs is 100%, and over 90% is attainable in practice. Note that this efficiency is high from very moderate power levels up to clipping, whereas the 78% maximum in Class-B is obtained at the onset of clipping. An efficiency of less than 50% is realised in practical use with music signals. The PWM amp's high power efficiency translates into less power consumption for a given output power but, more important, it reduces heatsink requirements drastically. Anyone who has built or seen a high-powered audio amplifier has noticed that big aluminium extrusions are needed to keep the electronics relatively cool. The loading on the power transformer is also reduced by a substantial amount, allowing the use of a smaller transformer for the same power output.

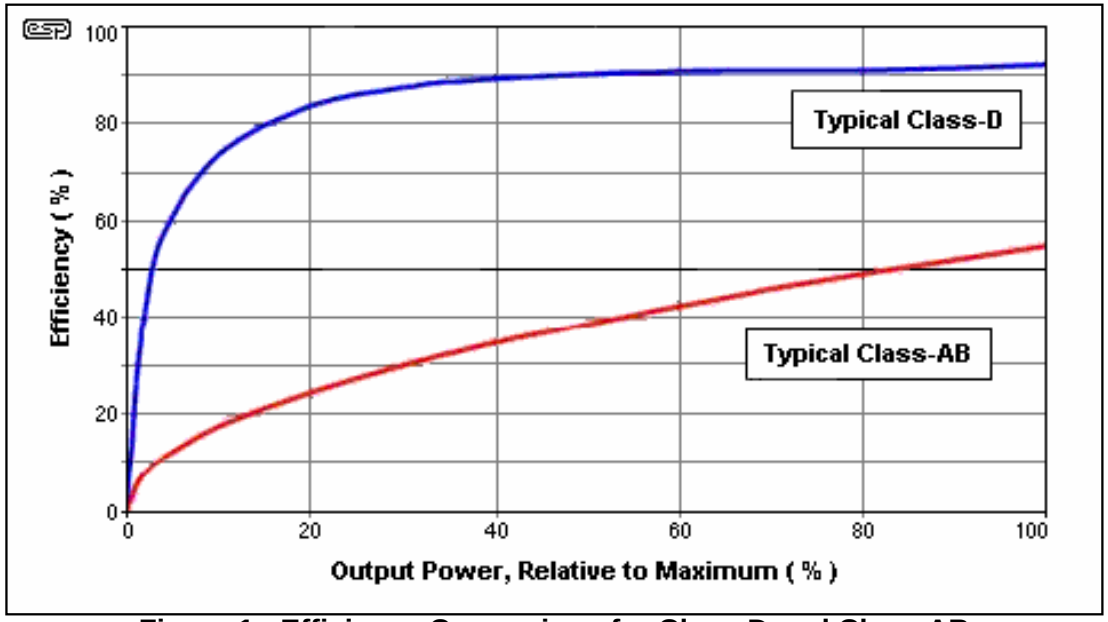


Figure 1 - Efficiency Comparison for Class-D and Class-AB

These heatsinks account for an important part of the weight, cost and size of the equipment. As we go deeper in the details of this topology, we will notice that a well behaving (low distortion, full range) Class-D amplifier must operate at quite high frequencies, in the 100KHz to 1MHz range, needing very high speed power and signal devices. This has historically relegated this class to uses where full bandwidth is not required and higher distortion levels are tolerable - that is, subwoofer and industrial uses.

However, this has changed and thanks to today's faster switches, knowledge and the use of advanced feedback techniques it is possible to design very good performance Class-D amplifiers covering the whole audio band. These feature high power levels, small size and low distortion, comparable to that of good Class-AB designs. (From now on, I will refer to Class-A and AB topologies as “classical”).

Complete 400W Full-Range Class-D Amplifier Module (Courtesy of ColdAmp)

From the DIY perspective, Class-D is rather unfortunate. Because of the extremely high switching speeds, a compact layout is essential, and SMD (surface mount devices) are a requirement to get the performance needed. The stray capacitance and inductance of conventional through-hole components is such that it is almost impossible to make a PWM amplifier using these parts. Indeed, the vast majority of all ICs used for this application are available *only* in surface mount, and a look at any PWM amplifier reveals that conventional components are barely used anywhere on the board. Since SMD parts are so hard to assemble by hand and the PCB design is so critical to final performance, DIY versions of PWM amps are very rare indeed (I don't know of any).

---

2 - How Class-D Works

In classical amplifiers, at least one of the output devices (let them be bipolar transistors, MOSFETs or valves) is conducting at any given time. No problem so far, but they are also carrying a given current where there is a voltage drop between collector-emitter / drain-source etc. Since P = V * I, they are dissipating power, and even if there is no output a small quantity of current must pass through the transistors to avoid crossover distortion, so some dissipation is present. As the output voltage increases, for given supply rails the voltage drop across the transistors will fall, but the current increases. At saturation (clipping), VCE or VDS will be low, but current is quite high (Vout / Rspk). Conversely, at low power levels, current is small but voltage drop is large. This leads to a power dissipation curve that is not linear with output power. There is a non-zero minimum dissipation (zero percent efficiency), and a point where maximum efficiency is reached ... about 78% in pure Class-B designs, 25% or less with

Class-A.

Class-D on the other hand, bases its operation in switching output devices between 2 states, namely “on” and “off”. Before discussing the topology specific details, we can say that in the “on” state, a given amount of current flows through the device, while theoretically no voltage is present from drain to source (yes, almost every Class-D will use MOSFETs), hence power dissipation is theoretically zero. In the off state, voltage will be the total supply rails as it behaves like an open-circuit, and no current will flow (that's very close to reality).

But how can our beloved audio signal be represented by an awful square wave with only two possible levels? Well, in fact it modulates some characteristics of this square wave so the information is there. Now we “only” have to understand the way the modulation is done and how to restore the amplified audio signal from it. The most common modulation technique used in Class-D is called PWM (Pulse Width Modulation) - a square wave is produced that has a fixed frequency, *but* the time it is in the “high” and “low” states is not always 50%, but it varies following the incoming signal. This way, when the input signal increases, the “high” state will be present for longer than the “low” state, and the opposite when the signal is “low”. If we do some maths, the mean value of the signal in a single cycle is simply ...

Vmean = Vhigh * D + Vlow * (1-D), where D = Ton / T, (duty cycle)

T being the period of the signal, that is, 1/Fsw (switching frequency).

For example, the mean value of a 50% duty cycle (both states are present for exactly the same amount of time) signal going from +50V to –50V is: 50 * 0.5 + (- 50) * 0.5 = 0V. In fact, the idle (no signal) output of a Class-D amplifier is a 50% duty cycle square signal switching from the positive to the negative rail.

If we modulate the input up to the maximum, we will have a near-100% duty-cycle. Lets put 99%: Vmean = 50 * 0.99 + (-50) * 0.01 = 49V. Conversely, if the signal is lowest, we need near 0% (lets use 1%), so Vmean = -49V.

PWM is usually generated by comparing the input signal with a triangle waveform as shown in Figure 2. The triangle wave defines both the input amplitude for full modulation and the switching frequency

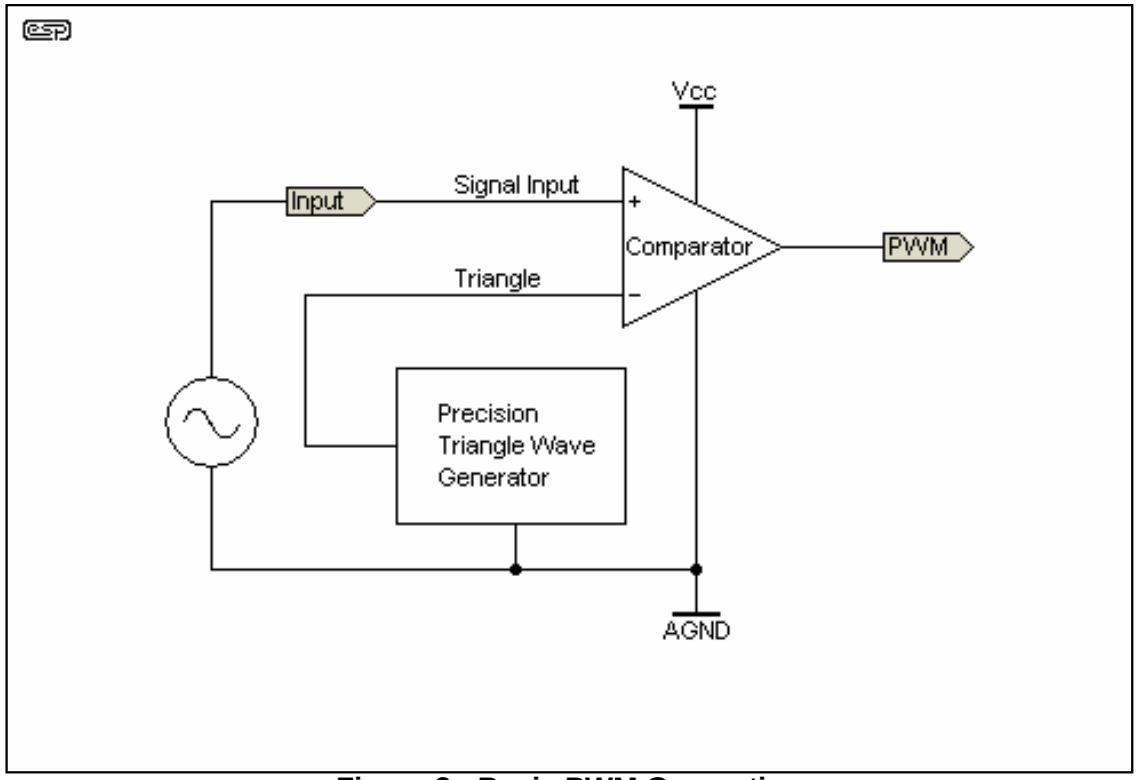


Figure 2 - Basic PWM Generation

Figure 3 shows a typical PWM signal modulated by a sine wave. Notice that it is designed so signals between –1 and 1V will produce 0% to 100% duty cycles, 50% corresponding to 0V input. The 'digital' output uses standard logic levels, where 0V is a logic '0' and 5V is a logic '1'. Because of this digitisation of the signal, PWM amps are sometimes erroneously referred to as digital amps. In fact, the entire process is far more analogue than digital.

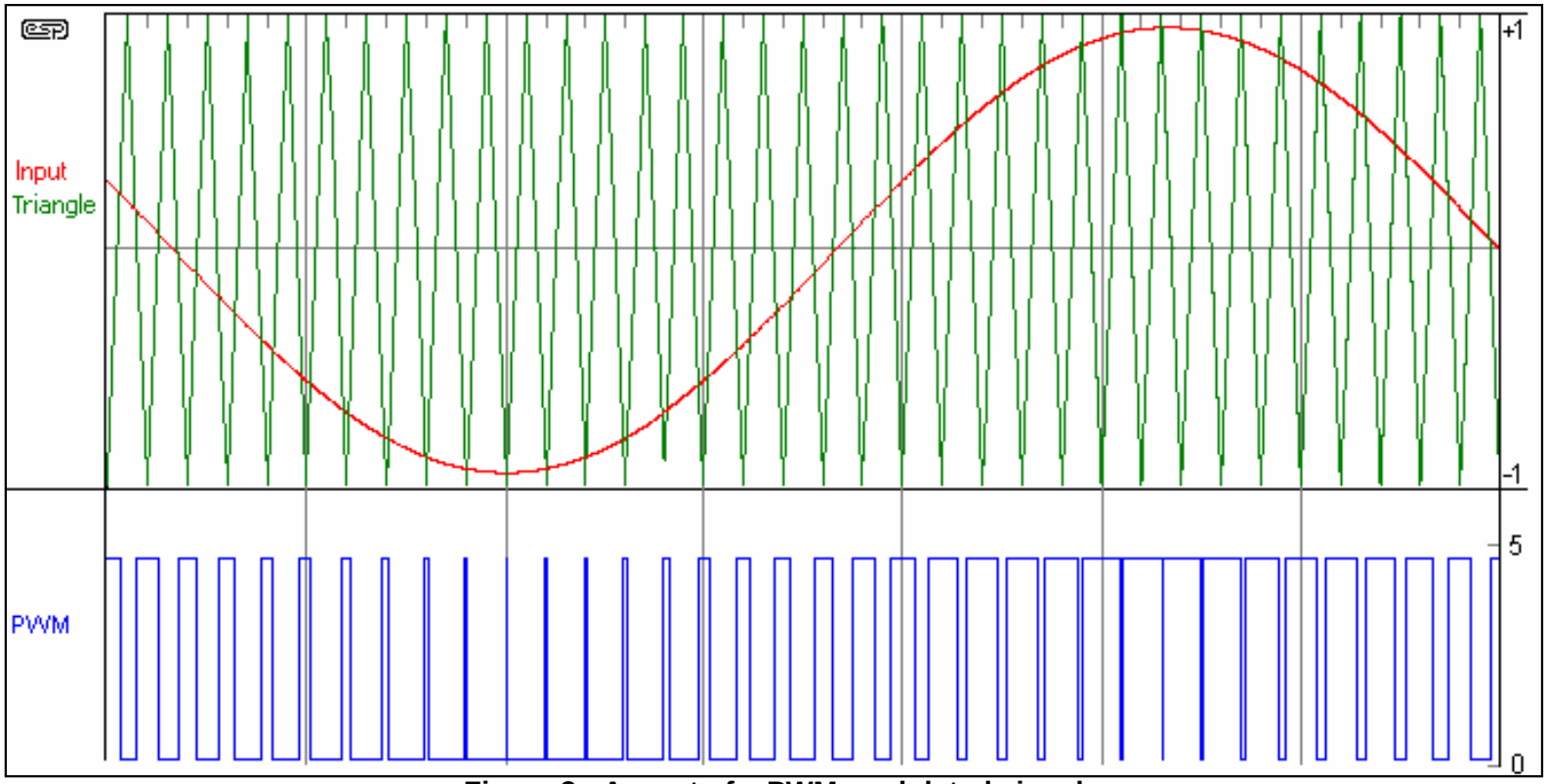


Figure 3 - Aspect of a PWM modulated signal

Notice that for a correct representation of the signal, the frequency of the PWM reference waveform must be much higher than that of the maximum input frequency. Following Nyquist theorem, we need at least twice that frequency, but low distortion designs use higher factors (typically 5 to 50). The PWM signal must then drive power conversion circuitry so that a high-power PWM signal is produced, switching from the +ve to –ve supply rails (assuming a half-bridge topology).

The spectrum of a PWM signal has a low frequency component that is a copy of the input signals spectrum, but also contains components at the switching frequency (and its harmonics) that should be removed in order to reconstruct the original modulating signal. A power low-pass filter is necessary to achieve this. Usually, a passive LC filter is used, because it is (almost) lossless and it has little or no dissipation. Although there must always be some losses, in practice these are minimal.

---

3 - Topologies

There are basically two Class-D topologies - half-bridge (2 output devices are used) and full-bridge (4 output devices). Each one has its own advantages. For example, half-bridge is obviously simpler and has more flexibility as a half-bridge amplifier can be bridged as with classical topologies. If it is not correctly designed and driven, can suffer from “bus pumping” phenomena (transfer

current to the power supply that can make it increase its voltage producing situations dangerous to the amplifier, supply and speaker).

Full bridge requires output devices rated for half the voltage as an half bridge amplifier of the same power, but it is more complicated. Figures 5a and 5b show both topologies conceptually. Obviously, many components such as decoupling capacitors, etc are not shown.

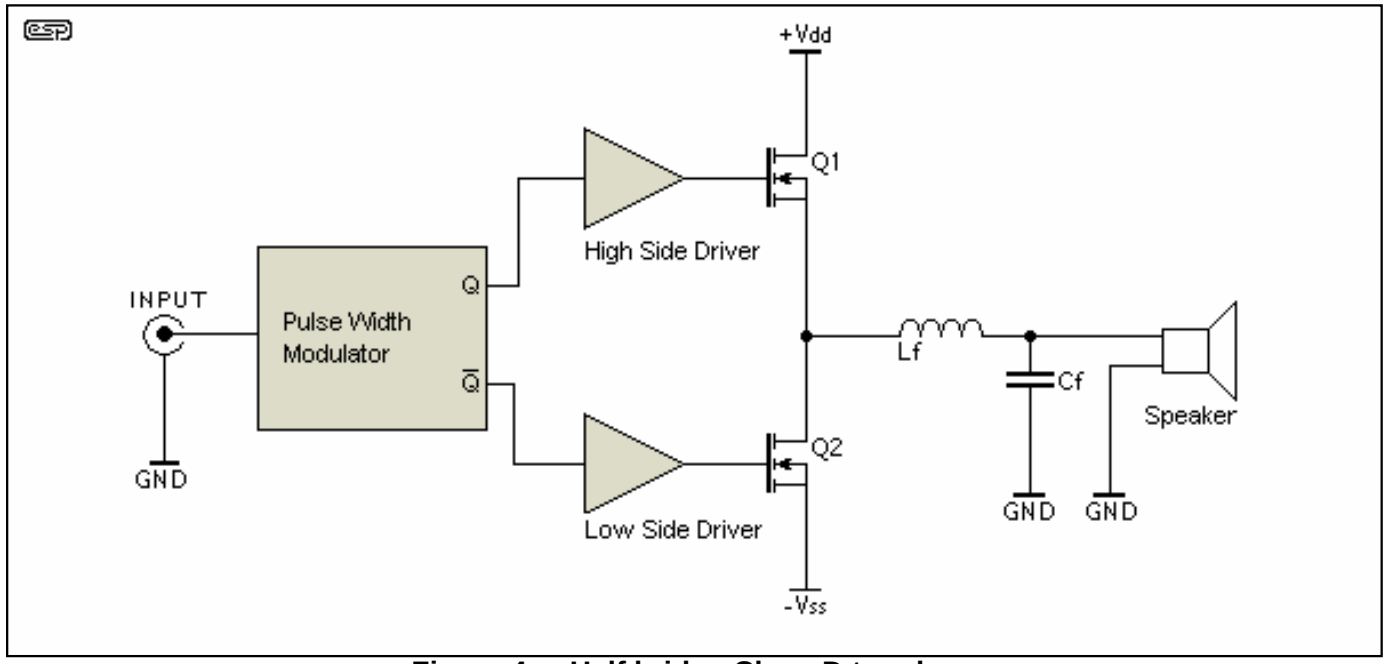


Figure 4a - Half bridge Class-D topology

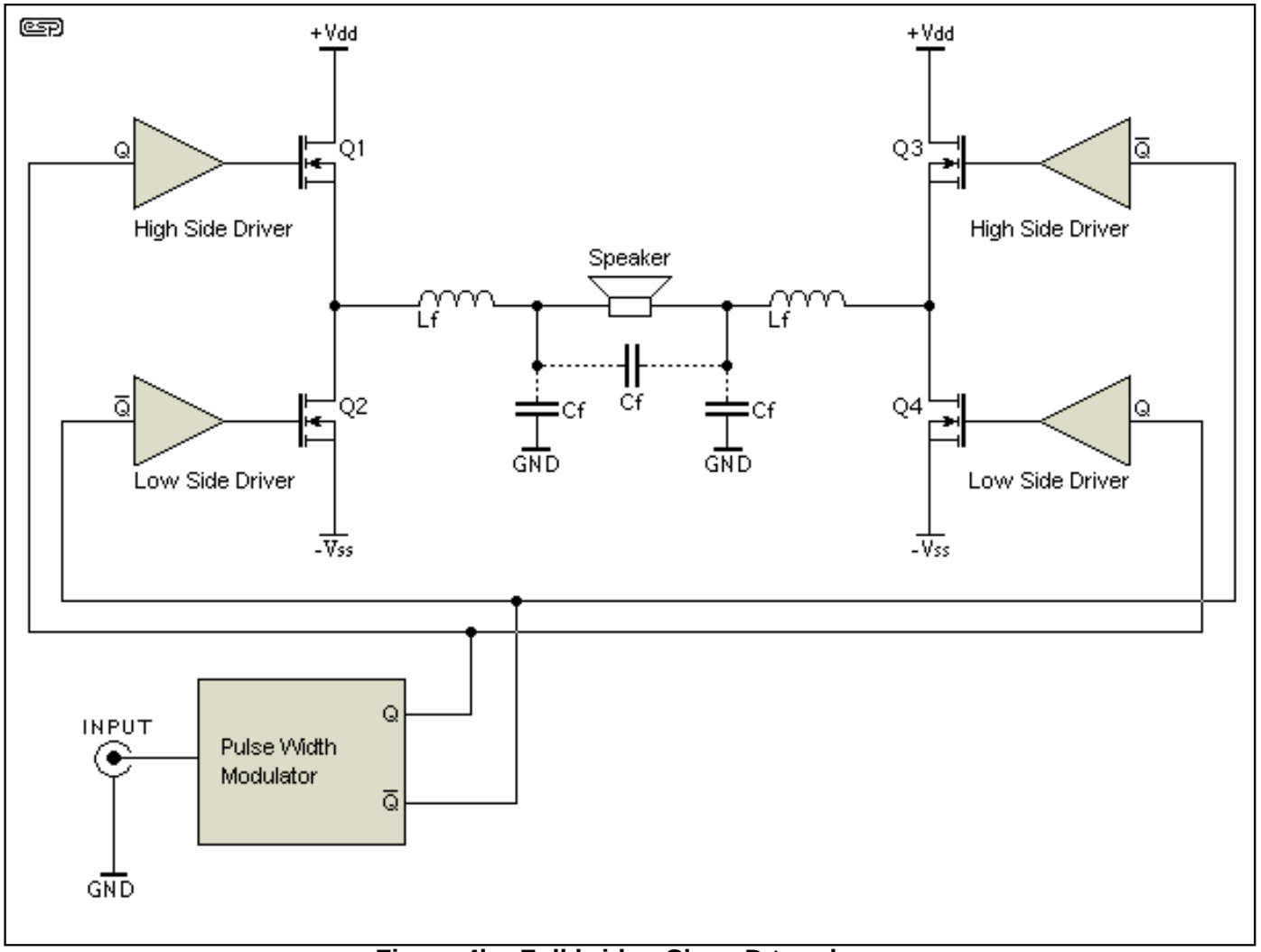


Figure 4b - Full bridge Class-D topology

Note that full bridge PWM amp needs only one supply rail - bipolar supplies are not necessary, but can still be used. When a single supply is used, each speaker lead will have ½ the Vdd voltage present. As it is connected differentially, the loudspeaker doesn't see any DC if everything is well balanced. However, this can (and does) cause problems if a speaker lead is allowed to short to chassis!

The filter may be implemented by means of a single capacitor across the loudspeaker, by a pair of caps to ground, or in some cases by both (as shown by the dotted lines connecting the caps).

For the rest of the document, we will concentrate on half-bridge topologies, although the vast majority of the ideas are also applicable to full-bridge designs.

Half bridge design

The operation of the half bridge circuit depicted in Figure 4a is as follows ...

When Q1 is on (corresponding to the positive part of the PWM cycle), the switching node (inductor input) is connected to Vdd, and current starts to increase through it. The body diode of Q2 is reverse biased. When Q2 is on (negative part of the PWM cycle), the body diode of Q1 is reverse biased and the current through Lf starts to decrease. The current waveform in Lf is triangular shaped.

Obviously, only one of the transistors must be on at any time. If for any reason both devices are enhanced simultaneously, an effective short-circuit between the rails will be produced, leading to a huge current and the destruction of the MOSFETs. To prevent this, some “dead-time” (a small period where both MOSFETs are off) has to be introduced.

Lf in conjunction with Cf and the speaker itself form the low pass filter that reconstructs the audio signal by averaging the switching node voltage.

Timing is critical in all this process: any error as delays or rise-time of the MOSFETs will ultimately affect efficiency and audio quality. So all the involved components must be high-speed. Dead-time also affects performance, and it must be minimised. At the same time, the dead-time must be sufficiently long to ensure that under no circumstance both MOSFETs are on at the same time. Typical values are 5 to 100ns.

The dead-time is a critical factor for distortion performance. For lowest distortion, the dead-time must be as small as possible, but this risks 'shoot-through' currents, where both MOSFETs are on simultaneously. This not only increases distortion and dissipation dramatically, but will quickly destroy the output devices. If the dead-time is too great, the response of the output stage no longer follows the true PWM signal generated in the modulator, so again distortion is increased. In this case, dissipation is not affected.

4 - Gate Driving

To ensure fast rise/fall times of the MOSFETs, the gate driver must provide quite a high current to charge and discharge the gate capacitance during the switching interval. Typically, 20 - 50ns rise/fall times are needed, requiring more than 1A of gate current.

Note that the schematics shown use both N-channel MOSFETs. Although some designs use N and P channel complementary devices, that is IMHO sub-optimal due to the difficulty of obtaining suitable P devices and matched pairs. So lets concentrate on N-channel only half-bridges. Note that, in order to drive a MOSFET on, a voltage above Vth must be present between its gate

and source. The lower MOSFET has its source connected to –Vss, so its drive circuit has to be referred to that node instead of GND.

However, the upper MOSFET is more difficult to drive, as its source is continuously floating between +Vdd and –Vss (minus drops due to on resistance). However, its driver must be also floating on the switching node and, what's more, for the on-state, its voltage must be several volts above +Vdd so a positive Vgs voltage is created when Q1 is on. This also implies a voltage shifting so the modulator circuit can communicate correctly with the driver.

This is one of the major difficulties of Class-D design: gate drive. To solve the issue, several approaches are taken ...

- Transformer gate drive: useful in half-bridge power supplies where duty-cycle doesn't vary widely. In audio amplifiers however, duty cycle ranges from 0% to 100%, so this method creates a problem because the signal is AC coupled. A DC restoration circuit (not shown) is needed.
- Discrete gate driving: some designs use transistors to perform both the level shifting and the MOSFET drive. Again, there is a problem: we need a voltage that is higher than +Vdd.
- Integrated drivers: there are a number of MOSFET drivers in the market, optimised for high speed, that can be used. Again, a voltage higher than Vdd is needed as well as level shifting.

Figure 5 (a, b & c) depict some possibilities for 'High Side' gate driving ...

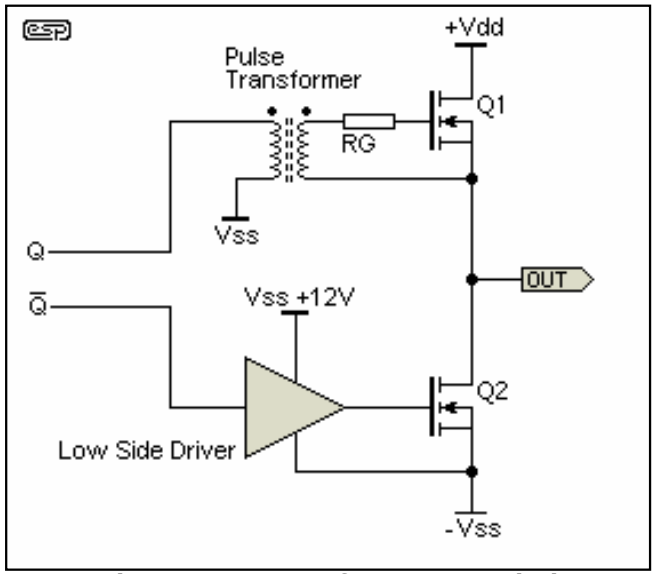


Figure 5a - Transformer Coupled

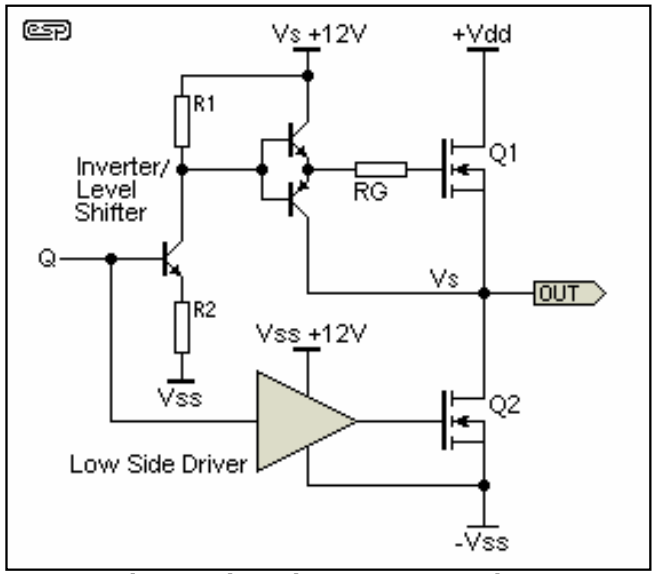


Figure 5b - Discrete BJT Driver

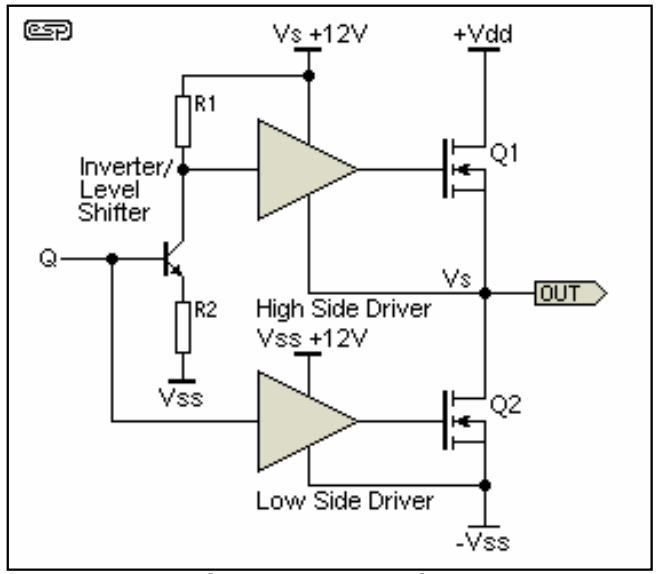


Figure 5c - IC Driver

Note that circuits in figures 5b and 5c have their PWM input referred to –Vss so may require previous level shifting of the

comparator output, that will normally be referred to GND. Fig 5a will require level shifting of the inverted PWM only, as the transformer input can be referenced to GND as shown. Many of the driver ICs available now have inbuilt level shifters, and these are optimised for speed. Remember that *any* delay introduced into the switching waveform can cause distortion or simultaneous MOSFET conduction.

We have still one problem to solve ... obtaining 12V above VS (the switching node). We can add another power supply, isolated from the main one, which (-) is connected to VS. This solution can be impractical, so other techniques are commonly used. The most widespread is a 'bootstrap' circuit. The bootstrap technique uses a charge pump built with a high speed diode and a capacitor. The output of the amplifier produces the switching pulses needed to charge the capacitor.

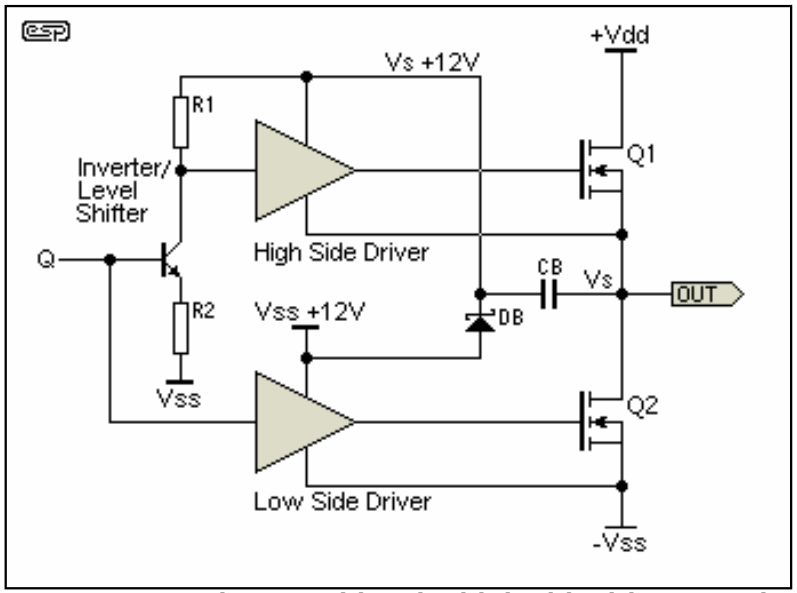


Fig. 6 Bootstrap capacitor provides the high side driver supply voltage

This way, the only auxiliary power supply needed is 12V referenced to –Vss that is used for powering both the low side driver and the charge pump for the high side driver. As the average current from this supply is low (although there are high current charging peaks during the switching events, they last only 20-50ns, twice during a cycle, so the average is quite low, in the 50-80mA range), this supply is easily obtained from the negative rail with a simple 12V regulator (paying attention to its maximum input voltage rating, of course).

---

5 - Level Shifting

As can be seen from the previous figures, in order to excite the MOSFET driver, the PWM signal has to be referred to –Vss. So, as the modulator usually works from +/-5 to +/-12V, typically, a level shifting function is needed. One can choose to shift the level

of the PWM signal and then generate the inverted version, or generate both outputs and invert both of them. It depends, for example, on the comparator type used (if complementary outputs are available, the decision is made).

A basic level shifting function can be performed with a single or two-transistor circuit similar to the one depicted in Figure 6 (before the high side driver). While this may work at low frequencies, it is important to simulate the behaviour of the comparator and level shifter, as they can introduce considerable delays and timing errors if not properly designed.

It is fair to say that the level shifter is one of the most critical parts of the circuit, and this is evidenced by the wide variety of competing ICs designed for the job. Each will have advantages and disadvantages, but in all cases the complexity is far greater than may be implied by the simplified diagrams.

---

6 - Output Filter Design

The output filter is one of the most important parts of the circuit, as the overall efficiency, reliability and audio performance depends on it. As previously stated, a LC filter is the common approach, as it is (theoretically) lossless and has a –40dB/decade slope, allowing for a reasonable rejection of the carrier if the parameters of the filter and the switching frequency itself are properly designed.

The first thing to do is to design the transfer function for the filter. Usually, a Butterworth or similar frequency response is chosen, with a cutoff frequency slightly above the audio band (30-60KHz). Have in mind that one of the design parameters is the termination load, that is, the speaker impedance. Usually, a typical 4 or 8 ohm resistor is assumed, but that would produce variations in the measured frequency response in presence of different speakers. That must be compensated for by means of proper feedback network design. Some manufacturers simply leave it that way so the response is strongly dependent on the load. Surely a non-desirable thing.

The design can be done mathematically or simply use one of the many software programs available that aid in the design of LC filters. After that, a simulation is always useful. Figure 7 shows a typical LC filter for Class-D amplifiers and its typical frequency response.

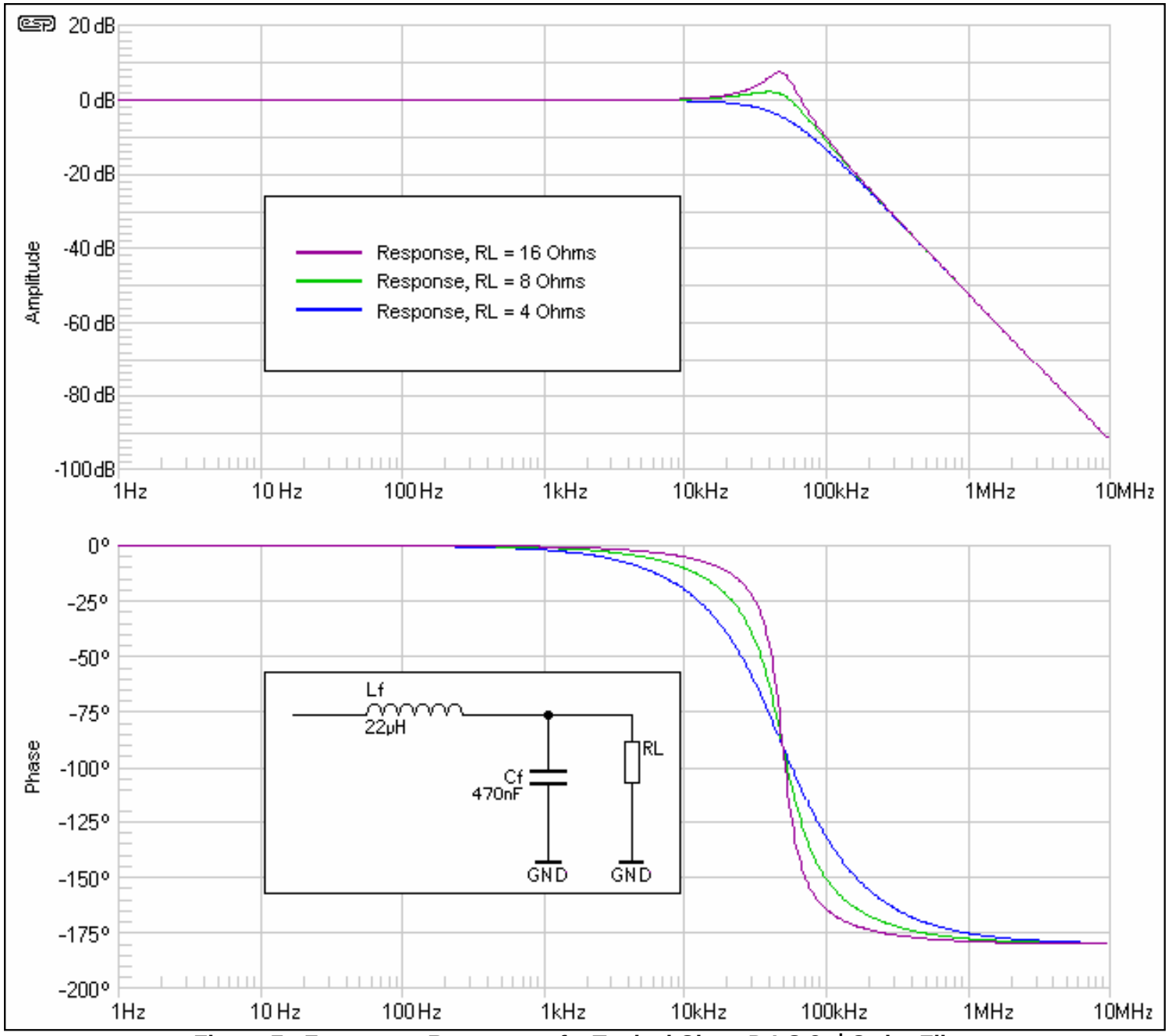


Figure 7 - Frequency Response of a Typical Class-D LC 2nd Order Filter

This simple filter has a –3dB cutoff frequency of 39KHz (with 4 ohm load), and suppresses the carrier as much as 31dB at 300KHz. For example, if our supply rails are +/-50V (enough for about 275W at 4 ohms), the residual ripple will have an amplitude of about 1Vrms.

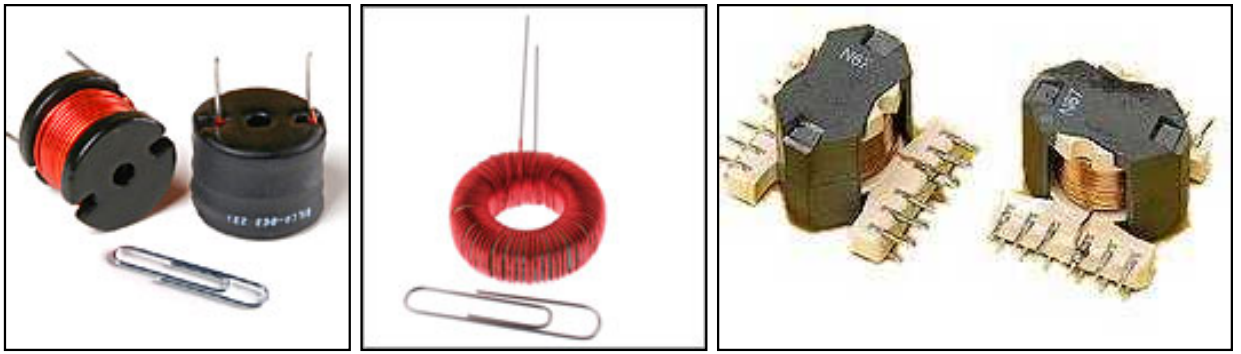
This ripple is, obviously inaudible, and 1V RMS will dissipate only around 200mW in a typical tweeter (not likely a problem, especially since the tweeters impedance will be a lot higher than 8 ohms at 300kHz). However, care must be taken as the speaker wires can become an antenna and affect other equipment. In fact, although a couple of volts RMS of ripple can seem low enough to run your speakers safely, EMI can be a concern, so the less carrier level you have, the better. For further rejection, higher order filters are used (with the potential disadvantage of increased phase shift in the audio band), although there are other clever ways to do it, as very selective bandstop or 'notch' filters tuned to the carrier frequency (if it is fixed, and that only happens in synchronous designs as the one described).

Well designed Class-D amplifiers have a higher order filter and/or special carrier suppression sections in order to avoid problems with EMI. As can be seen in Figure 8, the response is dependent on the load, and in fact the load is part of the filter. This is one of the problems to solve in Class-D designs. It doesn't help that a loudspeaker presents a completely different impedance to the amplifier than a test load, and many PWM amps have filters that are (and never can be) correct for all practical loudspeaker loads. Again, only a handful of good Class-D amplifiers use feedback techniques that include the output filter to compensate for impedance variations and have a nearly load independent frequency response, as well as to reduce distortion produced by non-linearities in the filter. Although passive components are thought to be distortion-free, this does not apply to ferrite or powdered iron cores that are used for the filters. These components most certainly do introduce distortion.

Now, The Filter Components ...
The output inductor has to withstand the whole load current, and also have storage capability, as in any non-isolated switching converter (Class-D half bridge design is in fact analogous to a buck converter, its reference voltage being the audio signal).

The ideal inductor (in terms of linearity) is an air-core one, but the size and number of turns required for typical Class-D operation usually makes it impractical, so a core is normally used in order to reduce turns count and also provide a confined magnetic field that reduces radiated EMI. Powder cores or equivalent materials are the common choice. It can also be done with ferrite cores, but they must have a “gap” where energy is stored. Wire size must also be carefully chosen so DC losses are low (requiring thick wire) but also skin effect is reduced (AC resistance must also be low).

Inductor core shape can be a drum core, gapped ferrite RM core, or toroidal powder core, among others. Drum cores have the problem that their magnetic field is not enclosed, hence producing more radiated EMI. RM cores solve this problem but have most of the coil enclosed, so cooling problems may arise as no airflow is possible. IMHO, toroids are preferred because they feature both closed magnetic field that helps control radiated EMI, a physically open structure that allows proper cooling, and easy and economical winding, as they don't need bobbins.



Coil Shapes ... Drum, Toroid & RM Style Coils and Cores (Wilco & Coilcraft)

Many core manufacturers such as Micrometals or Magnetics offer their own software, very useful to design the output inductor as they help choosing the right core, wire size and geometrical parameters. The capacitor usually falls in the 200nF to 1uF range, has to be of good quality too, in order to ensure good high frequency behaviour and low losses and of course has to be rated for the whole output voltage. Usually, polypropylene capacitors are chosen. Needless to say, you cannot use electrolytics!

---

7 - Feedback

As I have stated previously, timing errors can lead to increased distortion and noise. This cannot be skipped and the more precise it is kept, the better the design will perform. Anyway, open loop Class-D amplifiers are not likely to satisfy demanding specifications, so (negative) feedback is almost mandatory. There are several approaches. The most simple and common is to take a fraction of the switching signal, precondition it by means of a passive RC low pass filter and feed it back to the error amplifier.

To put it simply, the error amplifier is an opamp placed in the signal path (before the PWM comparator) that sums the input signal with the feedback signal to generate a error signal that the amps automatically minimises (this is the concept of every negative-feedback system, anyway).

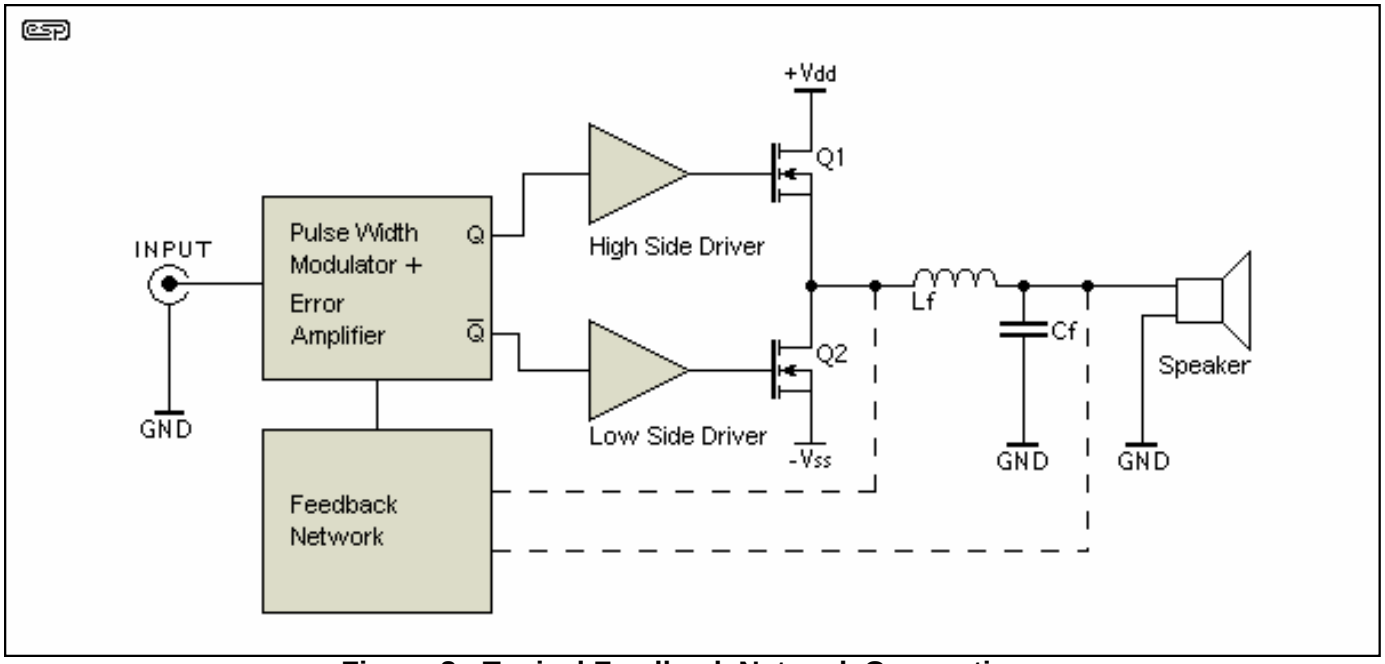


Figure 8 - Typical Feedback Network Connections

Although good results are obtained this way, there is still a problem: load dependency, due to the speaker being an integral part of the filter, hence affecting its frequency response as shown above.

Some more advanced amplifiers take the feedback signal from the very output, trying to compensate this. This way, a constant frequency response is obtained, with the further gain that the inductor resistance contributes much less to the output impedance, so it is kept lower, hence damping factor is higher (higher speaker control). However, taking feedback after the filter is not an easy task. The LC introduces a pole and hence a phase shift that, if not properly compensated, will make the amp become unstable and, ultimately, oscillate. Feedback may be taken from both the switching node and the filter output. Although this can give very good results, it is still difficult to maintain stability because of the phase shift through the output filter.

---

8 - Other topologies

Pure PWM (based on triangle generators, also called 'natural sampling PWM') is not the only way to go in order to construct a Class-D amplifiers. Some other topologies have arisen, many of them based on auto-oscillation, where the hysteresis in the comparator and delays between the comparator and power stage can be taken into account to design a system that oscillates by itself in a somewhat controllable manner.

Although simpler, these designs have some disadvantages, IMHO. For example, the switching frequency is not fixed, but

depends on the signal amplitude. This makes output notch filters ineffective, yielding higher ripple levels. Besides, when several channels are put together, the difference in switching frequency between them can produce audible beat tones that can become very annoying. This can also happen of course with synchronous design as the one described here, but there is a simple solution - use the same clock for all the channels.

Self oscillating designs in particular have some other difficulties like start-up: special circuitry may be needed that forces the amplifier to start oscillating. Conversely, if for any reason the oscillation stops, you could end up with an 'always-on' MOSFET, and thus a large amount of DC at the output, followed almost immediately by a dead loudspeaker. Of course, these issues can be solved with proper design, but the added complexity can void the initial simplicity, thus no gain is obtained.

Low distortion in a PWM amplifier requires a very linear triangle waveform, along with a very fast and accurate comparator. At the high operating frequencies needed for optimum overall performance, the opamps used need to have a wide bandwidth, extremely high slew rate, and excellent linearity. This is expensive to achieve, requiring premium devices. Some of these constraints are relieved somewhat by self oscillating designs (therefore making them slightly cheaper), but this is not an effective trade-off for the most part.

Clocked designs (fixed frequency) are not easier to make than self-oscillating or modulated switching frequency designs, but are certainly far more predictable and tend to have fewer problems overall. The ability to synchronise multiple amplifiers ensures that mutual interference is minimised. An 'advantage' claimed by the proponents of non-clocked and 'random switching' designs is that the RF energy on the speaker leads is spread over a wide frequency range, potentially making such amplifiers more likely (or perhaps less unlikely) to pass EMI testing. From an overall perspective, this is more likely to be a hinderance than a benefit, as it is no longer possible to optimise the filter network for maximum switching frequency rejection.

There are also PWM amps that claim to be truly 'digital', using One-Bit™ technology, or generating the PWM signal directly from the PCM data stream. Although the manufacturers of such amplifiers will naturally proclaim their superiority over all others, such self-praise should generally be ignored. Implementing feedback in a 'pure' digital design is at best difficult, and may be impossible without using a DSP (digital signal processor) or resorting to an outboard analogue feedback system. Including additional ADCs and DACs (analogue to digital converters and vice versa) is unlikely to allow the amplifier to be any 'better' than the direct analogue methods described in this article.

A relative newcomer to the scene is the Sigma-Delta modulator, however at the time of writing this still has problems (challenges in corporate speak). The main issue is that the transition rate is too high, and it must be reduced to accommodate real-world components - particularly the power switching MOSFETs.

The 'pure' digital solutions described above have another shortfall, and that's the fact that the number of different pulse widths is finite, and determined by the clock speed. A digital system can only switch on a clock transition. Based on currently available information, only around 8 x oversampling is possible if a digital noise shaping filter is added to the system. An analogue

modulation system has an effectively infinite number of different pulse widths, but this is not possible with any true digital implementation.

These latter comments cover a very complex area, one is outside the scope of this article. However, even the scant information above will give most readers far more information that is commonly available - especially from manufacturers of digital Class-D amplifiers.

---

9 - Some Final Notes

In conclusion, Class-D amplifiers have evolved a lot since they were first invented, achieving levels of performance similar to conventional amplifiers, and even better in some aspects, like an inherent low output impedance that allows effortless bass. All this, with the great advantage of high efficiency. Of course, only if they are properly designed.

However, although very attractive, Class-D designs are not very DIY friendly. In order to achieve a properly working design in terms of efficiency, performance and EMI, very careful PCB layout is mandatory, some component selections are critical and of course proper instrumentation is absolutely required.

This article has been written in order to throw some light about the internals, advantages and difficulties of this not very well-known (and even less well understood) technology. Everyone thinks that “Class-D” stands for “Digital”. I hope that after reading this article, no-one thinks that any more

---

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Sergio Sánchez Moreno and Rod Elliott, and is Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The authors grant the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Sergio Sánchez Moreno and Rod Elliott.

Page created and copyright © 04 Jun 2005

Elliott Sound Products **Valve Amplifiers - do they really sound different**

ESP Home Page

Articles Index

Page Last Updated 28 Nov 1999

Valve Amplifiers
- do they <u>really</u> sound different?

Despite a decade and a half of transistor development, many hi-fi enthusiasts have remained faithful to valve amplifiers. Now it seems amplifier manufacturers worldwide are having a second look at this once all-but-discarded technology.

This article is a partial reprint from the November 1977 edition of Australian Hi-Fi Review, with only the discussion section and review of the hybrid valve amp I designed. Please see the copyright notice at the end of this document.

Valve Amplifiers

WHEN Harold Leak introduced his company's (H.J. Leak & Co) first transistorised amplifier, the famous Leak Stereo 30, he remarked that in his opinion, eliminating the output transformer was the greatest advantage of the new model.

The Stereo 30 was a very significant development. It was the first commercial transistorised amplifier from a leading British manufacturer, previously renowned for state-of-the-art valved designs. Its introduction reflected the very sudden shift world wide from vacuum state electronics to solid-state, at a time when the thermionic valve was still capable of further development and exploitation.

It's easy to understand why amplifiers and hi-fi electronics in general became transistorised so rapidly. Transistors saved costs, and in terms of standard test procedures, measured far better. Total Harmonic Distortion (THD), for example, that still-revered but relatively unrevealing yardstick of performance for hi-fi amplifiers, was invariably lower with transistorised units. Power outputs were far

higher than from valved models of equivalent prices. Hi-fi amplifiers had now, it seemed, reached a peak of development, with minimal distortion and unlimited power availability (or so it seemed) to drive the new breed of inefficient speakers.

In the fifteen years or so since the great transistor revolution, the valve-versus transistor argument has constantly been debated. Many audiophiles regret having sold their old Quad 2's, their Fishers, their Scotts, their Leak TLl 2's, Stereo Twenty's and Sixty's and so on, having found the new transistorised designs less satisfying sonically despite vastly superior general specifications (right on! - Ed).

Until fairly recently, many of the problems associated with transistorised amplifiers remained unidentified, and there is little doubt that further problems have still to be identified and overcome. Yet there is no evidence that a transistorised amplifier (even one based on miniature integrated circuits) cannot sound at least as good as the very best valved models. Harold Leak's remark in connection with output transformers is one of the best arguments in favour of transistors, for it seems to be the output transformer more than any other component in a valved power amplifier which dictates performance quality. And good output transformers are very expensive indeed.

If we are to establish why a valved amplifier sounds better than a transistorised amplifier (and we're still far from convinced that this is so) we must look at the technique in both instances. It would appear that conventional methods of test and measurement give little real indication of how a particular component is going to sound, and this is probably because of the enormous difficulty of providing a controlled and quantified test signal simulating the sort of signal derived from musical sounds.

In a practical situation, an amplifier is faced with signals of astounding complexity, vastly different from the far simpler signals derived from test equipment.

In addition, an amplifier is also faced with a practical loudspeaker for a load, and an investigation of the behaviour of a speaker as seen by the amplifier reveals a situation which often makes the dedicated audiophile want to give up the whole idea of high quality sound reproduction and retire to a desert island! Many amplifiers particularly transistorised ones-can be excused for complaining audibly about the sort of load they are called upon to feed.

And at the amplifier front end, there is another interface problem, that of the source (invariably the pickup cartridge and its need for equalisation and good S/N ratio poses the greatest difficulties) and its effect on the preamp input performance.

There is a lot of evidence to suggest that a good transistorised preamp is superior to a good valved one, not simply in terms of measured performance but in terms of the audible result.

There is no doubt that most valved amplifiers contribute to a categorically different audible result from that of most transistorised amplifiers, in fact an experienced ear can almost invariably judge whether a valved or transistorised model is in use. This points to an influence by design and manufacturing technique on the result - which might seem obvious but which should not happen in theory.

Most high quality valved amplifiers of recent manufacture and design (within, say, the past couple of decades) use the so-called ultra-linear push-pull output stage, which inherently cancels out high-order-

harmonic distortion and has a number of other important desirable characteristics.

As its description indicates, this type of output stage employs a pair of valves in which one 'Pushes' and the other 'pulls' - a mechanical analogy being a two handled saw. Both valves operate throughout a complete cycle, swinging from positive to negative as demanded by the input signal. The use of two valves rather than one increases output amplitude to give more power than would be achieved via a single-ended output (quite feasible with valves). In fact many ultra-linear push-pull output stages will produce a useable signal with one of the valves removed, at the expense of power loss and increased distortion! (We don't recommend you try this, incidentally, if you own a valve amplifier).

Transistor Amplifiers

Most transistor output stages are rather different. Again, a pair of transistors is used, but in so-called complementary or (quasi-complementary) mode. In effect one transistor of the pair handles the nominally positive side of the signal while the other remains dormant. These roles ire reversed with the nominally negative side of the signal - the previously operating transistor handing over work to the formerly dormant transistor.

Some transistor amplifiers operate in a similar fashion to the type of valve units just described, although there are currently severe practical limitations mainly the voltage handling ability of power transistors-resulting in poor efficiency and excessive heat output.

Class A push-pull output stages are inherently distortion-cancelling and each amplifying device, be it valve or transistor, operates in a complementary electrical sense at all times. Deviations of performance from the ideal of either device are compensated by the other. But with so-called Class B stages, one transistor of a pair is constantly switching in and out of operation, handing over to the other transistor during its period of no operation. An obvious difficulty here is to achieve a smooth changeover or 'crossover' (not to be confused with loudspeaker dividing networks), and the main reason for the 'transistor sound' of most early (and unfortunately some present-day) transistorised amplifiers is/was poor crossover performance (where the switch from one transistor to another failed to give perfect signal continuity) leading to a spiky distortion most evident at low Output levels and high frequencies. Many transistor amplifiers actually had better performance at high output levels, quite the reverse of most good quality valved amps.

And then, of course, there is the output transformer. This is used in valved amplifiers to match the valves (which demand ahigh impedance load) to the loudspeaker, which is a low impedance device.

Transformers are extremely costly and in theory are responsible for a number of serious performance deficiencies including reduction of damping factor, phase shift toward the high frequency end of the spectrum, and, in a practical situation using a complex loudspeaker load and extended upper frequency response instability with consequent oscillation. No wonder Harold Leak was keen to eliminate this component!

On the other hand, the transformer serves as a 'buffer' between loudspeaker and amp proper. This can be advantageous with practical loudspeakers in avoiding the 'ricochet' effect of transistorised amplifiers

faced with back-EMF from loudspeakers resulting from unwanted diaphragm motion. This output from the loudspeaker can, in some circumstances, penetrate back to interfere with a signal passing out toward the speaker; a consequence of the inherently low impedance of transistorised amplifying circuits. Valves are inherently high impedance devices, and the effects of back-EMF from the speaker are therefore far less likely to achieve significant penetration at sufficient level to degrade performance. This must not be confused with damping factor per se which is generally rather better in transistorised amplifiers than in valved models.

A further factor in favour of transistorised amplifiers is long-term consistency of performance. Although total failure of transistorised amps seems to be more frequent than with valved units, repairs normally cost less and are quicker. Even so, overall reliability of transistorised amps seems to be better, especially in terms of performance consistency over prolonged periods.

Valves deteriorate at a fairly steady rate, and regular replacement is essential for top performance to be maintained. And, with very high quality amplifiers, this isn't always a matter just of replacing a valve - almost invariably, matched sets are required and it is also wise to optimise grid bias voltage level when valves have been replaced. Regular transistor replacement, on the other hand, is not necessary.

Generally speaking a transistor either works or it doesn't - and only requires replacing if it has an inherent fault or has been abused.

Of course, there are other fundamental differences between valved and transistorised amplifiers, and a study of these shows that a transistorised amplifier should be able to out perform any valved amp. This is certainly true of low-level amplifier stages, transistors generally having better signal-to-noise ratio, better frequency response in terms of overall bandwidth, better linearity across that bandwidth, and greater efficiency.

At the present time, hybrid valved/ transistorised amplifiers would seem to offer the greatest potential for best audible performance-transistors being used for all amplifying stages up to and including the driver stage, and a valved output stage.

In our experience, a top valved amp gives a sweeter, smoother result than a top transistorised amp, but this is a generalisation and we have heard at least one transistorised amplifier, currently being developed by Axxxxxx, a local manufacturer, which seems to have none of the audible deficiencies of either type. Nevertheless this (prototype) unit is still unhappy with some complex resistive/ reactive loads presented by multi-way loudspeakers, but the designer and development engineer are confident this problem can be overcome.

So much for technicalities. We've only scratched the technical surface here; there are far more behaviour characteristics of each type of amplifier which account for audible differences despite indications of identical expected performance from standard measurement techniques.

A--'S TRANSISTOR VALVE AMPLIFIER

Unfortunately delays in production prevented our reviewing A--'s fascinating new amplifier in this issue. This amplifier, as reported last month, is based on a valve output stage but uses transistor front-end and driver stages.

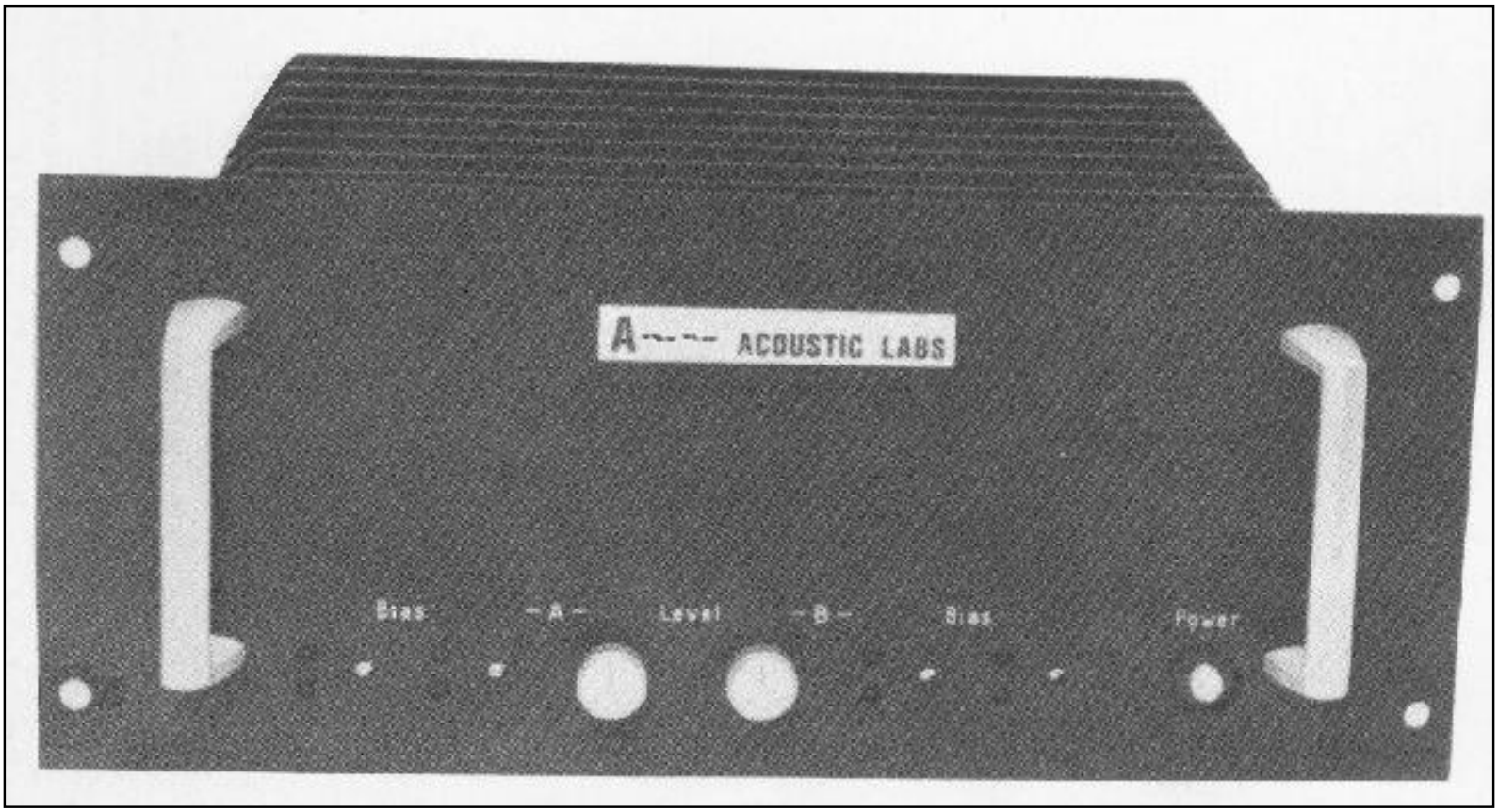


Figure 1- The Transistor / Valve Power Amp

An interesting feature of the preamp is a high-gain pickup input designed for use with low-output moving coil cartridges, plus of course the usual RIAA-compensated facility for cartridges of higher output. Following latest trends for top performance equipment, the preamp has no tone controls, and is fitted only with volume, balance, input select and tape monitor facilities.

The power amp has an ingenious system for optimising grid bias of each output valve, which contributes to minimum distortions and increased valve life. A pair of light-emitting diodes (LEDs) and a screwdriver-operated potentiometer associated with each valve is fitted to the front panel and adjusting the potentiometer for equal light intensity from both LEDs (of the pair) provides correct grid bias voltage. Level controls are fitted for each channel.

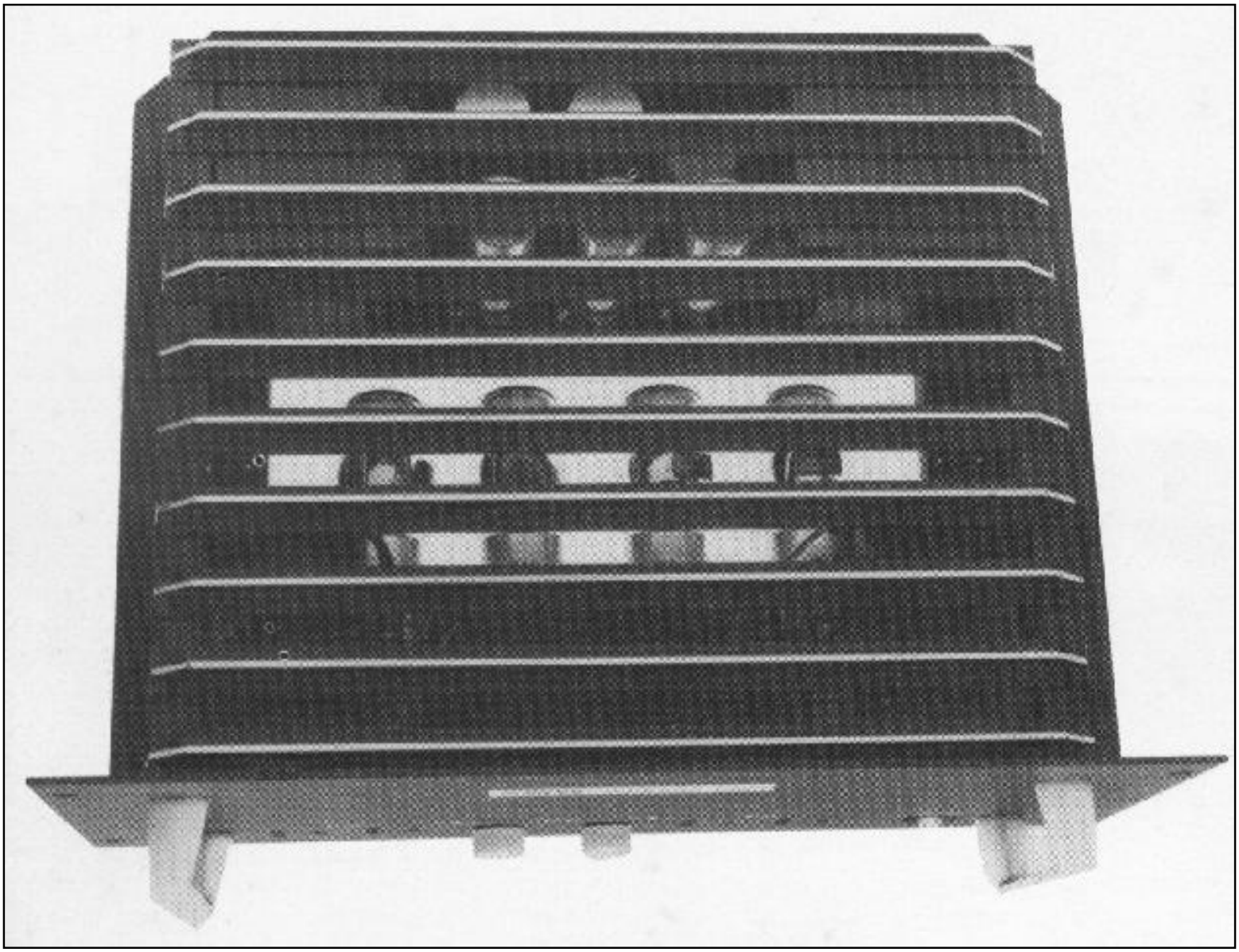


Figure 2- Top View of the Valve Power Amp

The power amp looks quite a monster and the sight of those enormous KT88 output valves, visible through the gaps in the T-section aluminium cover members is quite nostalgia-invoking.

The preamp, provided with a remote power supply which helps improve signal-to-noise ratio, has, by comparison, an unobtrusive appearance although we understand production models will be slightly different.

Power amplifier performance was excellent, low frequency sounds being particularly well defined and with good extension to extremes. There was no characteristic valve sound, the overall result being essentially neutral and with good definition of fine detail, impressive transient performance and excellent stereo imaging.

The preamp, whilst again having no specific performance deficiencies in terms of tonal balance, tended to sound a shade 'grainy' at times, especially when called upon to deliver high outputs. However this

was not unpleasant but rather gave a powerful, 'gutsy' sound which one observer described as giving the amplifier 'balls!' (blush, blush!).

Production is currently in full swing and samples of the amplifier should be available from AMW retailers during December. Rated output of the amp should be in the vicinity of 100 watts per channel, thanks to the use of very carefully wound and efficient output transformers - the prototype was claimed to yield some 60 watts per channel but sounded far more powerful than most amplifiers of this rating.

We're very much looking forward to seeing and hearing production samples of this very fine new Australian product, which upholds the firmly established AMW reputation for fine performance.

Price is expected to be around $2,300 for the preamp/power amp combination.

ESP / Rod Elliott's Notes: The electronics of the power amplifier and preamplifier referred to in this article were designed by me – I cannot take credit for the design of the power amp casing nor the specification for the output transformers, as they were the domain of my business partner of the time. The amp was designed on contract to A-- Acoustic Labs.

Also note that if some of the comments seem a little dated, this article was written in 1977, so they are indeed dated. The basic ideas are still the topic of debate, so although everything changes, it still stays the same (except the price, of course - if you could build one of these for $2,300 today you must have stolen the parts!).

Please note that I have reproduced the text verbatim - I do not necessarily agree with the comments regarding the difference between valve and transistor amplifiers, nor with all the technical matters raised.

ESP Home Page

Articles Index

Copyright Notice

The sections of the article shown are an excerpt from Australian Hi-Fi Review, November 1977

Not all of the article has been re-published here, as the remainder reviewed other amplifiers. Of the section re-published, this has been done without any alteration to the original text. Reviewers' names have been removed, as has the branding on the amplifier (the company that originally contracted the design is sadly now out of business).

Update Info: Changed page layout 28 Nov 1999

Elliott Sound Products **A-B Switch Box For Amplifier Comparisons**

Phil Allison
(Edited By Rod Elliott)

I have finally been able to add "Project X", thanks to Phil Allison. Somehow, "X" just seemed appropriate :-) Have fun.

Return to Projects Index

Return to Articles Index

Back to Home Page

Introduction

This is probably going to be a very controversial device. Its purpose is to prove people wrong and that is very confronting. If you don't wish to have your cherished beliefs about amplifiers and audio generally challenged then do not build or use this unit.

Many of you will know about the ABX system for doing audio comparisons. No doubt it is a very fine piece of design but out of reach for the average person. Some years ago I felt that a much simpler device would at least allow me to do comparisons on power amplifiers while the music played in a similar way to ABX. This device was the outcome. After using it for a few *seconds* my attitude to audio listening tests changed forever. If you are game for a challenge then try it yourself. It will cost you under $50 to build. It may be the best or worst fifty bucks you ever spent depending on your attitude.

The idea is that the listener, you, sits in the hot seat and concentrates on familiar music on your favourite speakers in your own lounge room with the ability swap power amps over without moving more than one finger. If there is even a small change in timbre or definition it should be instantly audible. The stereo image and any other factor can be checked precisely since you can sit totally still during the switchover. Well at least that was what was expected to happen.

Switch Box Design

The design is totally simple and I lashed mine up in a couple of hours and put it to use immediately. A couple of good quality relays, some hefty terminals and banana plugs and a long wire finishing in a hand held push button are the ingredients. The circuit is as shown and is self-explanatory. What is does however, is staggering.

Providing the two amps to be compared are of high quality (why would you be interested in anything else?) and of course fault free, the gains are carefully matched and have similar bandwidth the device permits instant and *seamless* switching of the amplifier outputs to the speakers at the push of the button in the listeners hand.

Because the relays switch in a couple of milliseconds there is **no** audible interruption. This surprised me at first and I tried my sine wave generator set to 100 Hz and pushed the switch. About half the time I could hear a faint “tick“ sound but this was not audible on musical program

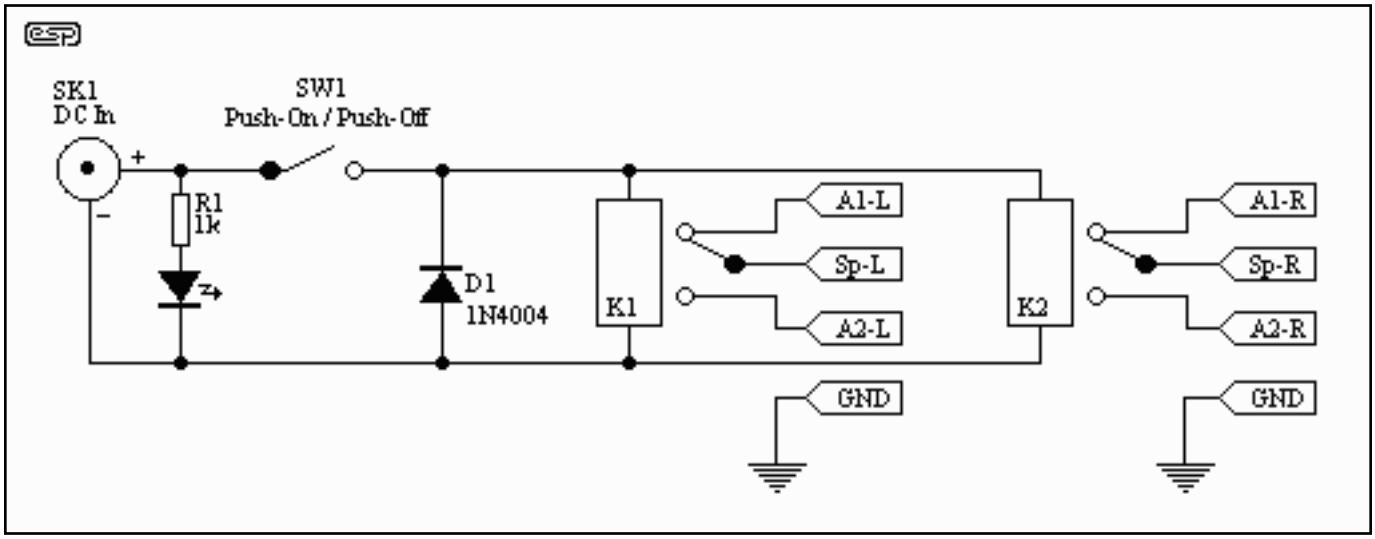


Figure 1 - The A-B Switch Box Schematic

A description of the circuit is hardly needed (but I'll give a brief one anyway). The DC voltage must be matched to the relays, and 12V is suggested as the most practical. The relays should be high quality, high current types, and DPDT relays can be used with the contacts paralleled for lowest resistance. Gold flashed contacts are desirable, but standard silver contacts should be quite adequate unless there is severe atmospheric pollution in your area.

The speaker grounds for the two amps were connected together at the box and the speaker grounds were coupled as well. Use short thick leads for connection to the amps. It is possible this might cause an earth loop hum with some combinations.

The switch (marked "Push-On / Push-Off") is the remote switch, and needs a lead that is long enough to reach the listening position. There is no real limit to the length, even with light duty figure-8 "speaker" lead, but in excess of 10 metres or so may cause some voltage loss. The LED is optional, and may be omitted. If fitted, be very careful that the LED cannot be seen from the hot seat, as it will dim slightly when the relays are energised, thus giving a visual clue - this you don't need!

Setting up the test

1. Build the box as shown in Figure 1 and place the two amplifiers close together. If one is on top of the other, then place some cardboard in between to prevent metallic contact and possible hum loops. (Make sure that you don't obstruct any air vents.)
2. Make "Y" leads to connect the left and right input signals to both amplifiers.
3. Use an audio generator or test CD to set gains to read equally +/- 1% on a digital multimeter at about 400 Hz. Leave the speakers unconnected while you do this.
4. Place the pushbutton on the “hot seat” or where you would normally listen to stereo from.
5. Check operation of the box by switching with no music and listen for the click of the relays as they changeover. There may be a faint click from the speakers too, due to small DC offsets. Trim these out if possible - the idea is to get a seamless change with no audible cues.
6. Put your favourite track on the CD player or turntable it you prefer vinyl.
7. Listen, and gently push the switch whenever you want to change amps.
8. Go and check why the box is not working.
9. If step 8 is not needed check your wiring for correct phase and levels again.
10. When you have everything right try it on your friends and family. They will likely think you have gone mad and or are playing tricks on them. The most likely comment is simply “the switch is not working”.
11. Sit quietly and contemplate what you have just found, and all its implications.
12. Don't blame me, I did warn you!!!

Notes:
When comparing amps of different power ratings, stay within the capacity of the smaller unit. Valve and transistor amps may be compared as long as the valve model has a damping factor greater than 25. Amps with subsonic filters may be distinguishable from those without on some material.

Matching the gain may be difficult if the amps do not have level controls. Solder a 10 kohm multiturn trim pot to the back of each RCA plug on the amp with more level to set the gain.

You ***must*** use a push-on, push-off hand held switch. Never use one that needs to be held

down to make the relays change over.

After the initial surprise (shock) wears off, try the old stop and restart A-B test method again. See what happens!

Editor's Notes

This is a contributed article, and the author (Phil Allison) has made this information available for the sake of audio, and not for profit or personal gain. The views expressed are not necessarily those of the editor, and no inference is made one way or another.

I do suggest that if you want to test this technique that Phil's instructions must be followed to the letter - even the smallest variation in level can invalidate any A-B test. Comparisons between valve and transistorised amplifiers are likely to show differences, only partly due to the higher output impedance of a valve amp, and extra care is needed to balance the levels with this combination.

It is also important that there are no visual cues that might alert you that one amp or the other is in operation. To be safe, place a screen of some sort between you and the amplifiers and switch box. As Phil has stated, if amplifiers are of different power ratings, make sure that neither amp clips (distorts) at any point during the tests. This will be immediately audible, and is not a valid test for an amplifier's sound quality (since it has none when clipping!)

If you don't want to be confronted by this switch box, build one anyway. It will allow you to make comparisons between amplifiers that are otherwise impossible to do with accuracy or repeatability. You might find that there are audible differences, or you might not. Either way, it gives you the ability to know (rather than assume or imagine) that one amplifier is different from another.

Return to Articles Index

Back to Home Page

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Phil Allison, and is Copyright © 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Phil Allison) and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allow that one (1) copy may be made for reference. Commercial use of this published material is prohibited without express written authorisation from Phil Allison and Rod Elliott.

Page created and copyright © 11 Aug 2000

ABX Double Blind Audio Tester

Steven Hill / Rod Elliott

Introduction

This project describes the construction of test equipment for double-blind or ABX testing of source components - preamplifiers, tuners, DACs etc. or even, if that is your particular vice, interconnects. It builds on the work done by Phil Allison described in Project X. I recommend that you read that project description before you commence this one. If you do not like what you read there, then you might as well stop reading at this point.

Double-blind and ABX tests do not allow the listener to know which component they are listening to, and furthermore don't allow the test controller to know either. This guards against visual cues to the audience (including body language).

There is information on the principles behind ABX testing elsewhere on the WWW, therefore, I intend to give only the briefest description here.

An ABX test allows the listener to select either A or B as many times as they like, and ultimately decide which of these is X, where X is randomly selected by the equipment to be either A or B, and the responses are logged for correlation when the test is complete. For example, a 1000Hz tone is assigned as A, and a 1500Hz tone as B. Random selection determines which of these is X. Listen to A then X, listen to B then X, and decide if A or B sounds like X (this would be a rather easy test for anyone to get right, of course- but if B were to be 1001Hz it may be more of a challenge).

True ABX testing is normally not easy and uses a microcontroller or a PC to interface to the switching module. This was not an option as the project is meant to be simple and inexpensive. The simple remote control (part 2) requires an operator to control switching between A and B and to whom the active channel is known. However, if you proceed to build the double-blind remote (part 3), then the unit is a true ABX comparator - you select the next test in the sequence with the 12-position rotary switch and then have the opportunity to decode if X is input A or input B. This is "double-blind" because the sequence is not known until the test has been completed.

With the basic unit, two pieces of equipment (for example, two preamplifiers) are under comparison (A and B). They are fed from the same source. The audience hears first A and then B. Thereafter, the test controller selects either A or B as X and repeats the test for up to (about) twenty iterations, changing from A to B at his or her whim, and the audience is required to write down which they think it is for each iteration. Thereafter, the results are analysed to find out whether the audience was able accurately to identify which piece of equipment was in use at each iteration of the test.

There is no requirement that A and B are selected the same number of times during any single test. In fact, a test controller once ran a whole test with A. None the members of the audience selected “A” for each iteration. I leave you to consider what that result says about people’s confidence in their ability to identify different components.

The project is in three parts of which you must build at least two. I recommend that you build only parts 1 and 2 to start with. You might find that your experience of this piece of test equipment is so infuriating that you will regret the time spent on building part 3 if you proceed with it immediately.

It is vitally important that the output of the devices under test is equal. A difference of 1dB is normally sufficient to bias the test, normally in favour of the louder channel. Therefore, you will need a test-tone CD for calibration purposes if using CD as your source or an LP with test tones if using vinyl. If comparing tuners, you will probably find that tuning in the white noise between stations is the only way you can calibrate the two pieces of equipment. You will also need a multimeter.

Unless you are using a (very expensive) true RMS digital multimeter, you should not use a test tone above 500Hz. With an analogue meter you can use a higher pitch but I do not recommend that you go higher than 1kHz.

Part 1: The controller box

The circuit diagram should be fairly easy to understand. There is nothing difficult about it, and no electronics are involved in the audio path.

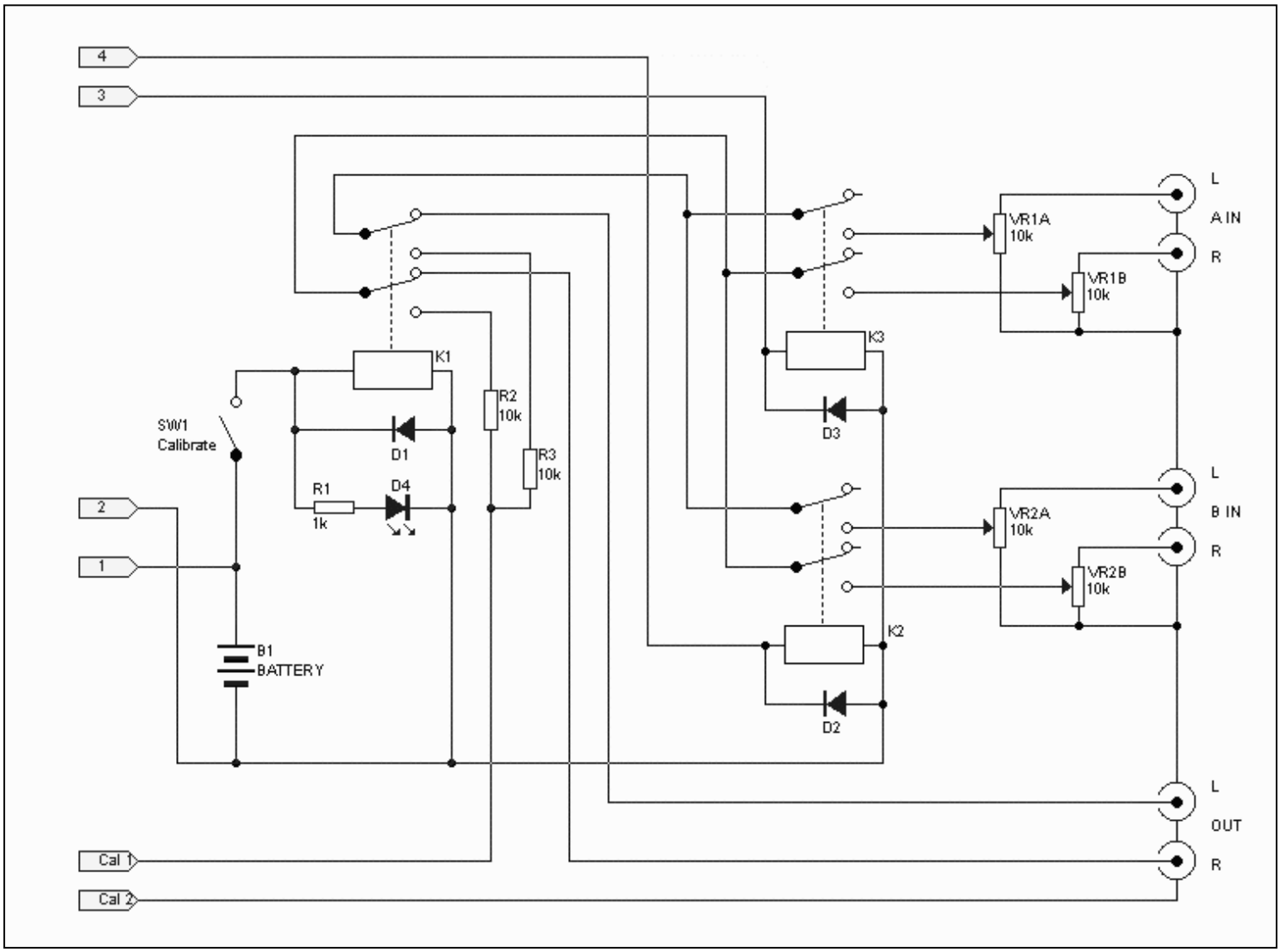


Figure 1 - The Switching Unit

Connections 1 through 4 go to either of the remotes and should be wired to a four-pin connector. The relays K1-3 can be low current types. The battery must match the voltage of the relays. I used 6V DPDT relays. K3 and K2 are used to select input A or B respectively. K1 mutes the output for calibration purposes. You could replace K1 with a switch if you wish but you then have to run more wires inside the box. The voltmeter connects externally to the two pins marked “Calibrate”. Pots are used on both inputs solely for consistency – if a pot is used on only one of the inputs, then some people may use this as “proof” that the device modifies one channel and not the other, and this can be used as an excuse as to why the results failed to correlate “correctly”.

The circuit should be built on a piece of Veroboard or similar and mounted in a shielded metal case. SW1 and D4 are mounted on the outside of the case, as are the connectors for attaching the voltmeter, the 4-pin connector for connection to the remote and the RCA plugs for the A and B inputs and the X output.

If you are proposing to use only the remote described in part 2 you do not need to take any precautions to ensure that K2 and K3 are inaudible at the listening position. However, testing using the part 3 controller requires that the relays be inaudible at the listening position. To do this you will have to mechanically decouple the circuit board from the case and also use some acoustic dampening material in the case. See note 2 below.

WARNING: I do not recommend that you omit the muting part of the circuit. When you come to use this device, you will first choose the music you wish to hear for the test and establish a realistic listening level Then you mute the output and use the 0dB level test tone to adjust the levels of both channels to the same voltage. You will probably find that the steady-state voltage is greater than 2V RMS. Unless you have monster loudspeakers, a tone at this level, fed through a power amplifier with a 30dB gain (which is common), will probably destroy your speakers and will do nothing for your hearing.

I have found that different pieces of music require different volume settings to establish a realistic listening level. An AB test where you are wincing at the volume or straining to hear the music is of no use. The calibration mute circuit has been included to speed up the recalibration between tests using different pieces of music.

That said, however, if you wish to go through the process of either (a) turning off your power amplifier(s) or (b) disconnecting your loudspeakers each time you need to recalibrate, you can omit this part of the circuit. However, you have been warned. (And you have forgotten to reconnect the speakers and/or turn on the amp(s) before proceeding with the test.)

VR1 (and optionally VR2) is included for trimming of the voltages during calibration. Obviously, the volume should be set using the preamps' volume controls (if those are what you are comparing). However, if you are testing preamps, both of which have stepped volume controls, you might find it difficult to match the voltages at your realistic listening level. That is what VR1 is for. If you don't expect to be doing this, VR1 can be omitted. For A-B testing of other equipment, such as tuners, VR1 will be required.

Part 2: The Manual Remote and its Operating Procedure

This is a simple selector which can be built in a small plastic box and is connected to the controller using four-core cable. Heavy duty cable is not required. The cable should be long enough that the operator can sit further away from the source equipment than a person in the normal listening position. SW1 is a 3-position rotary selector switch used to select A or B or null (center position). The channel selected is indicated by the relevant LED. SW2 is a DPDT toggle switch which flips the channel from A or B or vice versa. This can be used for simple A-B testing. I have found that there is an audible break when a toggle switch is used for SW2. You might find no break is audible if you use a rotary selector.

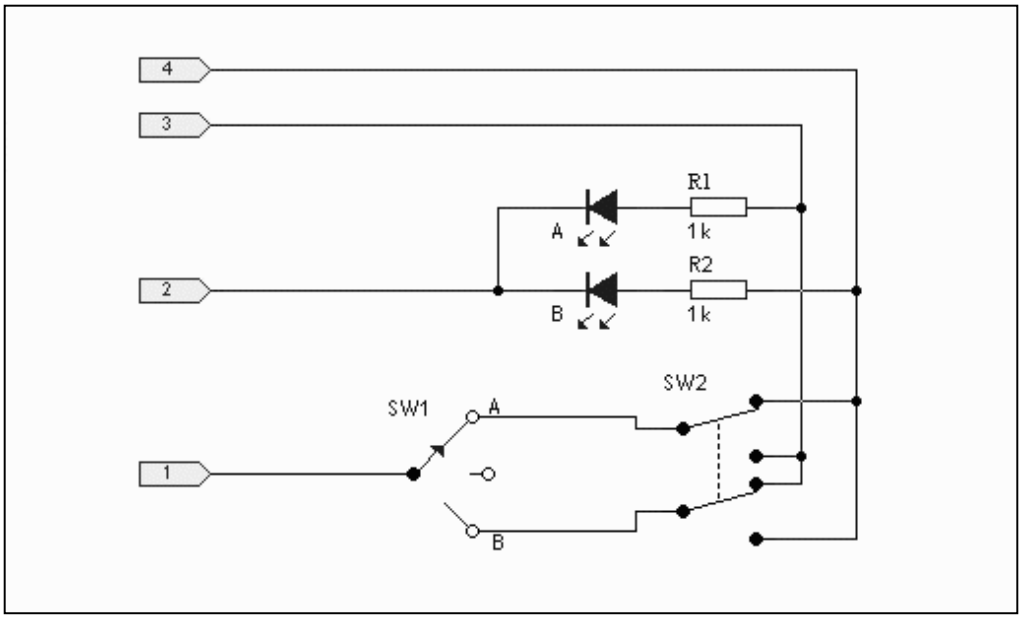


Figure 2 - The Basic Remote Schematic

To do an AB test with this remote you will need a (patient) assistant. The procedure (using as an example two preamps as the devices under evaluation) is as follows:

a) Use Y-splitters to connect the source equipment to each of the preamps. Connect the output of one preamp to input A and the other to input B on the controller box. Connect output X to your crossover or power amplifier.

b) Select one of the channels and play some music adjusting the volume on that channel to the desired realistic listening level.

c) On the controller box, mute the output using the switch provided. Connect a voltmeter to the calibration terminals. Insert the test tone CD, play a 0dB calibration tone and note the voltage. Use SW1 on the remote to change to the other channel and adjust the volume so that the same voltage is displayed. If stepped volume controls are in use, you might need to trim using VR1 on the controller. (Set VR1 to the maximum before calibrating and use it to attenuate the voltage.) Once the voltage is adjusted, remove the test-tone CD and replace the music CD and turn off the muting switch. Do not touch the volume controls again for the duration of the test.

d) Assume your normal listening position. You will need a pencil and paper to record your choices.

e) The person conducting the test (the operator) should sit out of view of the listener with the remote to hand. The listener must not be able to see the LEDs on the remote. Decide which one of you controls the source and how many iterations of the test will take place.

f) The operator selects first channel A and the listener familiarizes him/herself with the particular characteristics of the preamp connected to that channel. Channel B is then selected and familiarization of that that channel takes place.

g) The operator now selects either channel, the music is played and the listener to writes down whether A or B is his choice for X. The listener does not divulge to the operator his choice. The test is then continued for the number of iterations agreed upon, with the operator selecting either A or B. I recommend that, to avoid any unintentional bias, the sequence of changes be determined in advance of the test by some random operation (tossing a coin, throwing a die, for example). The operator should carry out this random operation prior to conducting the test.

h) The last iteration of the test having been concluded you may either check the results or return to step (b) with another piece of music.

The listener's choices are then compared with the operator's notes of the actual channel in use during each iteration. It should be immediately apparent whether any there is any correlation between the listener's choices and the actual selections.

If you do not have a patient assistant, or patience has run out, you might wish to proceed to ...

Part 3: The double-blind remote and its operating procedure

This remote uses cascaded SPDT slide switches to set up an A-B sequence which can be used blind for testing purposes and then read back to check the results. It should be apparent from the schematic how this works. The separate poles of SW1 and SW2 must each connect to lines 3 (A) and 4 (B). SW3 selects between SW1 and SW2. For each position on the rotary selector SW4, three SPDT slide switches are required. I have built this with a twelve-position selector using 36 slide switches.

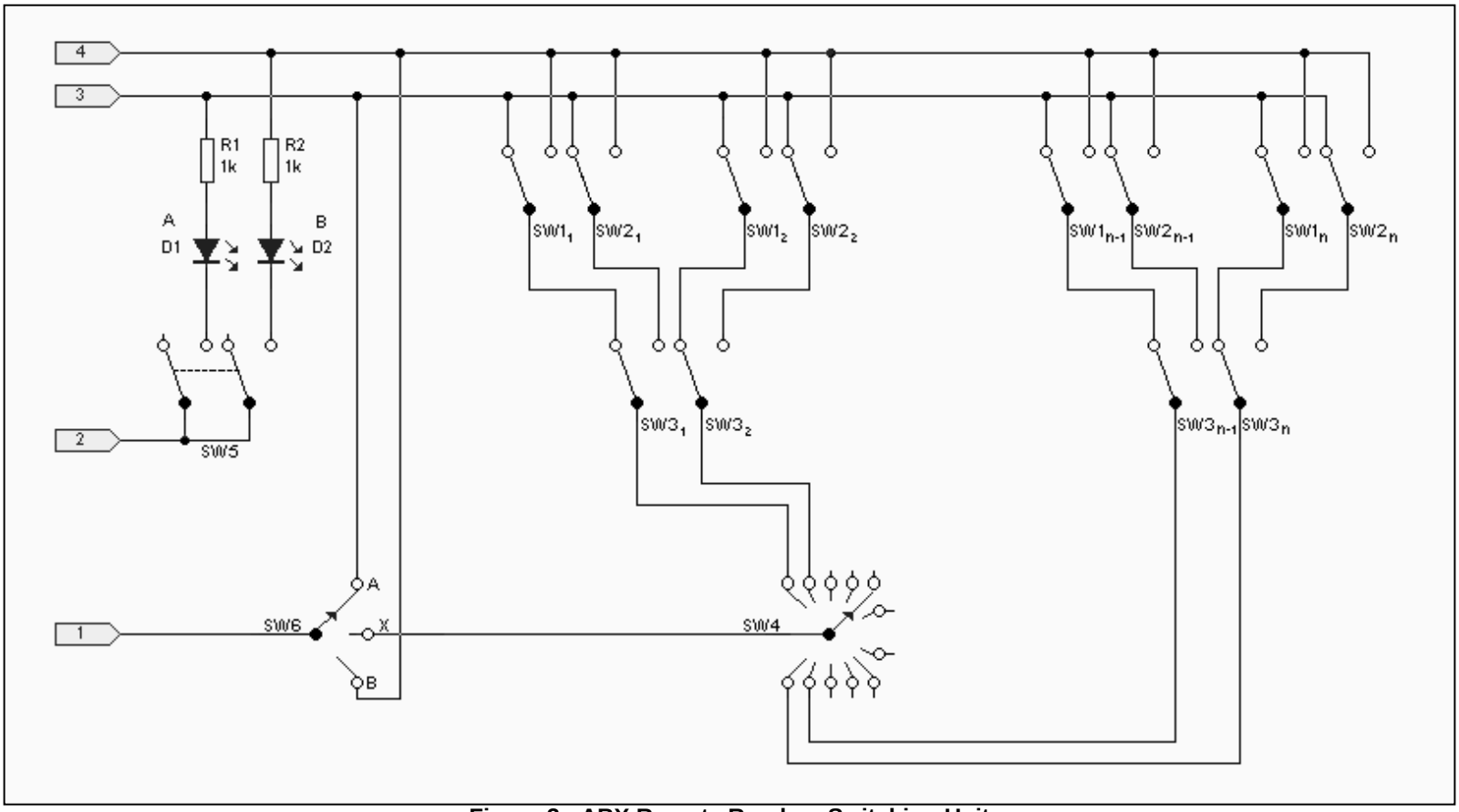


Figure 3 - ABX Remote Random Switching Unit

In the above, "n" is the number of poles on the rotary switch. If a 12 position switch is used for SW4, then "n" is 12, and "n-1" is 11.

The schematic, for clarity, shows the switches SW1, SW2 and SW3 connected together in a regular sequence. However, when building this remote, the switches are hooked up in an irregular fashion so that it is not possible from the outside of the box to identify what function any particular switch has, nor how they are wired in relation to one another. The separate poles of each switch (SW1 and SW2), however, must each connect to the A and B buses. If you do not do this you will introduce a bias in favour of one of the channels. The photo shows my own unit. Properly constructed, it should look like a mess of wires, as does mine. The orange and blue wires are the bus lines, the green wires the connections between the centre poles of the SW1s and SW2s and the brown wires go to the rotary selector.

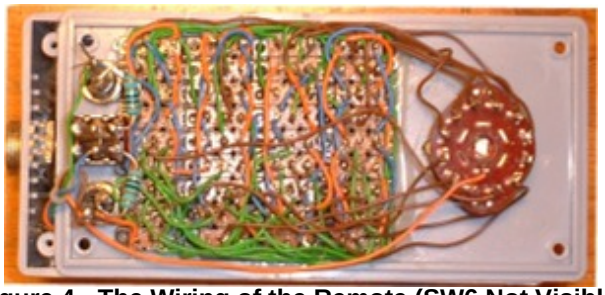


Figure 4 - The Wiring of the Remote (SW6 Not Visible)

If you decide to proceed with this part of the project, I must warn that this remote is tedious to construct and you must proceed with great care because trouble-shooting an incorrect hook-up or a dry joint can be difficult. You should first wire between the centre poles of the SW1s and SW2s to the SW3s, then run the bus wires to the outside poles of the SW1s and SW2s and finally connect from the centre pole of the SW3s to the rotary selector. Solid-core or magnet wire is recommended. The remote can be constructed in a plastic case.

So how do we use this abomination?

i) It is connected to the control box with the same four-core connector that you made up for part 2.

j) Before proceeding with the test, set up the test sequence by moving half of the total number of slide switches on the box. SW5 must be open at this time; no LEDs illuminated.

k) SW6 switches selects A or B for calibration as per (b) and (c). SW5 is closed (LEDs in circuit) during this step.

l) Open SW5 so that no LEDs are illuminated.

m) Assume your normal listening position and use SW6 to select A and B as in step (f)

n) Move SW6 to the X position. Is it A or B? Write down your choice. Proceed through all the other positions on SW4 marking down your choice each time.

o) When all positions have been sampled, close SW5, rotate the selector SW4 and observe which channel was in use for each iteration of the test. Correlate your results.

p) If you wish to repeat the test, proceed again from step (j).

Notes

1. It is probably legitimate to build part 3 with only an 8-position switch (24 slide switches) and run the test twice for 16 iterations (or a 10-position switch with 30 slide switches for 20 iterations). In this case, you would read back the results as in (o) note them down without correlating them, stop the music, perform step (j) and then proceed again from (m).

2. I mentioned in part 1 that, if you were considering building part 3 of the project you should ensure that the relays are not audible from the listening position. This is because, with an "off-the shelf" rotary switch, even thought they are nominally break-before-make, you will probably find that there is no interruption of the sound when changing positions even if there is a change of channel. If this is the case and if there is no change of channel, the relevant relay will not drop out. However, a change of channel will cause the relays to switch and you will hear it at the listening position, which is something of a give-away! If this troubles you, specify true "break-before-make" switches for SW1 and SW2 and then the relays will always click. (Actually, this is probably less troublesome than trying to insulate the controller box but has the downside that there might be an audible break in the music.)

3. You should be able to acquire all of the parts, including the cases, for less than $100. Parts 1 and 2 are simple to build. Part 3 will take a fair amount of time. However, the results of using this equipment may well astound or confound you and those to whom you demonstrate it. Did you ever wonder why your local high-end audio store does not have something so inexpensive like this so that you can compare your modest preamp, or whatever, with the mega-bucks piece of kit you have been drooling over ever since some reviewer's panegyric: "The best preamp I have ever heard for (only) the cost of a small family car."?

4. I mentioned at the beginning that you could use this to compare interconnects. How? Simply connect it in reverse: source into X, the cables under comparison to A and B then via a Y-splitter to the preamp. The presence of the Y-splitter and the controller box does not invalidate the test because the same "degradation" of the signal occurs to both cables. The author accepts no responsibility for domestic discord or other unpleasantness arising out of the use of this test equipment for such purposes.

5. ABX testing of CD players, turntables and cartridges etc. is problematic because of the difficulty of synchronizing them. However, the flip switch on the manual controller can be used to make a direct comparison between the two pieces of equipment.

I hope that this little project will amuse you. I should be happy to hear from anyone who builds it and takes it along to their "high-end" audio supplier for the purposes of comparison. Mail me on abxcomparator@hotmail.com.

Steven Hill, August 2002

Editor's Notes

Steven has done something I had thought would only really be possible using a microcontroller or a PC to achieve. The

random switching is of such complexity that it would be virtually impossible for anyone to know and remember the combinations created by each switch. Especially if assembled according to the instructions - the switches are not only capable of giving an excellent randomisation of their own accord, but if they are wired in a random fashion as you build the unit, the likelihood of being able to remember the combinations is extremely low.

Will it be worth the effort? Only you can answer that, and it depends on how serious you are about being able to tell the difference between pieces of equipment. Just like Phil's original Project "X", your first reaction may be that the unit is not working - the LED indicator that Steven included in the design to allow you to correlate the results will certainly prove that A and B are being selected, and if you are *really* unsure, you can always switch off one of the units under test.

All in all, this is an ambitious project, but one that every hi-fi reviewer should make (or have made) - I expect that if this were done, a great many of the glowing reviews we currently see would diminish. They may even vanish altogether.

Needless to say, the tester can be also used to verify that the expensive capacitors you bought really don't make any difference, or that all well constructed interconnects sound the same. This is all very confronting, but it is necessary if we are to get hi-fi back on track, and eliminate the snake oil.

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Steven Hill and/ or Rod Elliott, and is Copyright (c) 2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Steven Hill) and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allow that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Steven Hill and Rod Elliott.

Created 08 Aug 2002

The Audio Pages

Elliott Sound Products **High Fidelity AM Reception**

Approaches to Wideband and High Fidelity AM/MW Broadcast Band Reception

Copyright © 2006 - Felix Scerri VK4FUQ

Page Created 12 February 2006

Articles Index

Main Index

Contents

- 1 Introduction
- 2 Description
- 3 Construction Techniques
- 4 Conclusion

1.0 - Introduction

Although few would regard the AM/MW broadcast band as being capable of high fidelity reception, the reality is that most good quality AM broadcasters put to air quite a high quality audio signal, covering quite a wide audio bandwidth (to around 8kHz or so). The problem is that the majority of AM receivers in use today make use of the superheterodyne circuit principle (conversion to an intermediate frequency (IF) prior to detection), something that unfortunately results in very severe sideband cutting, with the result being audio bandwidth and quality comparable to an ordinary telephone channel.

Those who have heard AM under ideal conditions will know how good it can sound, with a clarity, bandwidth and subjectively superb quality that has to be heard to be believed. I have been tinkering with wideband AM receivers since my late teens, and in fact these were my first dabblings into the wonderful world of high fidelity. I have many fond

memories of laying in bed listening to the late night jazz music programs as reproduced through my home made 'crystal set' tuner feeding my old pioneer audio amp and old AR28 loudspeakers. I remember thinking, "that sounds so good". In those days we had no real FM service in our area, and AM on the medium wave broadcast band was all that was available. That was a very good incentive to see what could be done with wideband AM reception. Although I had only a very basic knowledge of radio and electronics at the time, I think I did a reasonable job with that very basic set-up. That initial passion has continued to this day! The great advantage of a simple 'crystal set' type of AM tuner, or similar is that all processing takes place at the (RF) signal frequency, and despite the limitations of this approach, the full modulation impressed on the carrier can be recovered - something not easily done with a superheterodyne receiver without much additional circuit complexity and refinement.

---

2 - Description

I mentioned the crystal set in the introduction. Apart from the charm of a receiver that requires no power source, a well designed crystal set tuner feeding an audio preamp/ power amp combination can give excellent results under ideal conditions. There is a lot more to diode envelope detection than one might think, especially if one is aiming for high quality detection. The process of diode detection is often explained in overly simplistic terms in text books, in my opinion. In reality, there are lots of factors that affect the ultimate potential quality of a simple RF diode detection system, such as the RF injection level and the loading on the diode (the so-called AC/DC ratio), which has major implications on the overall detector distortion profile, especially at high modulation percentages (very common practice these days in the broadcast industry in the endless quest for a 'loud' signal).

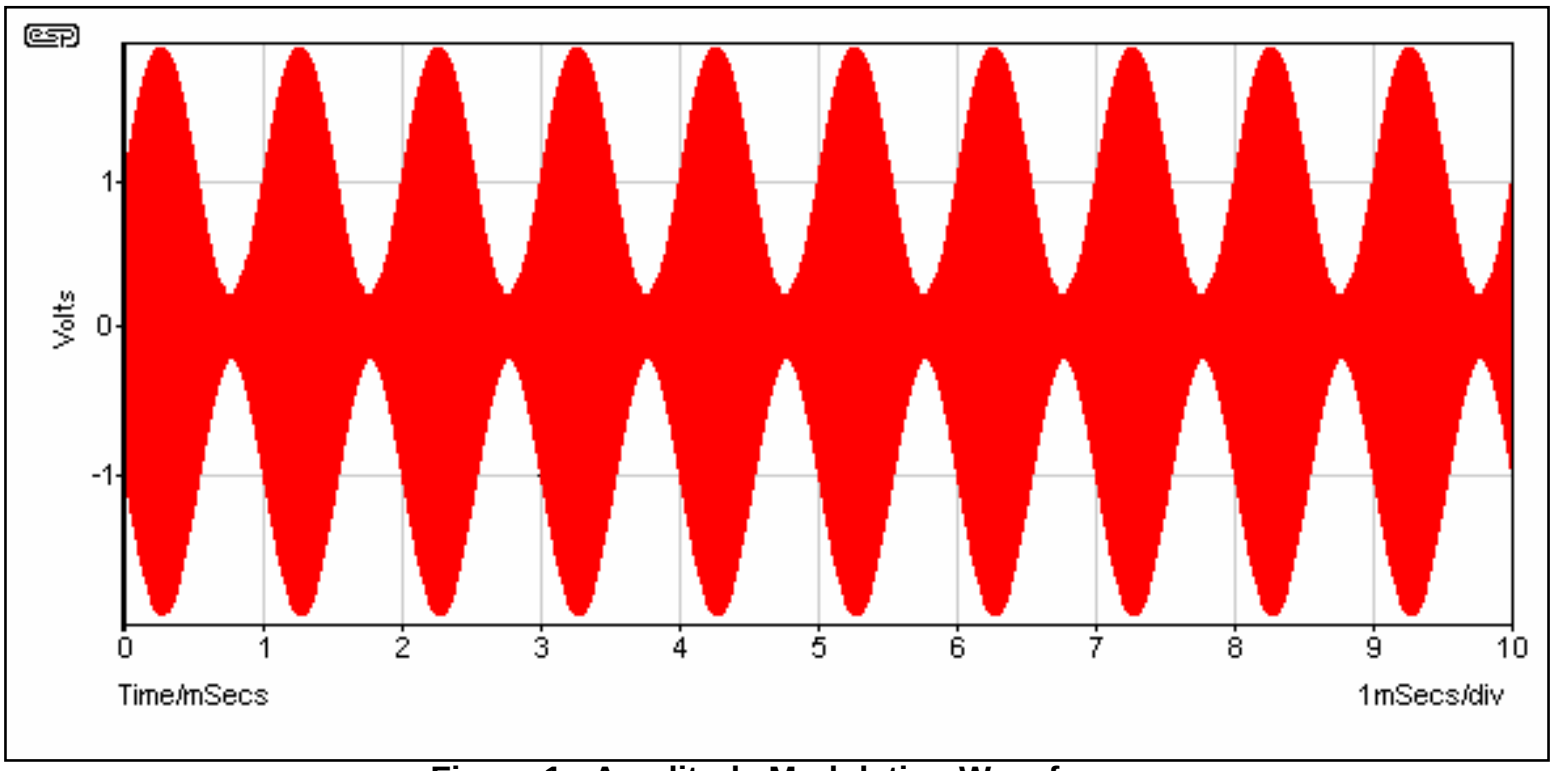


Figure 1 - Amplitude Modulation Waveform

The amplitude modulation waveform is shown in Figure 1. Maximum modulation level occurs when the RF signal falls to zero (maximum negative modulation), or is doubled (maximum positive modulation). Practical limitations mean that the minimum is usually around 5-10% of the static (unmodulated) RF power to prevent 'splatter' - a form of RF distortion that causes massive interference across the radio frequency spectrum. The maximum can be up to 150% of static power, and phase switching is often used to ensure that the highest level signals always increase transmitter power. This is possible because audio waveforms can be highly asymmetrical. The audio signal is *always* compressed and limited to ensure that over-modulation cannot occur.

Germanium diodes are considered mandatory in this service due to their low turn on voltage and consequent good sensitivity to weak RF signals. Although true germanium diodes are still being made, many electronics parts stores are now stocking germanium diode 'equivalents', diodes that are actually silicon schottky (hot carrier) diodes, fabricated to electrically resemble germanium diodes as sensitive RF detectors. They are actually pretty good. As with all silicon schottky diodes, they exhibit very little reverse leakage, very good weak signal sensitivity and low noise. They do possess a higher junction capacitance than most point contact germanium diodes, but this is not a problem in crystal set service. In a moderate to strong signal strength area, a simple diode detector of this type will work well. Combined with optimal diode loading and a selective tuned circuit front end, this kind of approach will indeed work well and provide a high quality audio program source for AM reception.

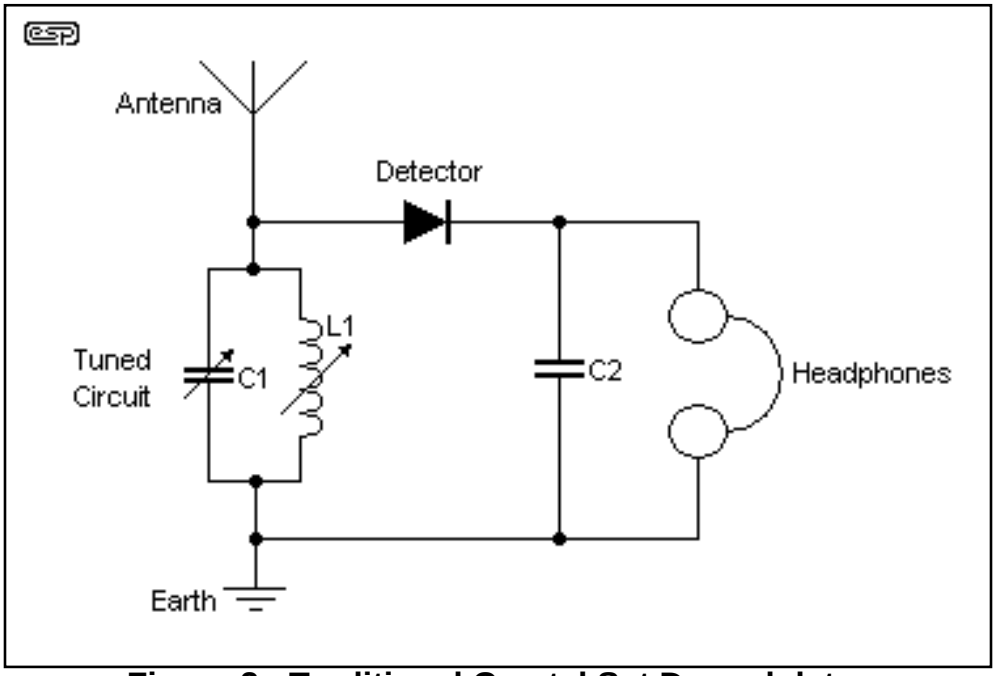


Figure 2 - Traditional Crystal Set Demodulator

Figure 2 shows the general arrangement of a 'traditional' crystal set. As shown, both the tuning coil and capacitor are variable, although many different arrangements have been used. Some sets used variable capacitance and a tapped coil, others a fixed capacitance and a variable coil. Still others used an RF transformer (so L1 has an additional winding) to create a better impedance match between the antenna (aerial) and headphones. Piezo-electric earpieces were commonly used because of high sensitivity and impedance. The audio signal can be taken to an amplifier rather than headphones.

The need for a selective front end and optimal diode loading are rather important. The relatively low impedance of the diode detector circuit will invariably heavily load its feeding tuned circuit, something which can cause problems such as station overlap and generally poor station selectivity in tuning. My preferred approach is to use double tuning, where using two separately tuned and coupled networks results in much improved 'nose and skirt' selectivity. This does complicate the design admittedly, but the overall selectivity improvement is definitely worthwhile. With single coil tuned circuit arrangements, a commonly used approach is to use coil taps in order to reduce the loading problems by connecting the diode and antenna to a point on the coil of lower impedance. This approach can work well, but it is not quite as good as two separately tuned networks in practice.

The diode loading aspect is actually a rather complex subject, but basically, the ratio of the AC and DC load on the diode output should be approximately equal for the best handling of high modulation depth with minimal detector distortion. This normally means a relatively low value diode load resistor, followed by minimal capacitive shunting as would invariably be needed for RF filtering purposes. A very high impedance for the following

audio preamp helps too. A following audio stage impedance of at least 250k is considered optimum. Actually, it is interesting to note that in these modern days of solid state technology, we are actually at some definite disadvantage as lower circuit impedances are generally the order of the day. Valve audio preamps of the past routinely possessed input impedances of the nominal required value.

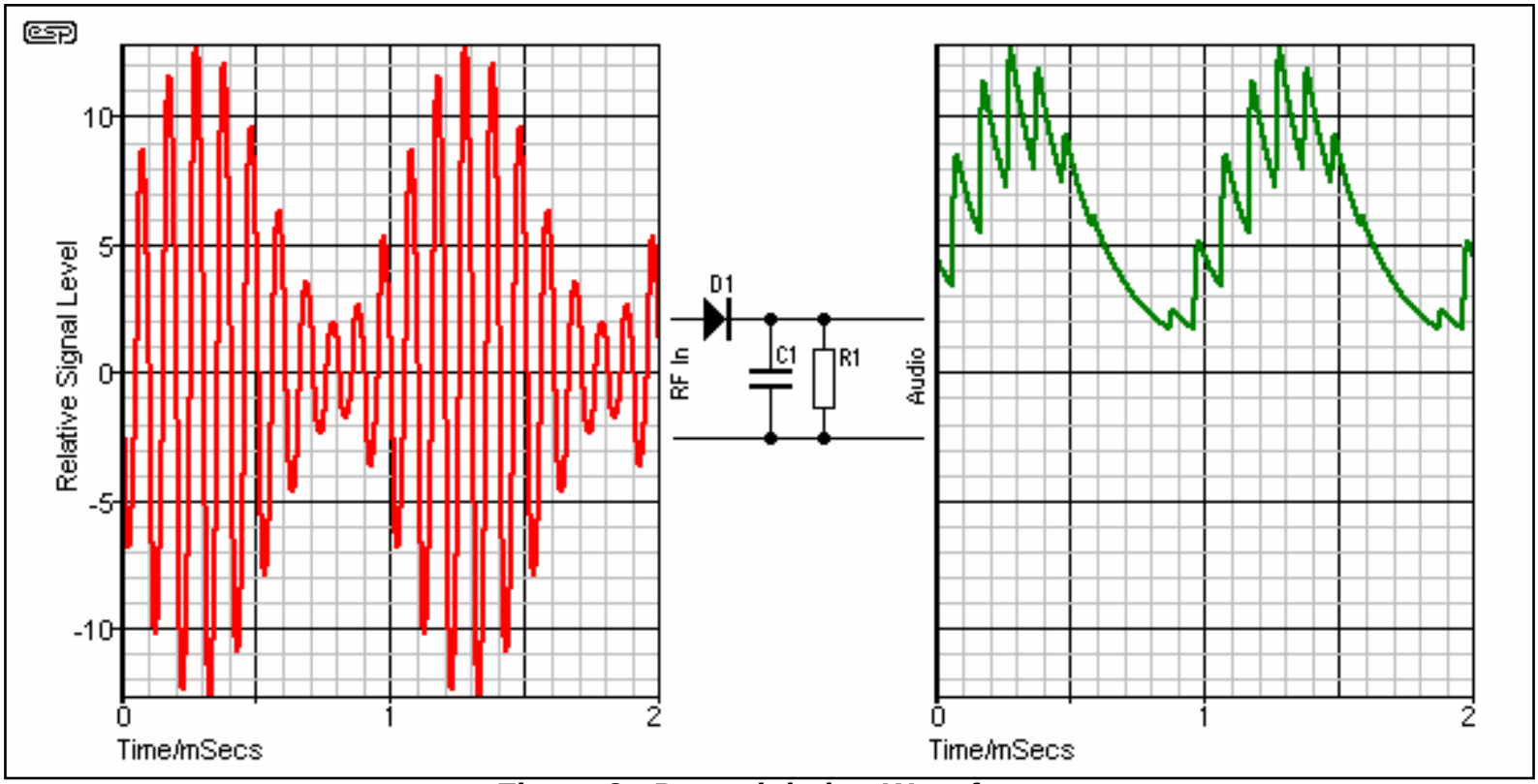


Figure 3 - Demodulation Waveform

The demodulation waveform is shown above. This assumes an ideal (or 'perfect') rectifier - i.e. a diode with no forward voltage drop, no resistance, and no junction capacitance). Because all of these parameters are included free with real components, an alternative diode demodulator is shown below. The RF waveform is shown on the left, but the carrier frequency is reduced to the minimum for clarity.

After passing through the diode, C1 integrates the signal. The cap charges and discharges slightly with each positive-going RF signal, and the remaining signal is the audio. It is always important that the capacitor value is worked out to be correct for the following load resistance (R1). If it is too large, that will create distortion - even if the diode is perfect. This is visible in Figure 3 at the negative-going audio peaks. If too small, excessive RF is applied to the following stage(s).

Note the DC offset in the demodulated audio signal. This is used in superhetrodyne (and some early TRF - tuned radio frequency) receivers to activate AGC (automatic gain control) of the RF stage(s) to ensure maximum RF level without excessive distortion.

Listening tests and evaluations show that proper diode buffering makes a very audible difference in a high quality AM tuner application. It is worth noting that diode detectors like to work with a moderate to strong level of RF input, and audible distortion does increase somewhat at lower signal levels. I have developed a novel method of using adjustable voltage bias with a 1N5711 UHF mixer hot carrier diode that does provide improved performance at lower RF signal strengths. This biased diode detector has an incredible ability to produce clean detection under even the weakest signal strength conditions by careful adjustment of the bias potentiometer! As an aside, it must be mentioned that the subject of diode detector distortion is a complex one, and much more involved than one might think. It is gratifying though, that good results can be obtained with simple circuitry.

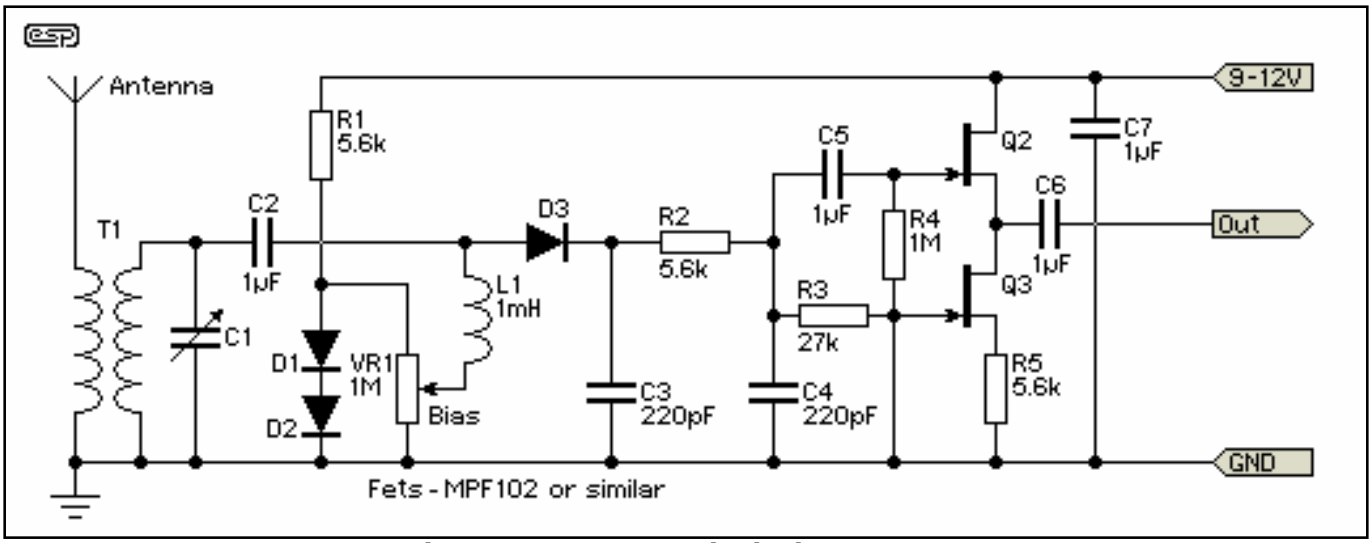


Figure 4 - Improved Diode Detector

Based on some simulations that ESP did with the circuits, a biased detector can reduce distortion from around 6% to 1% with a strong signal, with potentially greater improvements at lower levels. As noted though, this is a complex issue, and not one that lends itself well to simulation - unless the simulator is optimised for RF analysis. Most are not.

However one doesn't have to use a diode type of AM detector. There are other ways to achieve high quality AM detection. I have investigated a number of bipolar transistor and field effect transistor circuits that offer excellent general performance and other advantages as well. The 'infinite impedance' detector, using a field effect transistor is a personal favourite. It is based on a very old triode valve circuit from the early days of 'wireless'. FETs (Field Effect Transistors) are remarkable devices, especially in RF applications, and their close electrical similarity to valves is a very useful characteristic indeed. A FET version of the infinite impedance detector is easily made, and offers very high quality AM detector performance. A particular advantage results from the very high gate impedance of the FET into which the tuned circuit is fed. This effectively eliminates

the selectivity problems caused by diode detectors. I have used single tuned arrangements combined with infinite impedance detectors with excellent results in terms of good selectivity. Such is the advantage of a very high input impedance.

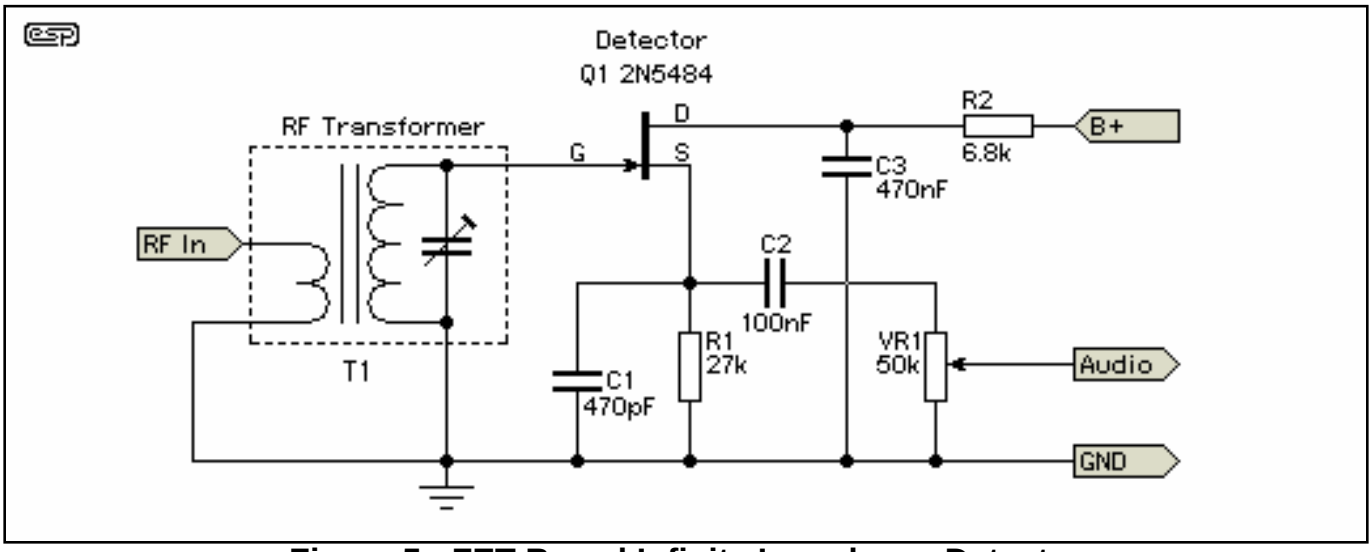


Figure 5 - FET Based Infinite Impedance Detector

A basic 'infinite impedance' detector is shown in Figure 5. While this will work very well, there are some refinements that improve performance. In my set-up, I make use of a high input impedance FET source follower modified with another FET constant current source for improved linearity and lower distortion. The second stage acts as a buffer between the detector and audio preamp input.

The version of the infinite impedance detector I have developed was adapted from the generic circuit that has appeared for many years in the ARRL Handbook, a long available and well regarded text, well known to Ham Radio operators all over the world. The circuit I've developed is more suited to tuner applications, and differs from the basic generic circuit in that the value of the source resistor is much higher, which overcomes a slight tendency to intermittent oscillation, has improved RF filtering, and includes a modified FET source follower output stage for improved audio drive into a following audio preamp stage.

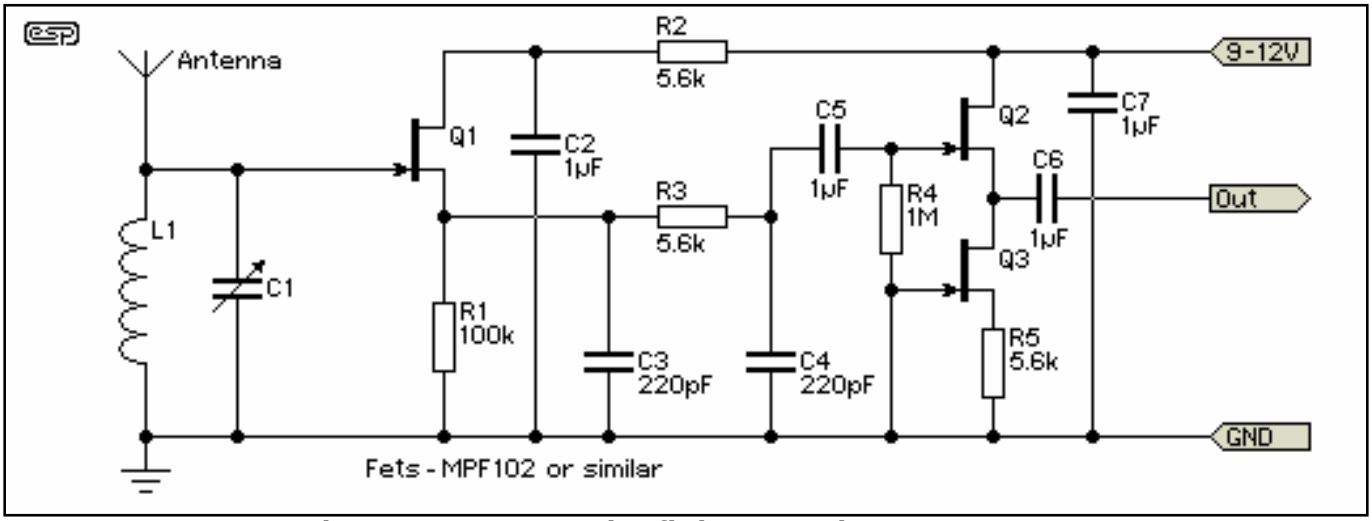


Figure 6 - Improved Infinite Impedance Detector

All round, the field effect transistor based infinite impedance detector is pretty amazing, offering high audio output, low distortion, excellent sensitivity and is quite free of the weak signal diode detector distortion tendency. It offers impressive performance for such a simple circuit. Just connect a tuned circuit and audio system, and one has quite a high quality AM program source.

All of these AM detector circuits offer potentially excellent recovered audio quaity, limited only by the quality of the original audio modulation. The infinite impedance detector offers the overall advantage of simple 'no fuss or adjustment needed' apart from tuning, but the biased diode detector has the advantage, by virtue of precise bias adjustment, of coping with any signal strength situation (strong or weak). This does require that the bias be individually optimised for each station received which can be a trifle annoying! However the audio quality under ideal bias adjustment has to be heard to be believed! Also as discussed earlier, due to selectivity considerations, a more complex "double tuned" input is considered desirable, however another option is to add a FET stage in front of the diode detector in order to relax potential selectivity concerns (as discussed previously). So many possibilities!

---

3 - Construction Techniques

Welcome to the wonderful world of building things at work at radio frequencies (RF)! Those who've only dabbled with audio gear will be slightly shocked by the way things are done at RF. It's basically a different world. At times, I've even wondered if Ohms Law applies at RF! At broadcast band medium wave RF, things aren't too critical, but it still pays to do things the right way. At RF, the effects of stray and 'incidental' inductance and capacitance become an important factor. Even a short piece of wire represents a substantial amount of inductive reactance at RF.

XL = 2 * π * F * L... and ...
XC = 1 / 2 * π * F * C

Where XL = inductive reactance, XC = capacitive reactance, F = Frequency, L = Inductance and C = capacitance

Things like excessively long leads are a no-no at radio frequencies - you will get around 5-6nH of inductance for each 10mm (1cm) of straight wire, depending on lead spacing and many other factors. My favourite method of construction is using a piece of copper clad circuit board material and wiring all components 'point to point' with all earth (ground) connections going directly to the copper surface of the circuit board material which provides a low impedance ground plane. Veroboard or similar is also quite ok at lower radio frequencies if used appropriately.

Pre-wound coils and ferrite rods can be purchased from the usual parts shops, along with suitable variable capacitors with a maximum capacitance of about 260 pf. The two outer leads of the variable capacitor need to be commoned together to obtain maximum capacitance. I prefer to wind my own coils in the interest of optimising coil Q (or coil quality). This is a very complex subject in itself, and many differing winding approaches are used, including the use of so-called litz wire. I use ordinary enamelled copper winding wire of about 0.315 mm thickness, along with spaced turns, where a slight gap (about a wire thickness) is placed between each turn. This helps with coil Q. A drop of super glue adhesive at the start and end of a winding is perfect for stopping a coil from unwinding itself. For the MW band, about 50 turns on a ferrite rod will be needed, however the exact number of turns may vary somewhat depending on local conditions and the desired tuning range.

For simple crystal set type of AM tuners, unless one lives close to a local transmitter, some kind of external antenna and earth connection will likely be required (just like in the old days). Remember that a simple crystal set tuner is a purely passive detector arrangement without any additional active front end RF gain, and the diode detector needs all the help it can get, hence the need for an external antenna. Depending on the method of connection to the tuned circuit, the antenna can affect the tuning range of the tuned circuit. With simple receivers like these, external factors can have a big impact. Remember too that the audio output level is entirely dependent on the incoming RF input level. A lot of empirical adjustment may be needed, but that's half the fun. An earth connection is also needed, but sometimes existing earthing through the audio gear is sufficient. I have an earth rod driven into the ground outside my window in the garden outside with a short connecting lead.

4 - Conclusion

Individual construction arrangement techniques are left up to the constructor, but remember the general guidelines when working with RF circuitry and also try to leave a little space around the coil(s) from any metal (enclosure etc) in the interests of not degrading the coil Q. Very importantly, when working with any type of external antenna, safety is absolutely paramount! I spent nearly four months in hospital after a serious accident when I fell from a roof, and I live with permanent spinal cord injury as a result. Don't repeat my mistake!

Despite the limitations of this kind of approach to wideband AM reception, I personally have found this a very satisfying way to listen to local AM stations with the sort of audio quality that no ordinary receiver can match. In fact, one may realise that your local AM radio station may have some transmission problem such as distortion or other defect. Such is the potential resolution of this kind of AM tuner. It is rather sad that many transmission defects will go completely unnoticed when listening on an average receiver with a poor audio bandwidth and lots of inherent circuit noise and distortion but are immediately noticed when listening on a high resolution system such as these AM detectors provide! In any case, once you listen to true wideband AM audio you will never think of AM in the same old way ever again!

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of the author (Felix Scerri) and editor (Rod Elliott), and is Copyright © 2006. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Felix Scerri) and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Felix Scerri and Rod Elliott.

Page created and copyright © 13 Feb 2006

The Audio Pages

Elliott Sound Products **Capacitor Characteristics**

Copyright © 2005 - Rod Elliott (ESP)
Page Created 24 September 2005

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Dielectric Loss
  - 2.1 Dielectric Absorption
  - 2.2 Electrolytics
  - 2.3 Distortion Tests
  - 2.4 Capacitors in General
- 3.0 Test Methodology
- 4.0 Parasitic Inductance in Bypass Applications
  - 4.1 Electrolytic Series Resonance
- 5.0 ESL and ESR
- 6.0 Food For Thought
- 7.0 Conclusion
- References

1.0 - Introduction

A great deal has been made of the 'sound' of capacitors. Dielectric losses (dissipation factor, dielectric absorption) feature heavily, with some fairly outrageous claims made as to the importance of these losses in amplifiers and other audio equipment. There are sites on the Net showing that different caps have different properties, and this is often used a 'proof' by

many people that the differences are audible. There are sites that seem to have impeccable credentials, but have managed to create nothing but FUD (fear, uncertainty & doubt) with wild claims of irreparable damage to the signal by using the 'wrong' kind of cap ... even as a supply bypass (yes, it's true - this claim *has* been made). In some cases you will read things like "listening tests have indicated ... (blah, blah, blah)". But where is the data? Who conducted the test? How was it conducted? Was the test ever *really* conducted at all? Most claims of this nature indicate that there is a hidden agenda, so beware.

What is often missed completely, is that capacitors used for signal coupling must have a very low impedance for all frequencies that one expects to pass through the system, and in general, the impedance (capacitive reactance) should normally be less than half the circuit impedance - for the *lowest frequency of interest*. For example, a coupling cap that is used at the input an audio amplifier may have a value of 1uF, with a following resistive load of 22k (this is fairly common in ESP designs).

The capacitor has a reactance of 7.9k at 20Hz, and 22k at 7.2Hz (this is the -3dB frequency). At this frequency, if 1V is applied to the input, 707mV will be 'lost' across the cap, and the amplifier will get an input signal of 707mV. The reason for the voltages not being 50% of the input voltage is due to phase. This is quite normal, and causes no problems. A double blind test of any two capacitors of the same value and reasonable construction will not reveal any audible difference - even if the music has significant very low frequency content, and the loudspeakers can reproduce it. At 40Hz, the capacitor has a reactance of just under 4k, and at 1kHz this has fallen to 159 ohms. These figures apply reasonably accurately at all voltages, impedances and frequencies.

Dielectric losses (dielectric absorption and dissipation factor are lumped together for my analysis) are blamed for 'smeared' high frequencies, thus implying that as the frequency increases, the problem gets worse. However, as the frequency increases, the amount of signal across the cap falls, so at the highest frequencies the capacitor is effectively almost a short circuit. The influence of *any* coupling capacitor diminishes as frequency increases, and is most significant at the lowest frequency of interest.

These effects are examined by a combination of simulation and actual testing - and to alleviate any concerns, no components were harmed in the production of this article (sorry). Simulation features heavily here, simply because most of the effects are extremely difficult (some are almost impossible) to measure. The resolution of the simulator is far greater than any known test instrument, but one has to be careful to ensure the models used act in the same way as real components.

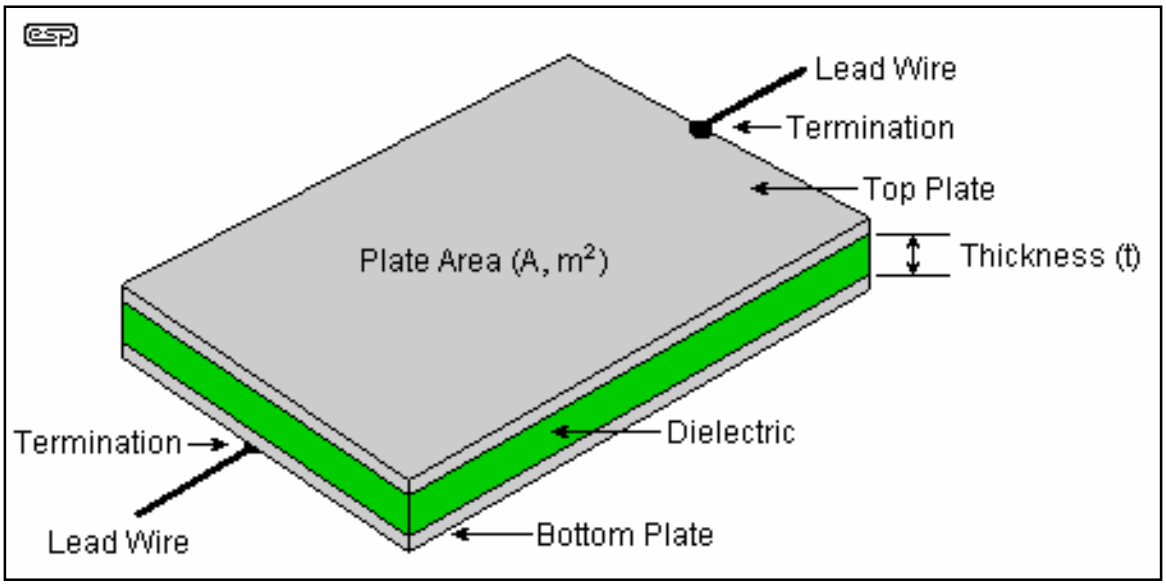


Figure 1A - Basic Capacitor Construction

Figure 1A shows the general form of construction for a capacitor. The plates shown may be metal foil, or more commonly for most caps, a metallised film. This is very thin and typically long and narrow, then it is rolled up and encapsulated. In some cases, the cap is made flat, with interleaved plates and dielectric. This allows the maximum capacitance for a given volume.

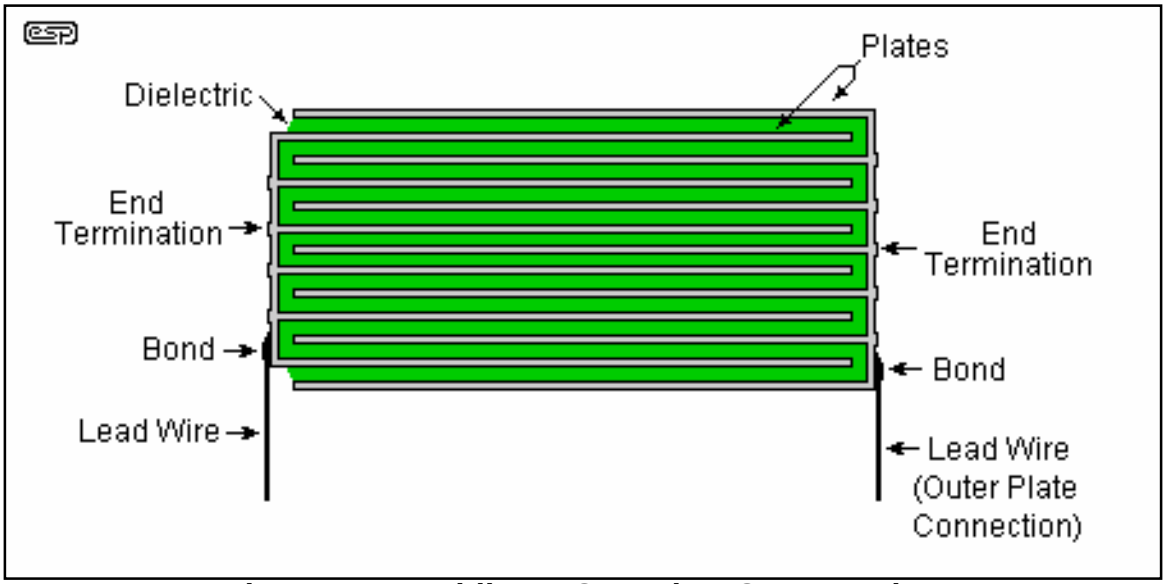


Figure 1B - Multilayer Capacitor Construction

Figure 1B shows the general construction of a multilayer cap, and is also representative of the cross-section of a traditional wound capacitor. With some capacitors, one end is marked with a band or is otherwise indicated as the outer foil. This can be useful for sensitive circuits, where the outer foil (or plate) end may be connected to earth (ground/ chassis) to shield the capacitor against interference. This is usually only ever needed in very high impedance circuits, or where there is considerable external noise.

Note the way that the ends of the foil are joined. This prevents the signal from having to traverse the length of the plates. Because one edge is joined in a 'mass termination', only the width of the plates (i.e. between the terminations, plus lead length) is significant for inductance.

The capacitance of a pair of plates is determined by the formula ...

> $C = 8.85 \times 10^{-12} kA / t$ where C = capacitance (Farads), k = dielectric constant, A = area (m^2) and t = dielectric thickness (m)

So, for example, a pair of plates of 0.01m^2 area, separated by 1um, and having an insulation with a dielectric constant of 3 (e.g. polyester), will have a capacitance of about 260nF. These plates might typically be a metallised layer of 10mm width, and having a length of 1m [1]. While this is probably not very useful, it may come in handy one day (or perhaps not). The dielectric thickness is mainly determined by the voltage rating.

Typical values for k (dielectric constant) are as follows ...

Material	**k**
Vacuum (reference)	1.00000
Air (Sea Level)	1.00059
Aluminium Oxide	7 - 12
Ceramic	5 - 6,000
Mica	3 - 6
Mylar	3.1
Polycarbonate	2.9 - 3.0
Polyethelyne	2.25
Polyester	2.8 - 4.5
Polypropylene	1.5
Polystyrene	2.4 - 2.6
Teflon	2.0
Snake Oil	Unknown

Table 1 - Dielectric Constants

This is just a small sample - see references for more. Only a few of the vast number of dielectrics available are useful, and only some of these are listed above. Of the many sites that give this information, there is considerable variation for many materials - this is to be

expected because of the range of different material formulations, even within the same chemical compound group. Snake oil has been included, but there is no actual data associated with it. Yes, this is in jest, but as you may discover, there is a great deal of snake oil used in the audiophile capacitor industry.

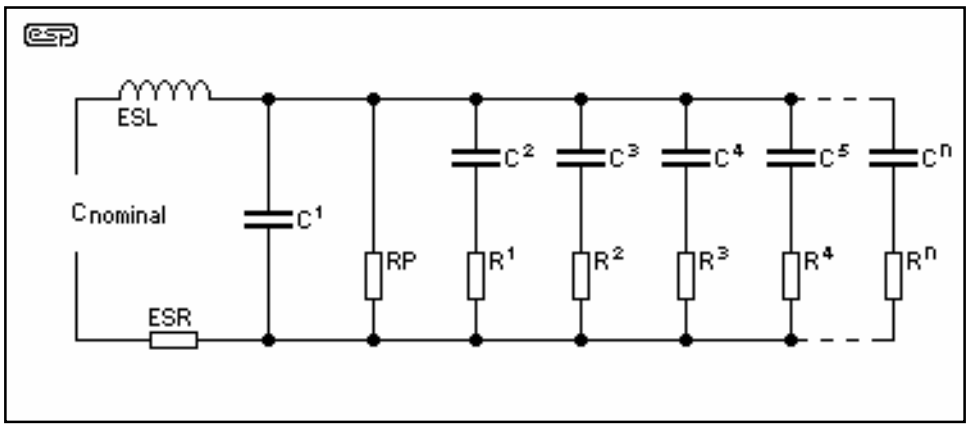


Figure 2 - Capacitor Equivalent Circuit

The generalised equivalent circuit of a capacitor is shown in Figure 2. The nominal capacitance is the value of C^1, with ESR and ESL (equivalent series resistance and inductance) in series. The parasitic capacitances (C^2 - C^n) and their series resistances represent the dielectric loss (resistance) and dielectric absorption (capacitance). These are infinite, with ever diminishing capacitance and series resistance. Figure 5 shows values used for simulation purposes.

It is important to understand the equivalent circuit of any component, because this allows you to simulate or measure the effects with the 'flaws' greatly accentuated. In many cases, it is not necessary to do either, since the effects will be quite obvious once seen for what they are.

2.0 - Dielectric Loss

The first thing to understand about dielectric loss, residual charge, series resistance and inductance, and all the other ills that afflict capacitors, is that they are quite normal, and appear in all real-world components. What is at issue is whether these cause a problem for normal audio signals at normal levels. There is no point testing capacitors by placing a 70V AC signal across them if this will never happen in the circuit being investigated. There is even less point doing this with capacitors that are rated at 50V ... DC!

While coupling capacitors are a primary target of the upgrade brigade, these are the most benign because of the very low voltages across them. Capacitors used in filter circuits are deliberately selected so that they cause the signal to roll off at the selected frequency, and this will be examined later in the article.

First, let's look at the voltage across a 1uF coupling cap connected to a 22k input impedance amplifier. At 40Hz, this is only 180mV for a 1V input, and by the time we get to 10kHz the voltage across the cap is down to less than 700uV. This is shown in Figure 2, and Figure 3 shows the circuit used for the measurement.

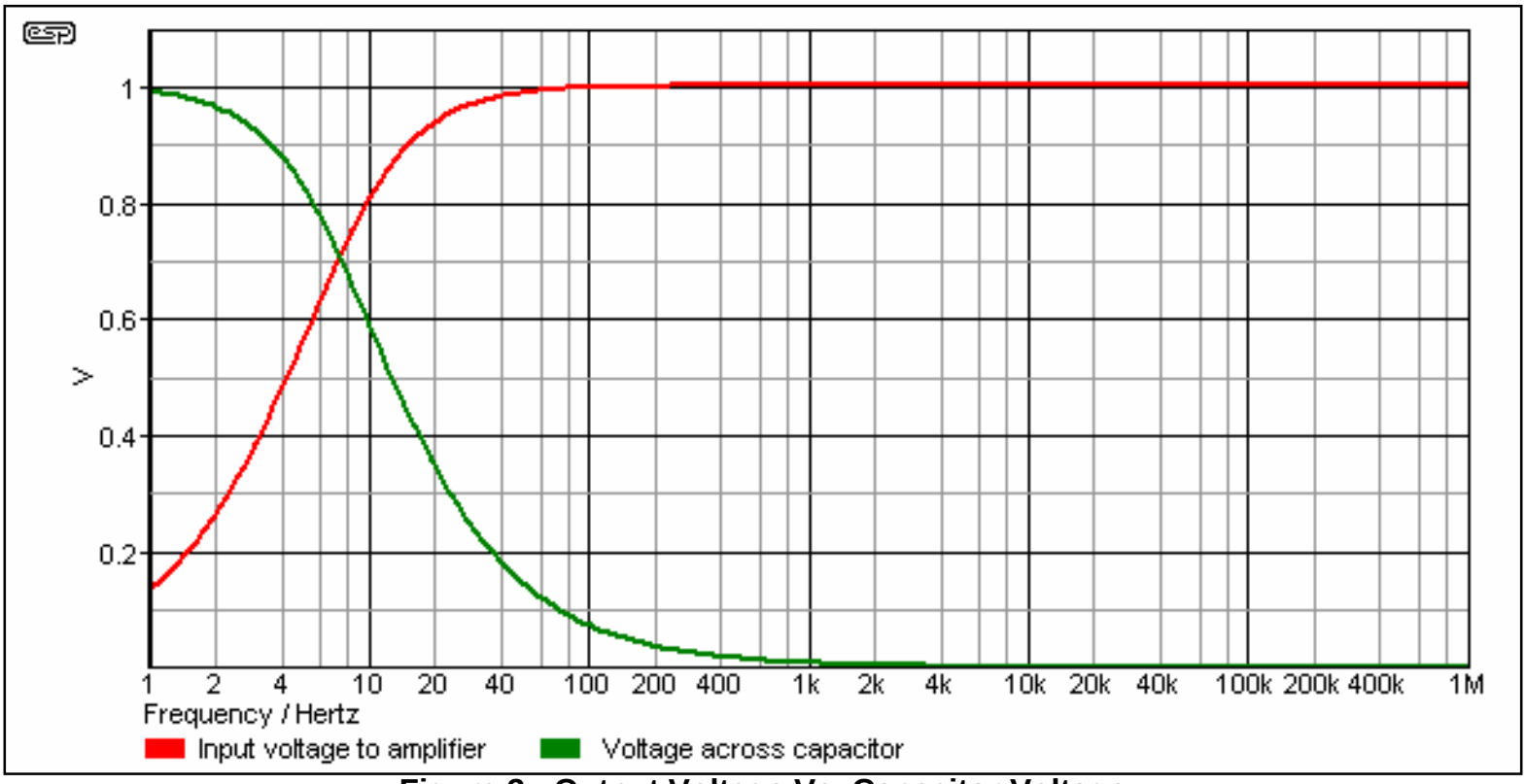


Figure 3 - Output Voltage Vs. Capacitor Voltage

The test circuit is shown below. It is simply a matter of measuring the voltage across the capacitor and the resistor. With a 1V RMS applied signal, each will measure 0.707V when the capacitive reactance is equal to the resistance. This is the low frequency -3dB frequency.

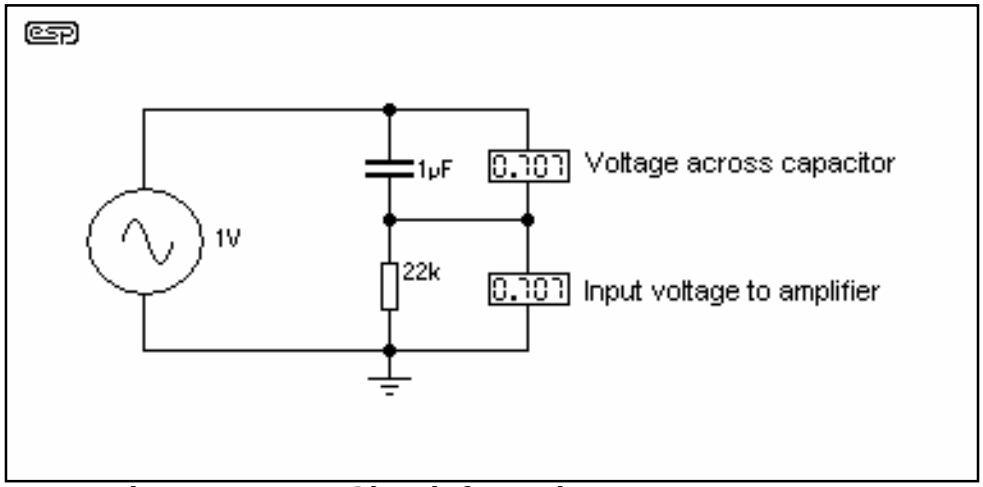


Figure 4 - Test Circuit for Voltage Measurements

Now, the caps used in a simulator are 'ideal' in that they do not have dielectric loss, series

resistance, insulation resistance (leakage) or any other undesirable parameters of a real capacitor. A simulated cap with these real parameters included is shown in Figure 5. The ESR (equivalent series resistance) is much higher than an actual cap, ESL (equivalent series inductance) is about typical, leakage resistance (via the parallel resistor) is much lower than reality at 100Megohms, and the dielectric loss components (the string of smaller caps with series resistance) deliberately exceeds that of most normal capacitors. This sub-circuit behaves like a capacitor with quite high losses - certainly it would be completely unacceptable as a tuning cap in a circuit operating at very high frequencies. In short, this is a dreadful capacitor. Perhaps these shortcomings might make it 'sound better', but it would need to be very expensive and perhaps unreliable to gain true acceptance. Just do a Web search for 'Black Beauty' capacitors - these are notoriously unreliable (especially early 'NOS'), sometimes unbelievably over-priced and should be avoided for anything more technologically advanced than land-fill (and yes, I do have personal experience with them).

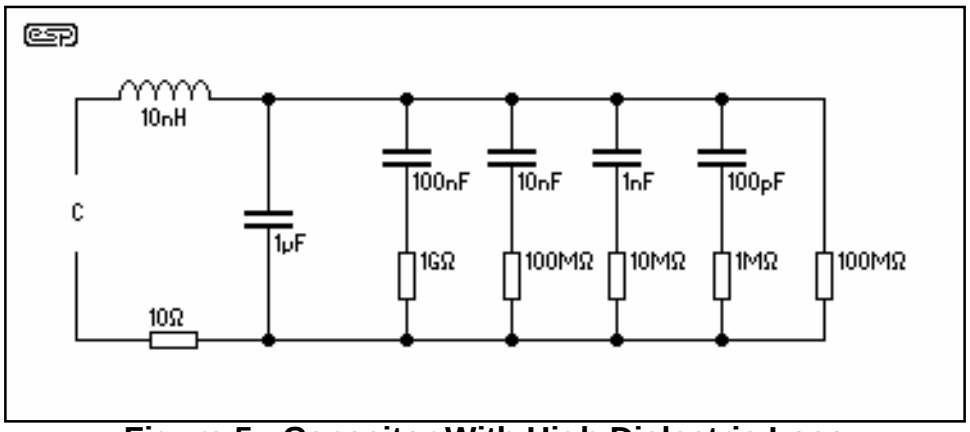


Figure 5 - Capacitor With High Dielectric Loss

This lossy capacitor (which is worse than any typical real-world component) is next used in the same circuit shown in Figure 4. When we look at the amplifier signal (across the 22k resistor) the frequency shifts up by 11mHz (milliHertz) and there is a loss of 3.3mdB (yes, milli decibel) at 10Hz, with a loss of 4mdB right through the audio band and up to 1MHz. This can be considered utterly insignificant. The vast majority of all loss is caused by the series resistance (which is exaggerated for clarity). Lest anyone think that the dielectric loss or leakage resistance may cause a phase variation, that too is insignificant. The phase angle at 10Hz is just under 36°, with the lossy capacitor being 0.047° different. Again, at higher frequencies there is no significant difference.

Ok, so there is very little change in overall performance when the lossy cap is used for coupling, but the losses should really mess up a filter circuit, right? Wrong, actually. There is virtually no difference at all. Although the difference can be seen with the simulator, most affordable real instruments don't have sufficient resolution to be able to see it. The difference between a 24dB/octave crossover filter built using ideal and lossy capacitors is so small as to be insignificant. The frequency changes by 1Hz, and the voltage difference at the crossover frequency is 0.044dB (44mdB). Many times this variation will result from normal

component tolerances, and even stray capacitance on the PCB itself could easily exceed the variation seen by the simulator. There is little to be gained by showing graphs with perfectly overlaid curves, but should anyone want to do their own simulations there is more than enough information here to allow that.

It is important to understand that the lossy capacitor appears (electrically) as an infinite number of small capacitors, each with its own series resistance. This can be built using real capacitors, with a lumped parasitic capacitance of perhaps one tenth of the value of the actual capacitance. Use a 1 megohm resistor in series with the 'parasitic' cap, using the general scheme shown in Figures 2 and 5. The 'losses' in this capacitor are far greater than any metallised film cap, yet using it in a circuit will not degrade the performance one iota. Dielectric absorption simply does not affect the way a capacitor passes the signal. Dielectric loss becomes a problem when significant (high frequency signal) voltage appears across the capacitor, but is rarely even measurable as a loss at audio frequencies and at levels typical of audio systems.

2.1 - Dielectric Absorption

So, we can conclude from this that the dielectric losses do *not* cause massive variations - in fact the variations are infinitesimal. But ... what of the charge storage of the dielectric? This is the phenomenon that allows a cap to recover some of its original charge due to 'dielectric absorption' (also known as 'soakage' [6]). This is part of exactly the same phenomenon that creates capacitor 'losses'. The lossy cap shown above has that effect too, and this is shown in Figure 6.

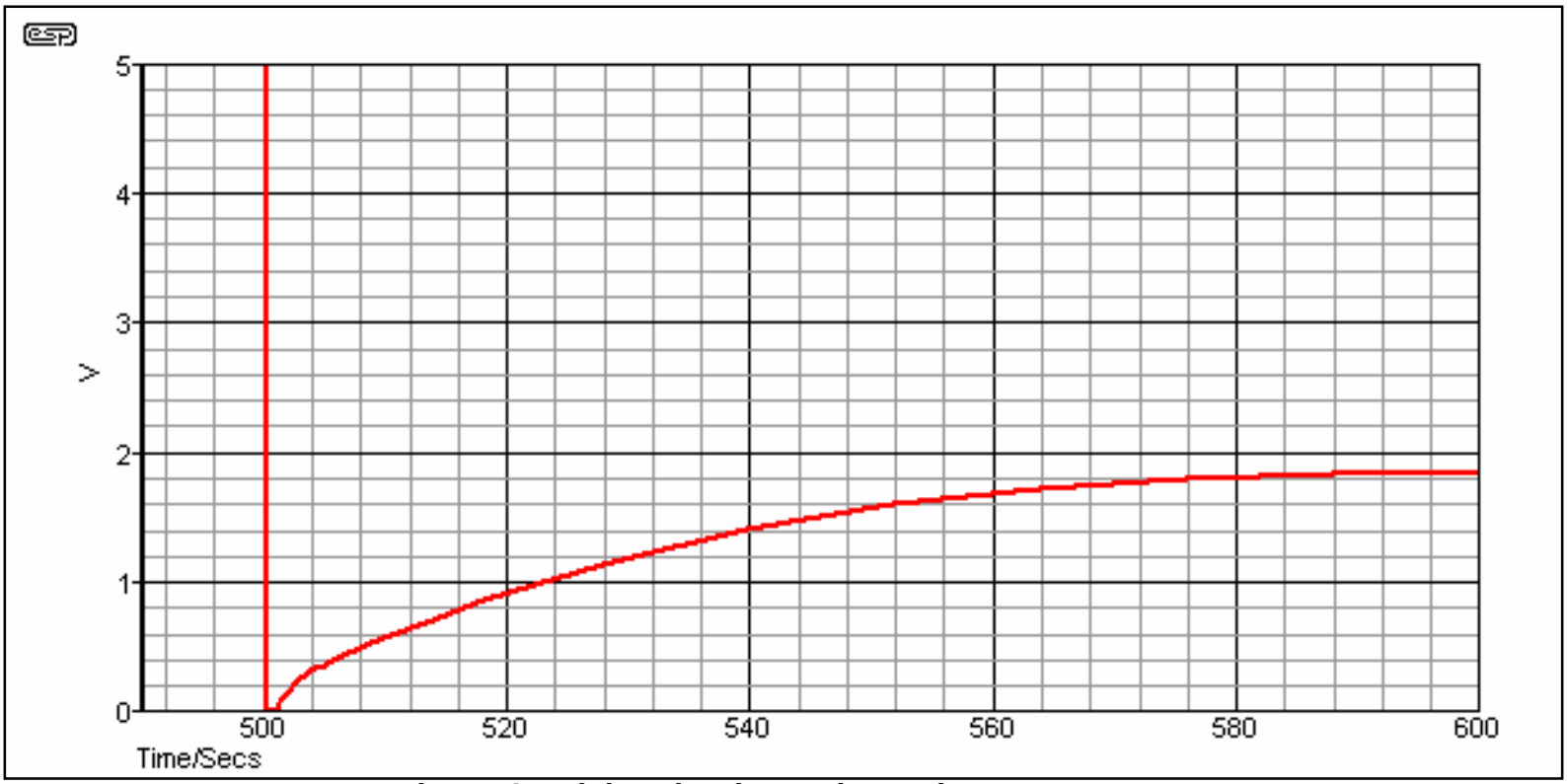


Figure 6 - Dielectric Absorption Voltage Recovery

The test circuit is shown below. This is a fairly standard test, but unless you are building a very low frequency filter or high accuracy sample and hold circuit, the effect is rather meaningless. It is interesting though. The simulated capacitor is the same as the lossy version shown in Figure 5. The official military specification test circuit for MIL-C-19978 (the test for dielectric absorption) uses an opamp wired to give almost infinite input impedance, because standard digital multimeters will not allow a useful measurement. The typical input impedance of a digital meter is 10 or 20 Megohms, and normal audio circuit impedances are much less than that - consequently any 'problems' caused by dielectric absorption will also be much lower than specifications indicate.

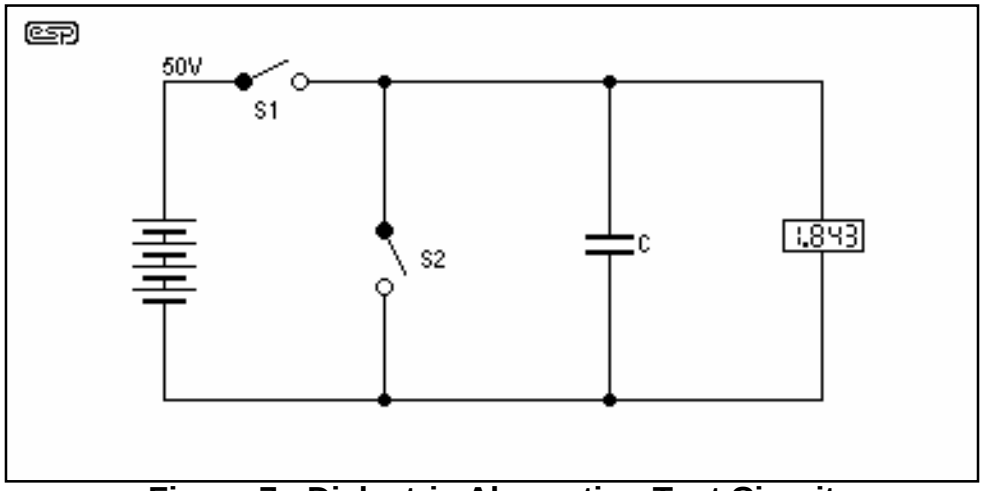


Figure 7 - Dielectric Absorption Test Circuit

The capacitor is charged for 500 seconds using SW1, then discharged (for one second) by

SW2. After the discharge, the voltage is seen to rise again, even though it was obviously zero for the duration of the short. Real caps do exactly the same thing, and if they were used in circuits having close to infinite impedance, it would be a problem. In long period sample and hold circuits, dielectric absorption *is* a problem, but in audio circuits it causes an almost immeasurably small loss of signal. Nothing more.

Once the cap is loaded with normal circuit impedances, the effect goes away almost completely. This assumes that caps will be charged then discharged in an audio system, but as covered above, that does *not* happen in normal audio circuits. Even in filter circuits, the effect is negligible - dielectric absorption does not magically create reverberation, sub-harmonics, background 'glare', 'whiteness' during silent passages, image smearing, ingrown toenails or cardio-vascular disease. Again, all this particular 'audio nightmare' (as some might have you believe) achieves is a tiny loss of signal.

With a 22k load resistor, the maximum 'recovered' voltage is 4.45mV, at 1.2ms after the short is removed (-81dB). Remember that this was after charging the cap to 50V for 500 seconds, then shorted for one second. This is not a normal audio circuit, and no audio circuit will subject a capacitor to anything even approaching the conditions used here.

Caps in audio circuits are simply not charged and discharged in this manner. To do so would cause signals to be generated that, after amplification, would mean instantaneous speaker disintegration. These tests are silly - they prove nothing, but are regularly hailed by some audiophiles as some kind of 'proof' that they can hear a difference because it can be measured. It is forgotten in the excitement that the signals and tests that form such proof will never occur in a real audio system that is not in the process of blowing up.

I have heard claims that the voltage recovery characteristic causes distortion similar to reverberation. What complete rubbish! If it were that simple to create reverb, one can be sure that no-one would have ever bothered with reverb springs, plates, or digital delays. Utter nonsense - it simply does not happen.

2.2 - Electrolytics

Electrolytic capacitors are definitely a problem though - there is any amount of proof ... Or is there ... ? Again, often claims are made based on tests that are irrelevant for audio. A popular myth is that electros have considerable inductance because of the way the foil is wound inside the can. This is nonsense - the foils are usually joined at the ends in the much the same way as with film caps. High frequency performance usually extends to several MHz [2], even with standard off-the-shelf electros and bipolar (non-polarised electrolytic) caps.

Electrolytics do have ESR (equivalent series resistance) as do all capacitors, but because of the nature of the internal chemistry of electrolytic caps it is non-linear. What is important here is not the non-linearity itself, but just how much signal is developed across the cap in normal (properly designed) circuits. We would be foolish to use electros in filter circuits, because they change their capacitance, ESR and inductance with varying temperature and frequency.

Electrolytics are not usually a problem with audio circuits, provided they are used only for coupling and decoupling applications. Because the AC voltage across the cap is so small (by design), the component's contribution is negligible. If you use electros for coupling, I would recommend that you use a value 10 times greater than needed for the design rolloff frequency. For example, if you were to exchange a 1uF film+foil coupling cap for a bipolar, the bipolar should be 10uF. This keeps the voltage across the cap to the absolute minimum at all frequencies.

2.3 - Distortion Tests

Imagine an electro, whose characteristics are so poor that it develops almost 10% distortion internally, with an applied voltage of 1V. This is a particularly bad capacitor, but it is sized so that the AC voltage dropped across the cap is 1% (10mV) of the applied voltage. This means that there is only 10mV AC across the cap, and the distortion across the load will be less than 0.1%. In reality, no electro will be that bad, so provided the voltage across it is kept to the minimum, distortion is not a problem.

Upon testing some 1uF 63V electros (polarised), the readings were interesting. My signal generator has a residual distortion of 0.02%, nearly all third harmonic. Connecting the 1uF electro directly across the output reduced the output voltage from 10V to about 5.5V RMS. This is because the generator has an output impedance of 600 ohms, and the cap was acting as a low pass filter. Figure 8 shows an equivalent circuit of the test setup.

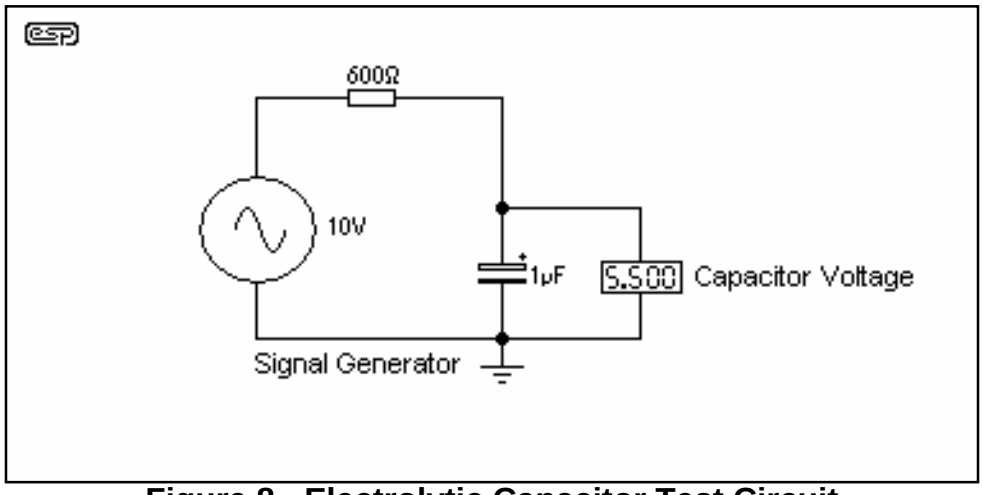


Figure 8 - Electrolytic Capacitor Test Circuit

The electro under test was unbiased, and with 5.5V AC at 400Hz across the cap, the distortion rose from 0.02% to 0.022% - a definite increase, but only small. At lower voltages such as 3V open circuit (about 1.6V RMS across the cap), the distortion *fell* to just over 0.015%. The reason the distortion appeared to fall is simple - the connection shown forms a low pass filter, which helps to remove the harmonics that make up the distortion component of the signal. A first order low pass filter will reduce the third harmonic sufficiently to make reading the difference quite easy. Based on a very similar test done using the simulator, the distortion should be about ½ the generator value, so the cap is still introducing some distortion. As the voltage (across the real capacitor) is reduced, so is distortion, until the noise limit of the distortion meter is reached.

Now, remember that this was using the electro in a way that was never intended. I subjected it to a relatively high applied AC voltage (where as a coupling cap it would have a great deal less - millivolts instead of volts), and it was unbiased as well. Even so, the increase in distortion was small, even with 5.5V AC applied, and it is safe to say that distortion was negligible below around 1.5V, and rapidly fell below the threshold that I am able to measure.

Attempting the same test described above with a polyester cap was a dismal failure - I was not able to measure the cap's distortion, only the attenuated distortion of the signal generator. As predicted by the simulation, measured distortion was about half that of the generator alone, using a 1uF cap at 400Hz, with 5.5V AC across the cap itself. I am satisfied that the polyester capacitor's contribution to measured distortion was well below my measurement threshold.

Various ceramics were a completely different matter though. A 0.22uF (220nF) ceramic was tried, as was a 100nF multilayer bypass cap, along with a few others. At any reasonable voltage, distortion was measurable - the worst measured distortion being 3% with 9V RMS across the cap. This was measured across the capacitor, so the actual distortion was worse than indicated because the capacitor was attenuating its own harmonics. I was unable to measure *any* distortion contributed by a 220pF 50V ceramic, even with 10V RMS at 100kHz across it.

2.4 - Capacitors in General

Electronics World did an epic series of articles written by Cyril Bateman [3], where he went to extreme measures to develop equipment to be able to measure the distortion of common capacitors. Again, this was done with an AC voltage applied across the cap, so the results are generally of far less importance for a coupling cap. The findings are useful for determining the usefulness of various caps in filter circuits (especially passive crossover networks) though, and he quickly disposes of a number of persistent myths, including (but

not limited to) the following:

- All ceramic capacitors introduce distortion
- Dielectric absorption compresses dynamic range and "smears" audio
- Polyester dielectrics are lossy and inefficient
- Capacitors look inductive at audio frequencies
- A capacitor's ESR is a fixed quantity

For anyone who wants to examine these findings in greater detail, I strongly suggest that you get hold of the original series of articles. In general, it was found that the distortion of capacitors was generally very low - well below that contributed by the majority of active circuits. There are very good and valid reasons not to use certain capacitor types in some applications, and equally good reasons to insist on their use in others.

For bypassing, so-called monolithic ceramics are very good, having a low impedance up to hundreds of MHz, assuring good supply bypassing at the highest frequencies. As frequency increases further, standard ceramics are usually preferred. Using them in an audio active crossover network would be a very bad idea though, because their capacitance tolerance is not good, and the value can also change with applied voltage and temperature. Some ceramics (high k types being the worst) may be microphonic due to the piezoelectric properties of the ceramic substrate. I have encountered this (many years ago) in some guitar amps, where large (19mm diameter) 0.22uF ceramic caps were used in the tone control circuit. While newer (and much smaller, both physically and electrically) caps are less likely to suffer from microphony, it is worth bearing in mind. As with dielectric absorption, microphony is more likely to be a problem in high impedance circuits. In most audio applications it will rarely be an issue, but ceramics in general are not recommended for filters, or as coupling caps in audio circuits. This is because of wide tolerance and capacitance variations with frequency and temperature.

G0G or NP0 ceramics have very low temperature coefficients, and are generally useful in many areas of audio where small values are needed. In particular, they can be used as RF suppression, or as the Miller cap in power amplifiers. While it is generally thought that polystyrene or silvered mica (for example) would be better, this may be more of an expectation than a reality. This is something I have tested, and have been unable to measure any difference in distortion between a polystyrene and ceramic cap. Frequency response is essentially unchanged, as is slew rate.

Electrolytics are excellent for power supplies, and most places where high values of capacitance are needed. They are unsuitable for filters, because they have wide tolerance, should be biased, and may change capacitance depending on applied frequency. Bipolar electros are excellent where high values are needed, and no polarising voltage is available. Because of wide tolerance, they too are unsuitable for filter circuits. The distortion caused by

bipolar electrolytic caps used for signal coupling (including their use in feedback networks to ensure unity DC gain) is low to immeasurably low if they are selected to have minimum AC signal voltage developed across them at all frequencies of interest.

When it comes to high current applications (such as passive loudspeaker crossover networks), there will be significant voltage across the cap and current through it. It pays to use high quality capacitors that can withstand the voltage and current that the caps will be subjected to - this generally means polypropylene, polyester, or perhaps paper-in-oil (if you must). This is an area where dielectric loss may cause the caps to heat up with sustained high power, and the devices used need to be stable with time and temperature. Do not necessarily expect to be able to hear any difference between these (high quality) types in a blind test though, as you may well be disappointed.

One thing that may be very important for passive crossover networks is the material used for the 'plates' of the capacitor. Metallised film caps may not be the best choice because of the resistance of the film itself. The film is usually extremely thin, and it may not have a low enough resistance to allow the full current required. I have not experienced any problems with this, but a film and foil type is more suited to high current operation than a metallised film construction. This topic is mentioned on capacitor manufacturers' websites, and I recommend a search if you want more information about current handling capacity.

Bipolar (non-polarised) caps are (IMO) simply unsuitable for use in passive crossovers, because they are so small for their capacitance that heating is far more likely ... whether because of power lost in ESR or dielectric losses. Wide tolerance also means that the network will probably not be right unless it is tweaked, and it will change with time anyway.

3.0 - Test Methodology

The standard (subjectivist) test method with capacitors (indeed, with many electronic components) seems to be to exchange a standard unit with one often costing a great deal more, then to proclaim that ... "Yea indeed, behold the huge difference", and "Lo, see how great is the improvement". As I have noted many times, this is flawed reasoning, and any such test is utterly invalid. Nothing can be gained from this except a continuation of the 'pure subjectivist' dogma.

In a properly conducted test, the test methodology will force the listener to determine if there is a difference between two pieces of equipment (or even any two components), and do so without knowing in advance which is which, and, to do so with *statistically significant accuracy*. This is usually taken to be around 70% - the listener must pick 'Part A' from Part B' correctly *at least* 70% of the time.

According to the claims one might hear from some people regarding their favourite capacitor (or anything else), any blind test should score 100% accuracy, such is the difference heard. Sadly, it seems that in any blind (or ABX) test, the difference fades to nothing, and test results are nearly always inconclusive - it cannot be said with certainty that a difference was heard or not. I cannot understand how something can be claimed on one hand to be 'chalk and cheese', yet cannot be reliably identified as soon as the visual cues are taken away. This should alert everyone to the fact that experimenter expectancy and/or desirability are the overwhelming factors, and that the components themselves are sonically virtually identical.

The actual testing of components must be done with care. The components must be tested in a manner that reflects the way they will be used in practice, or, if this fails to yield any measurable result, the degree to which the part is pushed beyond its ratings must be explained. The report should then extrapolate (in as far as practicable) the measured results at elevated operating conditions to the expected result at normal levels. That this is rarely (if ever) done is fair warning of the likelihood of erroneous data being propagated.

I have never been able to measure the distortion of a capacitor that is used sensibly in a real circuit. This is partly because the equipment I have does not have the extraordinary resolution needed to be able to measure such low levels of distortion, and partly because the active circuitry and system noise will usually predominate. There is little point trying to measure signals that are -100dB below the 1 Watt level, or even worse, at -100dBm (i.e. referred to 775mV).

For example, let's look at distortion at -100dB referred to full power of an amplifier. Assume that the loudspeakers are 90dB/W/m and the amplifier is 100W. The peak SPL is 110dB (unweighted) at 100W, and you might be blessed with an exceptionally quiet listening room - let's say 30dB SPL. If you have distortion artefacts at -100dB, then with a peak SPL of 110dB, the distortion will be at 10dB SPL unweighted (110dB - 100dB). Your very quiet listening room is a full 20dB noisier than the distortion components!

Based on my own observations, as well as those from many others (Bateman, Self, et. al.), capacitor distortion in any real circuit will generally be (much) less than 0.001% ... that's a level of -100dB. Testing and obtaining good results at these levels is highly problematical. Circuit noise, residual distortion and even a tiny bit of corrosion on a connector will increase the measured distortion dramatically. Cyril Bateman was forced to build specialised test equipment to measure the distortion, and while anyone can do the same, it is time-consuming and expensive to do so.

Returning to the use of a cap for signal coupling ... let's assume a seriously non-linear capacitor as shown in Figure 9. when used with significant current through the cap (A), the simulated distortion is 1.26% at V1. When used for coupling, distortion is zero - there is

simply not enough current through the non-linear circuit to cause a problem. Real tests show the same behaviour - the 1uF polarised cap that so happily gave measurable distortion before shows none that I can measure. The simulator also shows zero distortion at V2 when the non-linear cap is connected as a coupling capacitor (B). It is not until the load resistor (22k) is reduced sufficiently to cause significant voltage across the cap that distortion becomes measurable. For example, when the 22k resistor is reduced to 1.5k, distortion rises to 0.0076%. At 600 Ohms, distortion is 0.85%. The diode shown is 'ideal', so it will conduct at very low voltages.

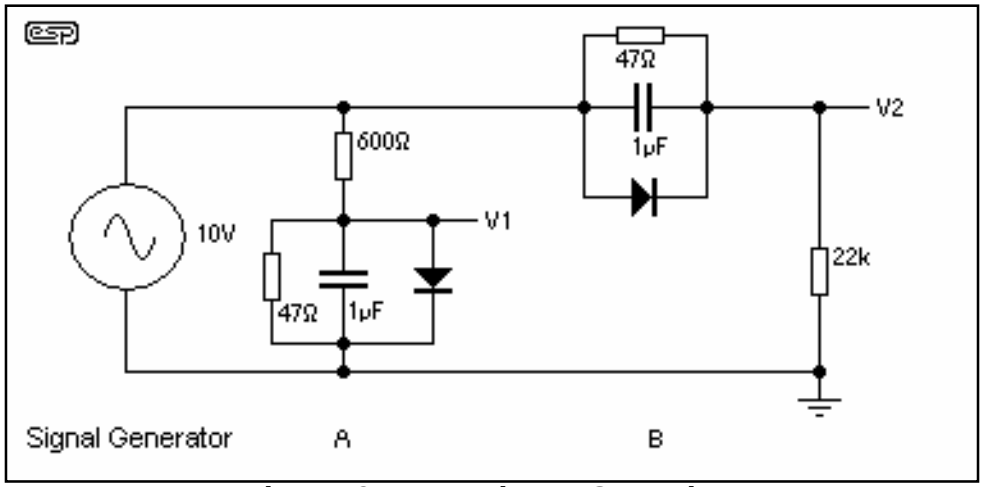


Figure 9 - Non-Linear Capacitor

Although this is obviously a simulated experiment to show the general principle, reality (including test results on the same electrolytic cap that produced measurable distortion before) is very close. If a capacitor is going to cause measurable distortion, then the signal voltage across it (and also the signal current flowing through it) must be significant. If neither of these is true (little or no voltage across the cap, and little current through it), then it is quite reasonable to expect that the contribution of the component at that frequency is negligible, and any inherent distortion it produces must also be negligible.

4.0 - Parasitic Inductance in Bypass Applications

All capacitors have some inductance, but what is often overlooked is that the leads are the primary cause for this. To minimise the inductance, keep the leads as short as possible, and keep them as close together as possible. When two conductors are run in parallel, they form a capacitor. By maximising the (capacitive) coupling, you automatically reduce the inductance. Loudspeaker cables have been produced that have extraordinarily low inductance, despite the fact that they are quite long, and should (in theory) have high inductance as well. Not so. They have high *capacitance* (sufficient to give many amplifiers severe heartburn), but inductance is low. The closer the conductors are spaced and the greater the physical area, the greater the capacitance and the lower the inductance.

Now, consider a conventional wound film and foil (or metallised foil) capacitor ... even if the plates were not joined at each end to form a (relatively) solid block (see figure 1B, or do a Web search for capacitor construction), the capacitance would be at the required value, and inductance would *still* be negligible. The mechanism that supposedly causes internal inductance has never been demonstrated for film caps, but a great many measured results have neglected the capacitor lead length, resulting in erroneous figures. The errors can easily exceed an order of magnitude with a poorly set up experiment.

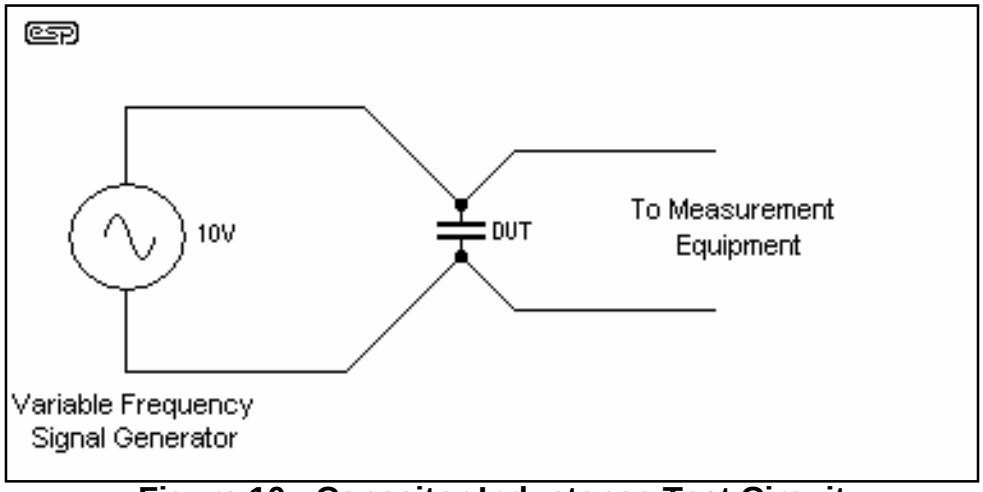


Figure 10 - Capacitor Inductance Test Circuit

The measurement must be taken from a point as close as possible to the capacitor. If the measurement is taken even a few millimetres away from the capacitor itself, it will include the lead inductance. This is made worse by spreading the legs of the cap to allow convenient connection. The inductance of the leads can be calculated by [4] ...

> $L_{dc}= 2L[\ln(2L/r) - 0.75]nH$ where L_{dc} is the low-frequency or DC inductance in nanohenries, L is the wire length in cm, and r is the wire radius in cm.

The inductance is not great ... about 5-6nH per cm (centimetre), but it is still significant at very high frequencies. With a 1uF cap (hardly massive), a mere 10mm of lead length (6nH) creates a series resonant circuit at close to 2MHz. Increase the capacitance to 10,000uF, and it is now 20kHz. This is not capacitor resonance, it is a resonant circuit formed by the capacitor and the *external* inductance of the capacitor's leads. For bypassing applications, the resonant circuit so formed *does not* reduce the effectiveness of the bypass capacitor if it is 'too big'. In reality, power supply bypass capacitors will supply the current required by the circuit regardless of the 'self resonant' frequency, so small values of capacitance do *not* mean better bypassing.

Figure 11 shows a simulated power supply and switching circuit, with inductive leads to the MOSFET and its load. Given the 'self resonant' frequency of the capacitor and lead inductance (about 35kHz with a 1,000uF cap), one would expect that the pulse performance

would be woeful, but it is essentially unchanged as C1 is changed from 1uF up to 10,000uF. If the value is reduced (less than 1uF), then performance does suffer. Lower capacitance *does not* give better bypass performance.

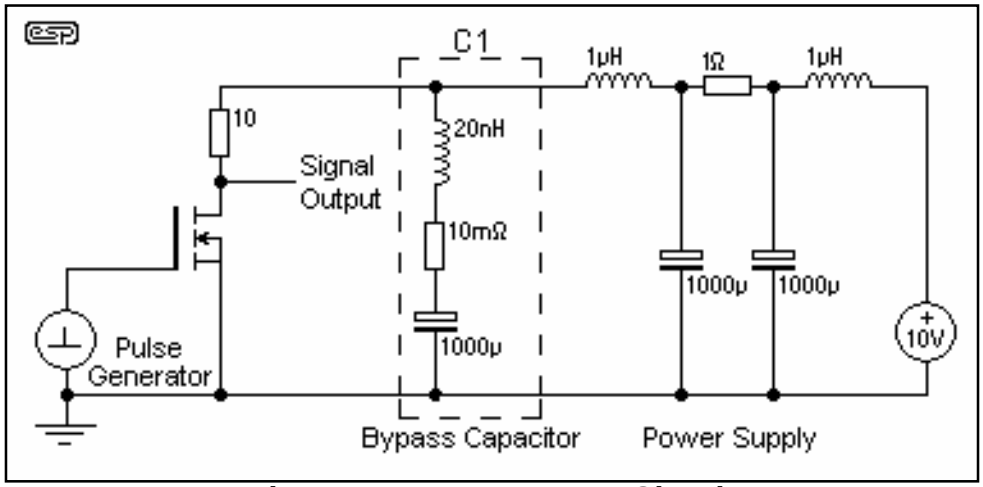


Figure 11 - Bypass Test Circuit

In Figure 12, you can see the waveform of the switching pulse with (red) and without (green) the series inductance and resistance - a comparison between a real and a perfect capacitor. Note that there is very little difference. This, despite the fact that in theory the 'combination' capacitor has a series resonance of 35kHz, and the switching speed is many, many times that. Using a much smaller capacitor (such as 100nF) is a disaster, allowing the circuit to ring, and develop excessive back-EMF. Feel free to perform the test using real components - you will get very similar results!

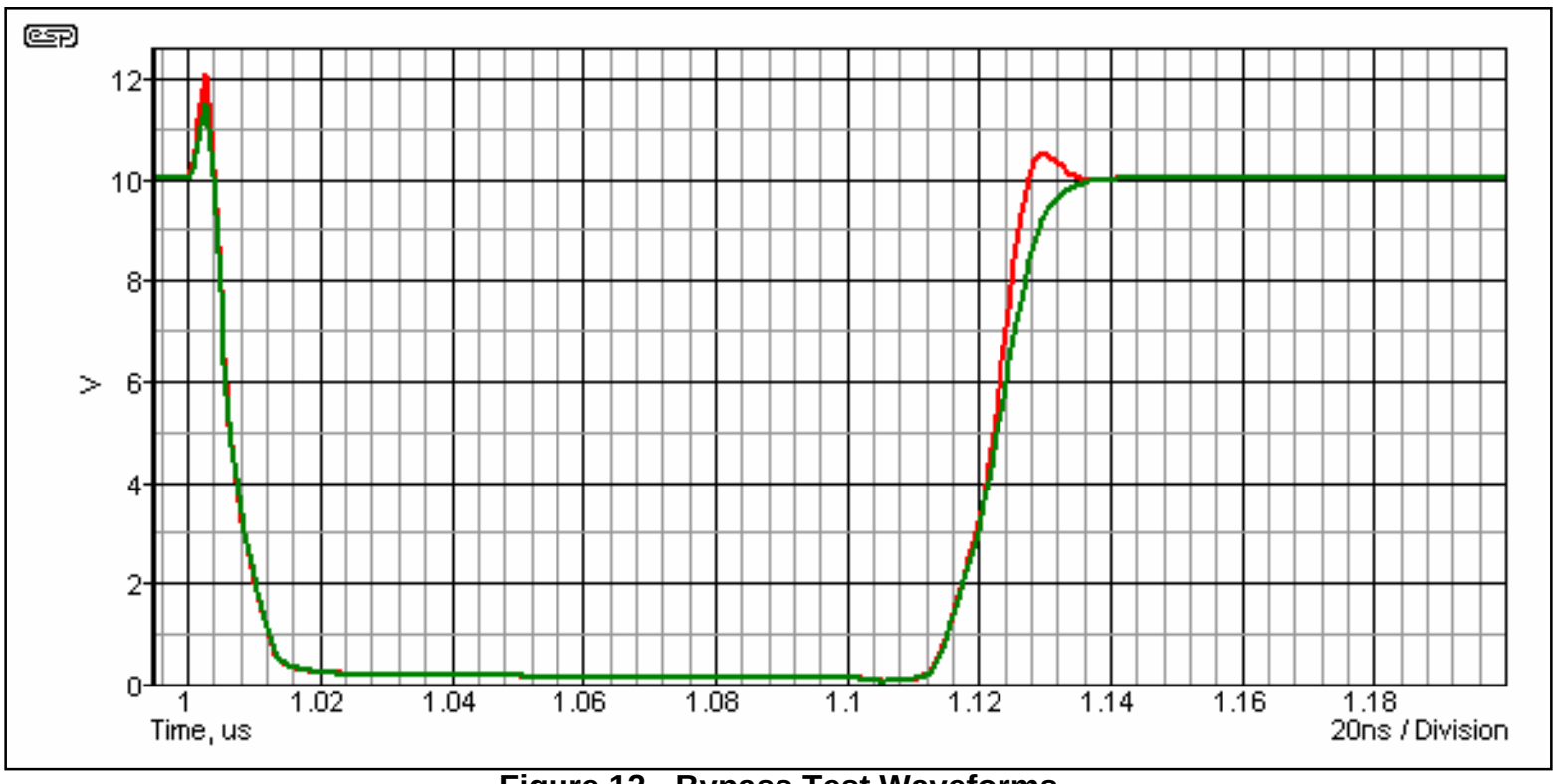


Figure 12 - Bypass Test Waveforms

An important thing that is often missed is that the resonance formula ($f_o=1/2\pi\sqrt{LC}$) only *implies* that higher capacitance values cause lower 'self-resonance' and worse high frequency performance. This is largely untrue - the ability of the larger capacitor to supply instantaneous current demands is not impaired, so the idea of using a small cap ("they have a higher self-resonant frequency") in parallel with a big cap is essentially nonsense - more capacitance equals more energy storage. The concept of 'self-resonance' in this context is flawed thinking, and leads to silly designs (100nF caps in parallel with 10,000uF electros for example) that generally achieve nothing useful, other than using more components.

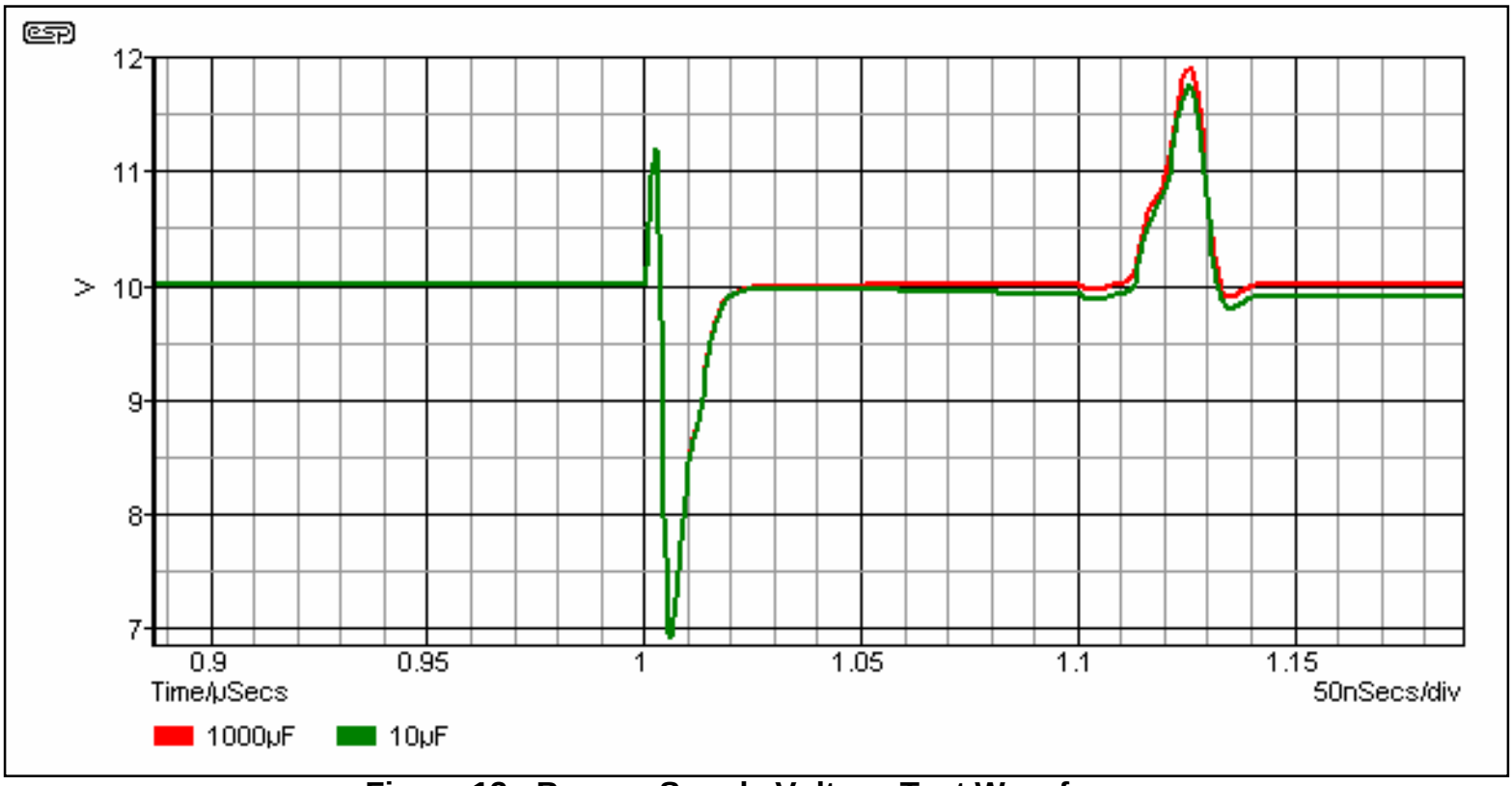


Figure 13 - Bypass Supply Voltage Test Waveforms

In Figure 13, you can see the difference between using a 1,000uF (red) and 1uF (green) bypass capacitor, measured at the positive supply to the switching circuit. The 1,000uF cap *should* show a sluggish response because its 'self resonant' frequency is so low. As power is demanded (MOSFET switched on), there is no difference at all, although recovery is a tiny bit slower. Not fully visible is the fact that the low value cap causes a damped oscillation, whereas the higher value does not. So, do low value caps 'work better' as bypass? ... No, they most certainly do not.

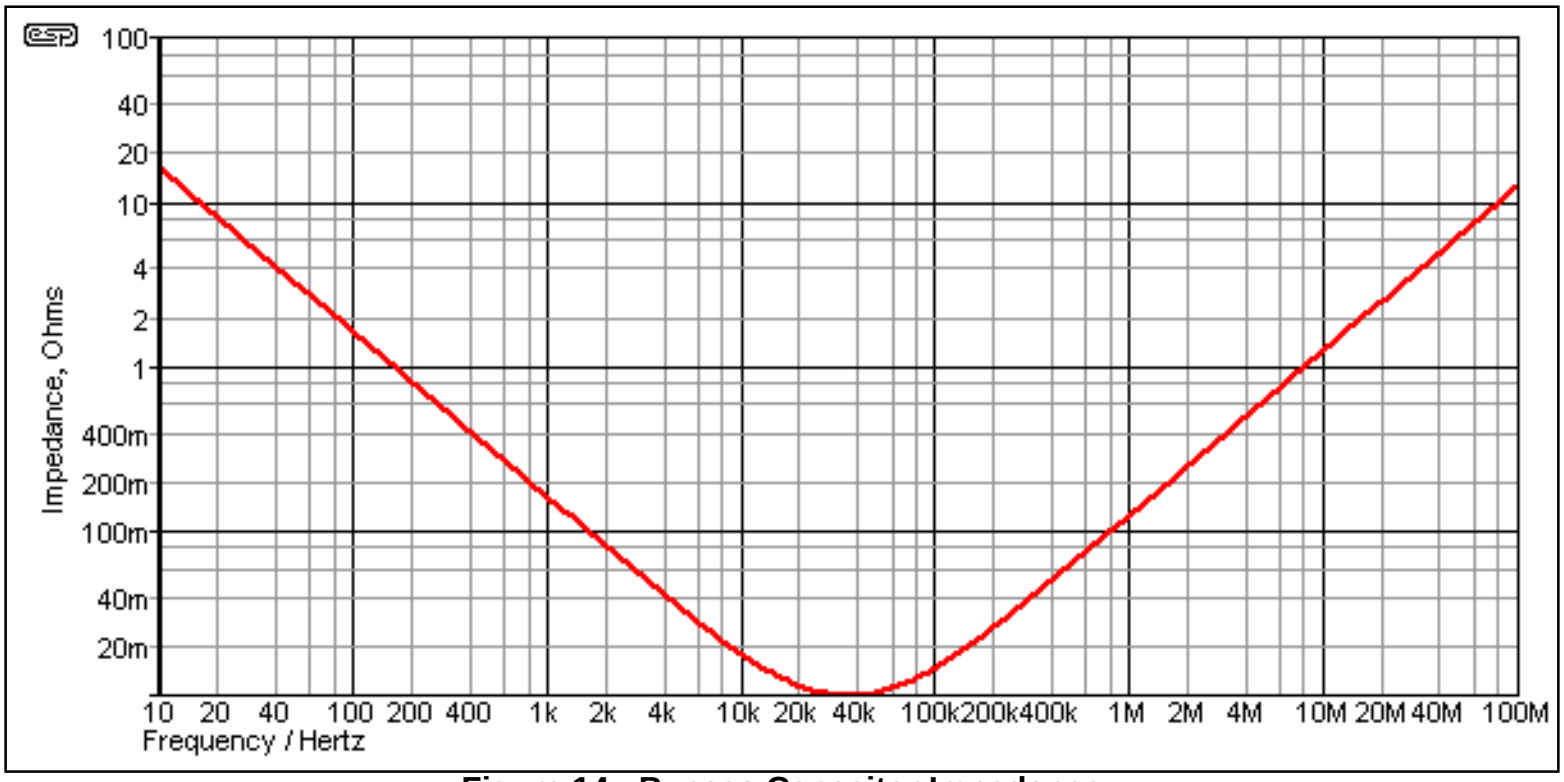


Figure 14 - Bypass Capacitor Impedance

Figure 14 shows the impedance of a simulated 1,000uF capacitor with 20nH series inductance and 10mΩ series resistance. The 'self resonance' frequency is 35kHz, with a minimum impedance equal to the series resistance (ESR). Even at well above the resonant frequency it is *not* an inductor, despite appearances. This is commonly claimed, but is false logic. The impedance is increasing, but until such time as the inductive reactance becomes significant (with respect to the circuit impedance) the composite circuit is still a capacitor. Even at 1MHz, the total impedance is only 125 milliohms. Although the 125mΩ is almost all inductive reactance, it cannot be considered significant - that is a somewhat vague term that is usually taken as around an order of magnitude compared to the load. In this case, the load is 10 ohms, so 1 ohm is 'significant'. This occurs at 8MHz. It is very important to understand the difference between a supply bypass application and a tuned circuit or other electronic function.

In a normal circuit (such as a series tuned circuit for example), when the applied frequency is the same as the resonant frequency of a capacitor and inductor (including leads, PCB tracks, etc.) the tuned circuit is no longer reactive - it is resistive! The resistance is equal to the sum of all component and lead resistances (including ESR). Below resonance, the circuit is capacitive - above resonance, inductive. Series resonance in a capacitor may result in rather unexpected behaviour in high frequency circuits (including digital), depending on the specific application.

It is obvious that capacitor leads should be kept as short as possible, and it would be an advantage if manufacturers stopped spreading and kinking the leads of monolithic ceramics

in particular, as this introduces a (small) additional inductance because of the lead length and reduces the maximum operating frequency. It is quite obvious that lead (and PCB track) inductance must be considered for very high frequency circuits - or for circuits that are capable of very high frequency operation even though they are used at much lower frequencies (audio amplifiers come to mind).

Some interesting observations are made by Ivor Catt [5], where it is maintained that the vast majority of capacitor claims are false. His information on bypass caps (in particular) goes against all 'conventional logic', yet the simulation described above validates his theory. A couple of his more notable quotes are ...

- If you cut off a capacitor's legs at the knees, you will double its self-resonant frequency
- A well behaved capacitor keeps its legs close together

Ivor is considered eccentric to many in the electrical/electronics fields (some may say that is an excessively generous description), but his data cannot be dismissed out of hand. Particularly when a simulation shows that a capacitor, even with series inductance, *can* supply the instantaneous demands of a switching circuit. This is despite that fact that the switching occurs at a frequency that is well above the 'self resonant' frequency of the capacitor.

Another of Ivor's contentious claims is that a capacitor is a transmission line. Based on the tests conducted (see ESL & ESR below), there is much to commend this model, even though it has been scoffed at by many who (in my opinion) should have been thinking more clearly. A length (any length) of coaxial cable appears to be capacitive at low frequencies, and at a frequency determined by its length, shows series resonance - it becomes (almost) a short circuit for that frequency. Above the resonant frequency, the cable is inductive. The primary difference is based not on any of the counter-claims that I saw to the suggested model, but because the very construction of a coaxial cable is such that it has vastly lower capacitance than any real capacitor, so the resonance is very high Q. In addition, the cable's capacitance and inductance are optimised for the circuit impedance. Capacitors are optimised for capacitance (what a surprise), so generally use a dielectric that is far thinner and has somewhat greater losses than coax. That does not change the basic model though - it simply means that the characteristic impedance of any given capacitor is dramatically lower than that of any 'normal' transmission line.

It is probable that if a capacitor were to be laid out flat rather than rolled up in the normal way, its inductance will not increase by anywhere near as much as the pundits might imagine. This, despite the fact that when it is rolled up, the entire edge of each plate is joined, so the 'length' of this transmission line is the *width* of the metallised film (or separate foil). This agrees quite well with the measured or calculated internal inductance of almost all capacitors, and this is easily verified by anyone with access to basic RF test equipment.

Electrolytic Series Resonance

The series resonance of an electrolytic has to be considered in conjunction with the circuit impedance. In real life devices, it is actually quite a broad null, often extending over several decades of frequency. This is readily apparent from looking at manufacturers' data, or by measurement. Measurement is actually quite difficult, since a significant current must be applied to be able to see the results. This requires an amplifier with very wide bandwidth indeed, and although some esoteric audio amps may be capable of providing sufficient current to obtain a worthwhile reading, most cannot.

Something that is somewhat easier is to apply a squarewave at (or near) the approximate series resonant frequency. Although the graphs in Figure 15 are simulated, the simulation is based on actual measurements, using a 15,000uF 35V electrolytic capacitor. This component has a series resonant frequency of around 40kHz, however, this comes with caveats. It is a very broad resonance (as expected), and with 10V RMS applied via a 10 ohm resistor, I was able to obtain a readable trace on the oscilloscope.

The following graphs show two traces ... the first (red) is the waveform across the cap with 50mm of lead between the cap and the measurement point. The second (green) is with the oscilloscope probes placed as close to the capacitor as possible. A mere 50mm of lead equates to approximately 20nH of inductance, but as seen in Figure 16, that is enough to double the amplitude of the spike at the leading edge of the waveform. It is also enough to lower the 'self resonance' quite dramatically - the low frequencies are unchanged, but the high frequency (where the impedance starts to rise) is moved up by nearly 400kHz. The minimum amplitude difference is because of the lead resistance (simulated as 10 milliohms).

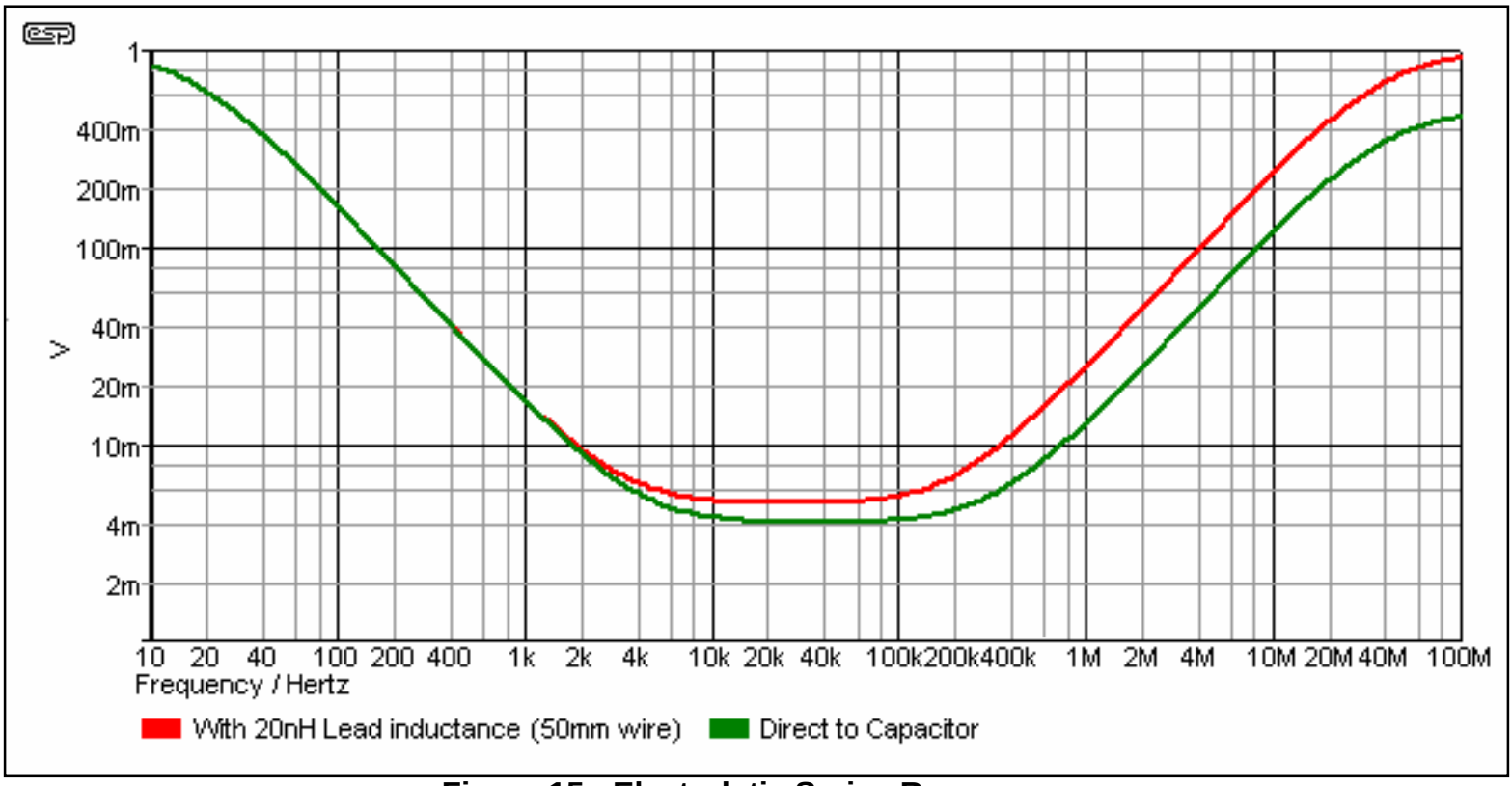


Figure 15 - Electrolytic Series Resonance

These simulations agree quite closely with the measured results, so even though there will be some variance, it will be less than that obtained from different samples of real-world components. As with the pulse response, there was zero measurable difference when a 220nF film cap was added in parallel - either with the extended leads or without. The simulation agrees for the most part (but only if a 'real world' capacitor with losses is used), and I have not included these data.

In a real circuit, there is a possibility that a small film capacitor in parallel with a large electrolytic may cause ringing (damped oscillation) at a frequency determined by the series inductance of the electro and the capacitance of the additional film cap. This is more likely to degrade performance than improve matters. It is possible to simulate this easily, but it is not so easy to measure because the frequency will be very high, and the impedance still very low. Because this possibility is rather remote, if it makes you feel better, by all means add a parallel film cap. Don't expect to hear a difference in a blind test though, because you almost certainly will not.

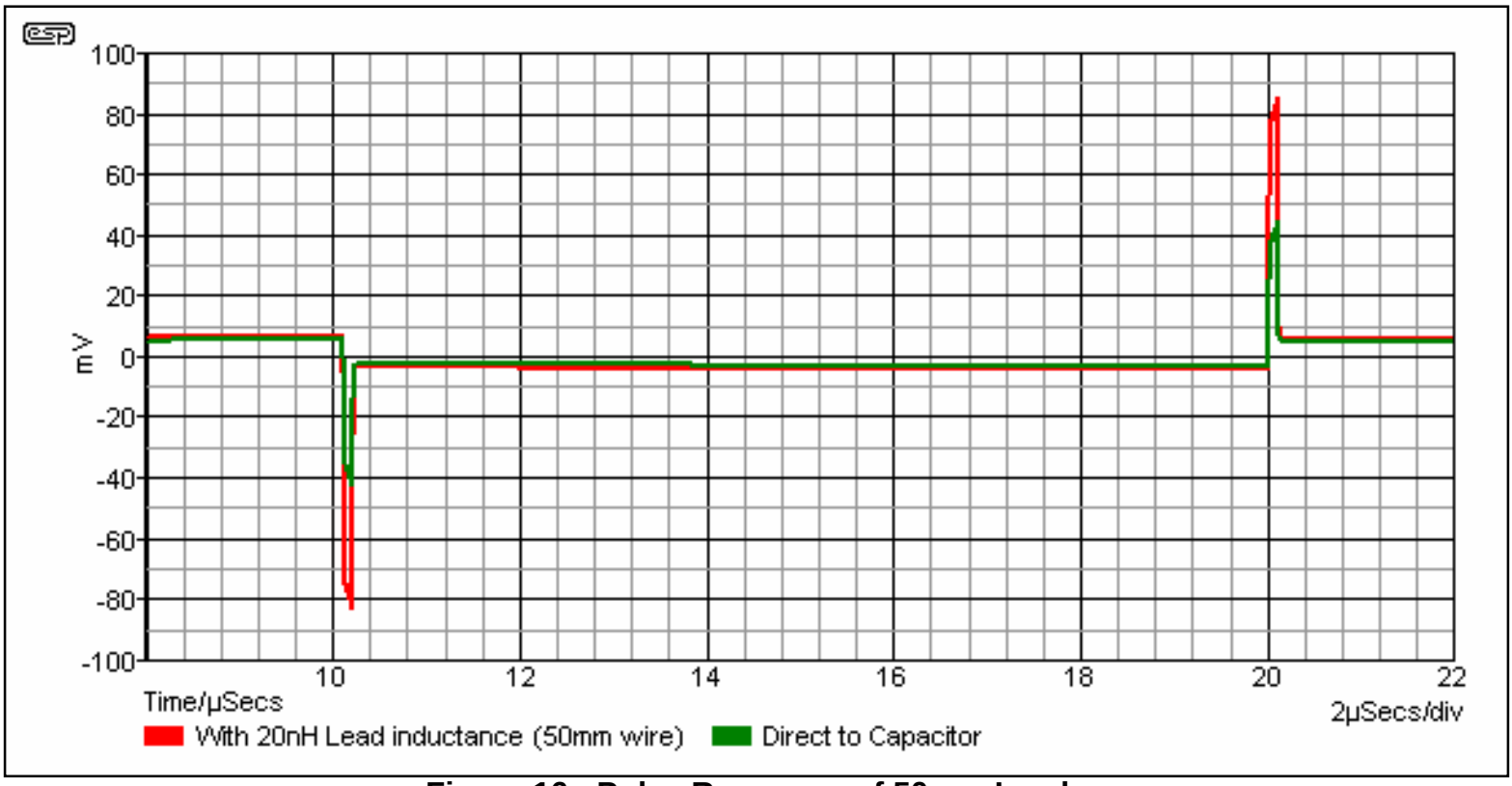


Figure 16 - Pulse Response of 50mm Leads

The pulse response is interesting. This is as close as I could get to the actual measured waveform, and contrary to common belief, adding a parallel capacitor (in this case 220nF) *did not change the measured pulse waveform one iota*. The impedance of the film cap is simply much higher than that of the electro, so it cannot have any significant effect on the end result. There *is* an effect, but it is immeasurably small. The impedance (capacitive reactance) of an ideal 15,000uF cap at 1MHz is 10mΩ, but we must add the ESR to that result. According to the simulation, the total impedance is 134mΩ at 1MHz - inductive reactance is responsible for most of that. By comparison, an *ideal* 220nF capacitor will have a reactance of 723mΩ at the same frequency - more than 5 times that of the electro.

Earlier, I made the comment that for resonance to actually work as expected, the circuit impedance must be 'significant' compared to the resonant impedance. It is time to examine exactly what is significant, and what is not. The resonant frequency of a capacitor and inductor is given by the equation

$$f_o = 1 / 2 \pi\sqrt{LC}$$

This is a general formula, and while it holds true in all cases, the Q (quality factor) of the resonance is dependent on the circuit impedance and component losses. In the case of electrolytic capacitors (especially large ones, where the resonant frequency is comparatively low), the capacitance is massively dominant compared to inductance. For this reason, electros will rarely (if ever) appear as a resonant circuit in any power supply or coupling

capacitor application.

The coupling cap is a good one to examine, because this is an area where it is often thought that a parallel capacitor will assist with high frequencies. If we assume a 4,700uF capacitor, having 100mΩ ESR and an inductance of 100nH (this is much worse than a real capacitor), its 'resonance' is at 7,341Hz. The test circuit is shown in Figure 11. As a coupling capacitor, it might appear to have inductance above the self resonant frequency, or so it would seem. Not so, as Figure 12 shows. In fact, the frequency response into an 8 ohm load remains substantially flat up to 4.5MHz (0.5dB down), and is -3dB at 12.7MHz.

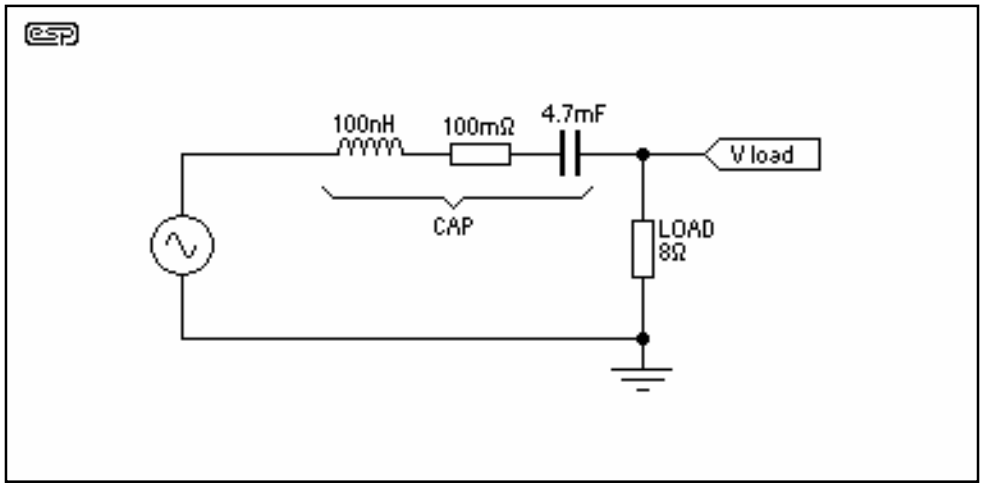


Figure 17 - Electrolytic Coupling Capacitor Test Circuit

In Figure 17, the 100nH of inductance in the capacitor is totally insignificant compared to the circuit impedance (4 ohms) and the capacitance. The speaker lead will have a great deal more inductance that the capacitor, so the frequency should be even lower than calculated. It doesn't happen ... well, actually it *does* happen, but the effect is so infinitesimally small that it can only be measured by simulation. In the example given, the difference between the voltage across the load resistor at 100kHz is reduced by less than 250udB (micro decibel), compared to the 7.3kHz 'self resonance' frequency.

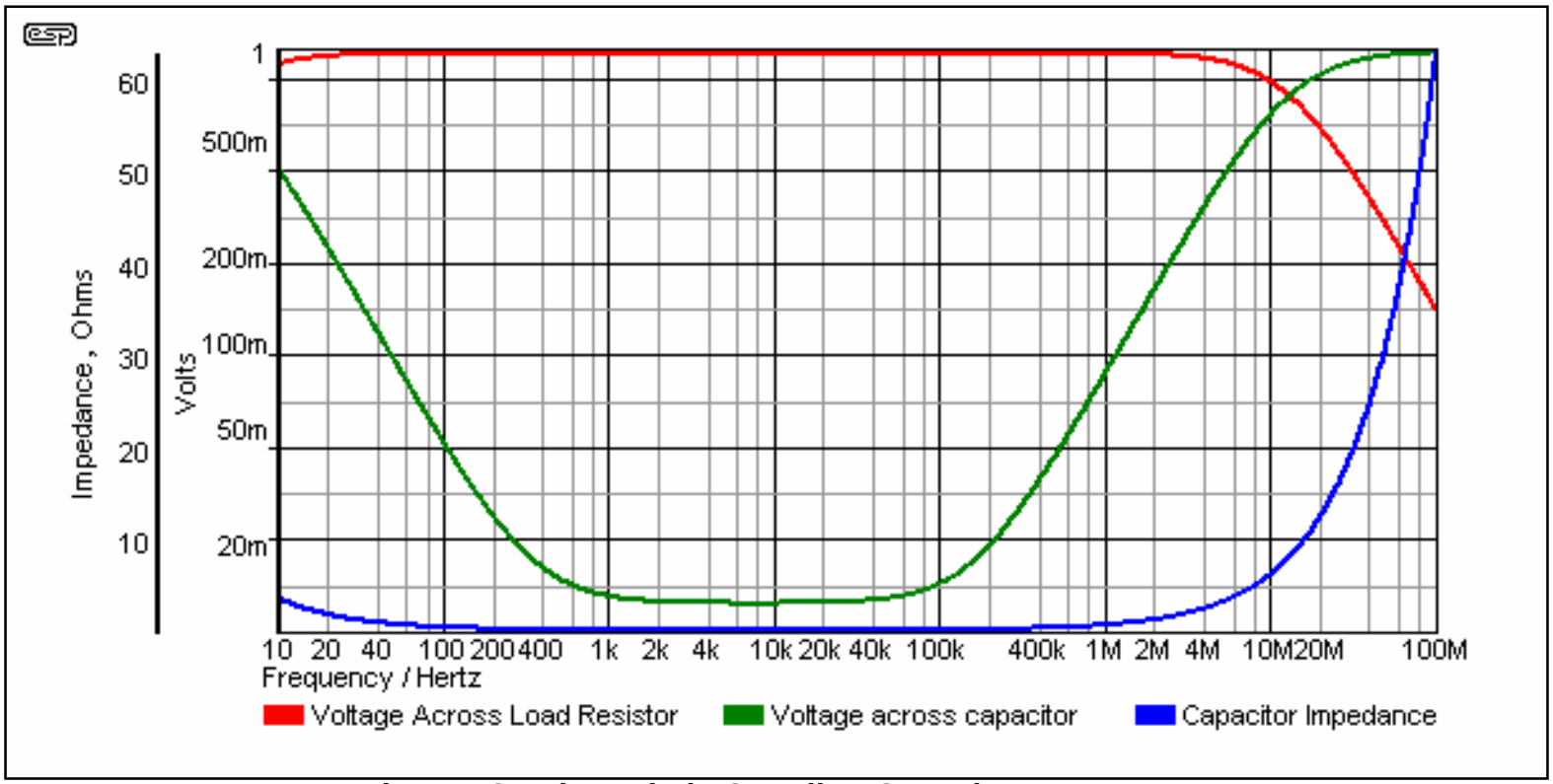


Figure 18 - Electrolytic Coupling Capacitor Response

For self resonance to be noticeable, the circuit impedance needs to be in the same general range as the capacitor's inductive impedance at resonance. As can be seen from the graph, the inductive reactance only reaches 8 ohms at 12.7MHz (the upper -3dB frequency). Will a 100nF polypropylene cap (or any other type) in parallel be of any use whatsoever? From the above, we can safely say "no". Its reactance will be equal to the 8 ohm load at 198kHz, but at that frequency, the electro has a total impedance of about 200mΩ, making the influence of the small parallel cap insignificant.

5.0 - ESL and ESR

ESL - Equivalent Series Inductance

Claims have been made that most capacitors *must* be inductive, because they are made from a wound sandwich of film and foil, or metallised film. Because it is usually wound (in a flat coil), logically, this leads to inductance. The problem with that theory is that it assumes that the termination is made to the foil at the end only, but a quick check of manufacturer data will show that this is generally not the case. The vast majority of capacitors are made so that the foil or metallisation projects from each side (one 'plate' on one side, the other 'plate' on the other side). Each end is then connected so that all sections of the plate are joined together. There is no longer a 'length' associated with the plate, and only its width becomes significant for inductance. When distortion is measured in a film capacitor, it is almost always the method used to connect to the foil that causes the problem, rather than the dielectric or foil material.

Aluminium is the most common metal for both foil and metallisation, and aluminium is notoriously difficult to attach to anything with good and reliable conductivity. That cap makers have made them as good (and as reliable) as they are is testament to the effort that goes into capacitor manufacture.

The ESL (equivalent series inductance) of any given capacitor is related more to its physical size than anything else. A larger capacitor will almost always have a greater inductance than a smaller version of the same capacity. Usually, the lead length is of far greater importance for high frequency operation.

To check the general principle, I decided to test a roll of telephone jumper wire as a capacitor. This is a fairly large roll of twisted pair (Cat-3), with the diameter of the roll being about 130mm, and 51mm high. The wire is twisted (as you would expect for twisted pair), and the roll contains about 80 metres of wire. Insulation is 0.25mm PVC, and wire diameter is 0.5mm. All in all, this should be an appalling capacitor.

Being a coil of wire, one would expect a high inductance and therefore low self resonance. The measured values were ...

Capacitance	9.67nF
Dissipation Factor	0.059
Self Resonance	303kHz
Inductance	28.5uH

The capacitance was measured at 9.67nF with two different meters, and DF (dissipation factor) was 0.059 ... not especially wonderful, but far better than I expected. Remember, this is a *coil* of twin wire, with the connection made at one end only. Inductance is calculated based on the self-resonant frequency - it is obviously much greater than a normal capacitor, but that is expected due to the physical size of the 'capacitor'. Needless to say, the fact that the connection was made at only one end doesn't help matters, but remember that this is a physically large coil of wire - one would expect that self resonance (and inductance) would be far worse than was the case.

Joining the ends together gave the following ...

Capacitance	9.73nF
Dissipation Factor	0.059
Self Resonance	1.0MHz

Inductance	2.6uH

This is a significant improvement to the inductance (roughly an order of magnitude), and also gives a small increase in capacitance. As you can now well imagine, by joining the entire edge of each capacitor plate, inductance is reduced to almost nothing, and only the physical size of the cap will influence the inductance. This can't be applied to my 'coil capacitor', because I only have access to each end, rather than the edges of the 'plates'. I think it is safe to assume that the wire coil performs far better than might have been expected, especially with the ends joined together. It also has far thicker insulation and smaller plate area than any real capacitor, both of which increase inductance and reduce capacitance.

In case anyone was wondering, the inductance of the coil used above is 15.6mH, with a series resistance of 6.4 ohms. This was measured with the two wires connected in parallel. Self resonance is at 27MHz. Not useful in the context of this article, but worth including.

ESR - Equivalent Series Resistance

While inductance is not affected by the dielectric material, ESR is - it is dependent on the dissipation factor (DF) of the insulation material, as well as the resistance of the leads, plate material and plate terminations. Because DF varies with frequency in most common dielectrics, so too does ESR. However, ESR is rarely a problem in most audio circuits. It is important in passive crossovers used in high powered systems, or for other applications where capacitor current is high. ESR (like all resistance) creates heat when current is passed, so for high current circuits the ESR is often a limiting factor.

ESR is very difficult to measure with low value capacitors, because the capacitive reactance is usually a great deal higher than the ESR itself. In general, it is safe to ignore ESR in most electrolytic and film caps used in signal level applications (such as electronic crossovers, coupling capacitors and opamp bypass applications). ESR becomes very important in high current power supplies, switching regulators/supplies and Class-D amplifiers, many digital circuits and any other application that demands high instantaneous currents that are supplied by the capacitor.

6.0 - Food For Thought

The above is food for thought indeed. I have had several e-mails from readers (within a day of the article being published), and further comments should be made to clarify a couple of important points. Much ado was made above about coupling caps, and these are a favourite of the upgrade brigade. It is not uncommon to see circuit boards where the constructor (or 'upgrader') has used caps for which the PCB was never designed. As an example, look at

Figure 19. This is not at all uncommon, but what is not understood is the potential for possibly major problems to be introduced.

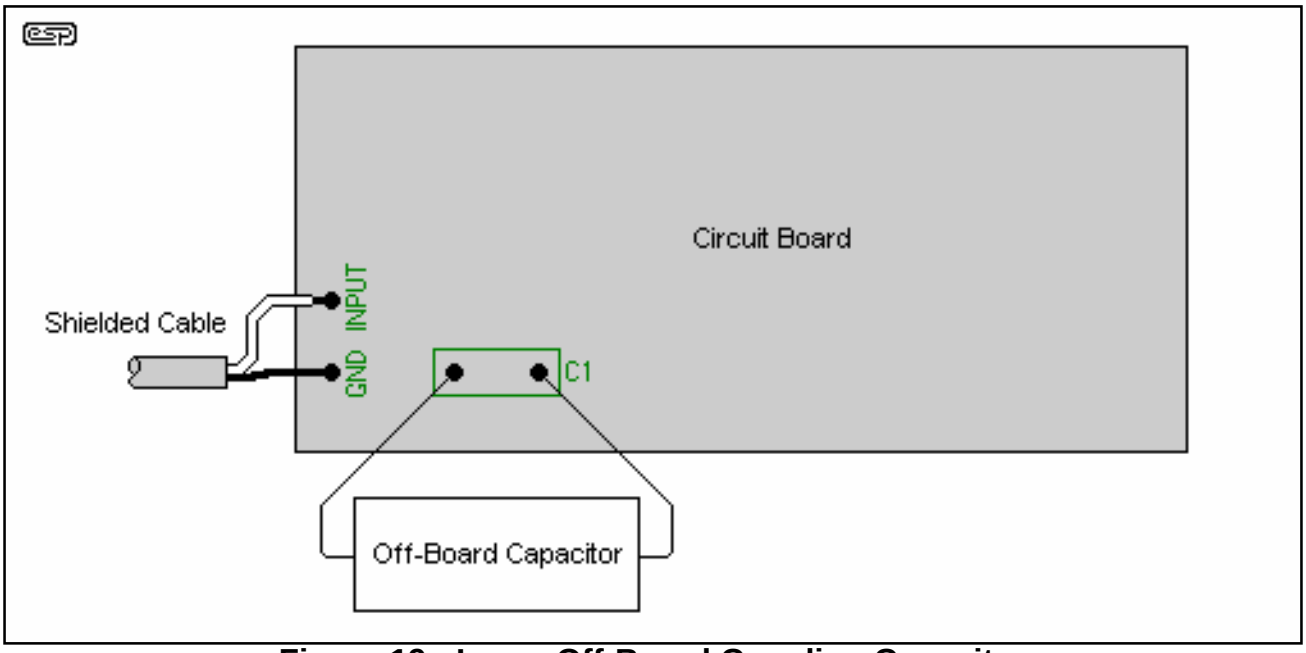


Figure 19 - Large Off-Board Coupling Capacitor

At first glance the diagram looks alright. Everything is connected where it should be, so where's the problem? Notice that the input signal is connected to the PCB via a shielded lead. The PCB may have a ground plane, but even if not, the connection between the shielded input lead is nice and short, and connects to C1 on the board. The space allowed is sufficient for a cap as originally designed.

Now, someone comes along with a massive (physically) cap that was sold as polypropylene (but could easily be polyester). It won't fit on the board, so is installed as shown. Look at the length of unshielded lead between the input terminal and the rest of the circuit on the PCB. Remember that the *entire* capacitor is part of the unshielded circuit, not just its leads. Even if the cap is marked so you know which is the outer foil, that doesn't help either, as any noise picked up will be coupled through regardless (this is what the cap is *for*!). This arrangement has the potential to pick up considerable noise, and if part of a power amplifier may even be sufficient to cause oscillation. It goes without saying that noise or oscillation will *not* improve the sound, even though the owner may think that it has done so.

The likelihood of noise or oscillation depends on many factors of course, and these may not be an issue (or not at a level that is audible). The mechanical reliability is also highly suspect, especially if the oversized cap has not been fastened such that it cannot move relative to the PCB. Had the on-board cap been installed in the position shown, its size is much less, and the board would have been tested with it in position - any problems would be immediately obvious.

Another point made is that series resonance can be used to your advantage. In the presence of a strong RF signal, normal bypassing may be insufficient to prevent the RF signal from getting through an audio system. If you know the frequency, then it is not difficult to tailor a ceramic cap and appropriate lead length to create almost a dead short for the interfering signal.

For example, if there is an AM CB transmitter (these operate in the 27MHz band) nearby that insists upon interfering with your audio system, you need to know exactly where it is getting into the audio path. Once this has been determined (not easy, but certainly possible), you can deduce the necessary capacitance and inductance using the standard formula ...

$$f_o = 1 / 2 \pi\sqrt{LC}$$

If we assume (say) a 1.2nF capacitor, then it works out that a series inductance of 28nH is needed. With roughly 6nH/cm, the cap needs to have leads about 46mm long, and with two leads than means 23mm each. The leads need to be as widely separated as possible, and some adjustment of the frequency is possible by pulling the leads wider apart or pushing them closer together. In its basic configuration, the combination of cap and leads has an impedance of less than 1 ohm between 25MHz and 30MHz, and is resonant (effectively a short circuit) at about 27.5MHz.

Maximum effectiveness is achieved when the circuit is tuned as accurately as possible, but this normally requires specialised RF test equipment. In general, a calculation will get you close enough for the circuit to be effective, and a bit of tweaking should enable you to get almost total rejection of an unwanted signal.

This last point is only made as a generalisation - RF is by its very nature sneaky, and deliberately using capacitor+lead resonance to solve a problem is just one of many techniques that need to be tried to solve an RF interference problem. No one method will work in all cases, and serious problems may need a combination of different suppression tricks. This applies both for preventing RF getting into a circuit, or preventing it from getting out (and therefore causing interference elsewhere).

7.0 - Conclusion

If wine or pharmaceuticals were tested the same way as audio, we would be in a very sorry state indeed. To be valid, all tests must be conducted blind, where the tester does not know which product they are using, or preferably double-blind, where neither tester nor controller knows which is which. That sighted tests are not only tolerated but encouraged is testament to the level of disconnection from reality that many 'magic component' believers obviously

suffer.

Unfortunately, there are many who will search the Net for 'proof' of their current theory, and will use or misuse any data they happen across to further their argument. That the data quoted may be out of context, flawed, or simply a load of codswallop is immaterial. Once these rumours start they take on a life of their own, and it then becomes almost impossible to get the discussion back into the land of reality. This technique has been shown many times with the 'great cable debate(s)',and much the same has happened with capacitors and other generally benign components.

A popular piece of disinformation that really irks me is the claim that ceramic caps should not even be used for bypass applications in audio. This is drivel, and is totally unfounded drivel at that. The purpose of bypass caps is to store energy that ICs need on a short term basis, swamp PCB track inductance to ensure that opamps don't oscillate, and to ensure that digital circuits don't generate supply line glitches that produce erroneous data. There is absolutely no 'sound' associated with DC supply rails. Opamps don't care if the DC comes from a battery, solar cell, or rectified and filtered AC (sine or square wave, any frequency). DC is DC - it has no sound, and it contributes nothing to sound unless it is noisy or unstable. Supplies may be completely free of noise, or might be relatively noisy (especially where digital circuitry shares the same supply). Provided all noise (including voltage instability) is of low enough level that the opamp's (or power amp's) PSRR (power supply rejection ratio) prevents the noise from intruding on the signal, supply noise is immaterial. As always, a blind test will reveal any genuine difference, and a sighted (non-blind) test will reveal the expected result - real or imagined.

To reject ceramic bypass caps, which have the best high frequency performance of nearly all types, is sheer lunacy. This is especially true when discussing simple DC power supply lines. PCBs have capacitance too, and the standard fibreglass material used is fairly lossy - it is certainly useless for very high frequency work at several GHz. Maybe that ruins the sound too - I *have* heard such tales, and they can be discounted out of hand. Still, these fairy stories circulate, are perpetuated by those with a vested interest in separating people from their money, and will continue for as long as anyone is silly enough to believe it.

Just as misleading are claims that all/most electrolytic caps are resonant within the audio band (or at least below 100kHz). Again, while this may be true, it is meaningless without context and simply indicates that the person conducting the test is leaving it up to the designer to determine if that will cause a problem in the circuit being developed. Always remember the importance of the lead inductance - that is usually far more important than a few nH of capacitor inductance.

Large electrolytic caps may well resonate (if a very broad impedance minimum can be considered resonance) within the audio band, but with impedances of well under 0.1 ohm

overall, this can hardly be claimed as a problem. Adding a film cap directly in parallel achieves nothing, because its impedance is many times greater than that of the electrolytic for all frequencies of interest. If there is *any* distance between the large electro and the film cap (for example leads running from the power supply to an amplifier PCB), the caps are no longer in parallel. They are separated by an inductance determined by lead length as well as some resistance, and the extra cap will help to damp out the effect of the inductance. Again, the larger the local bypass cap, the better it will perform.

One thing you can count on ... if anyone wants to sell you 'special' capacitors, designed to replace 'inferior' types (such as polyester, aka PET, Mylar®, etc.), then you know that there is a problem. These vendors are cashing in on the audio snake-oil bandwagon. Like cables, many of their offerings are likely to be of good quality, but at many times the genuine value of the part. Others will be perfectly ordinary parts that have been re-badged. For example, there are many capacitors sold as polypropylene that are actually PET or Mylar. It seems that no-one has ever heard the difference, simply believing that it *is* polypropylene, so therefore sounds 'better'.

I have never seen the specifications for snake oil as a dielectric, but I expect it to have rather poor performance overall. With 'magic' components, in the end everyone loses. DIY audio is supposed to be fun, not an endless search for the mystery component that will make everything sound wonderful. Sad news ... that component does not exist.

"The best cap is no cap" is claimed by some. I would much prefer to ensure that no DC flowed where it is unwelcome by using a cap than to allow a fully DC coupled system to try to destroy speakers given the chance. Perform all the blind tests you can with capacitors used in real circuits. Having done this, if you still think there is a difference (and can demonstrate it to others in a blind test), then you will probably be the first to do so.

If you wish to let me know that I am wrong, feel free to do so ... but *only* if you have conducted a blind A-B test and can provide some verifiable data to substantiate your claim. I regularly get e-mails from people who claim that they can hear the difference between components, leads or whatever, but in every case thus far, no blind A-B test method was used. I am not the least bit interested in hearing about the results of *any* sighted (non-blind) test, because such tests are misleading and simply verify existing opinion. In fact, the 'result' of the entire test is only an opinion, as there is never any data to substantiate the claim.

Electronic equipment is designed using facts and mathematics, not opinion and dogma.

References

1 Capacitor ESR Ratings, Transtronics
2 Improved Spice Models of Aluminum Electrolytic Capacitors for Inverter Applications, Sam G. Parler, Jr.

Cornell Dubilier
3 Capacitor Sound - Jul, Sep, Oct, Nov, Dec 2002, Jan 2003 editions of Electronics World (formerly Wireless World), Cyril Bateman
4 Inductance Of A Straight Wire, Tim Healy, Santa Clara University
5 Self resonant frequency of a capacitor, Ivor Catt
6 Understand Capacitor Soakage to Optimize Analog Systems, Bob Pease, National Semiconductor

---

Other Useful Links ...

Introduction to Capacitors, Capsite 2005 (Highly Recommended)
Characteristics of Capacitors, Murata
Straight wire inductance calculator, Research Solutions & Resources
Dielectric Constant Reference Guide, Clipper Controls

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 24 Sep 2005
(US Spellings are included to assist in searching ... metallization, aluminum, polarized)

Elliott Sound Products **CD vs. SACD vs. DVD-A**

by Niklas Ladberg

Introduction

For a while I have been interested to find out how good the new high resolution formats Super Audio CD and DVD-Audio really are? I have visited some demonstrations, but not been impressed which may be due to other causes than the sound formats.

After reading many replies at different forums, it seemed like many audiophiles considered SACD to be better than DVD-Audio. I also thought so until recently read what Ing. Öhman wrote in the Swedish Audio Technical Society * journal.

(*A non-profit organisation for sharing interest and knowledge in audio and sound reproduction)

The following are quoted from what Ing. Öhman wrote in the journal:

"It is nothing less than a tragedy that Sony/Philips system SACD still is considered to be a real competitor to DVD-A, though it has lower real resolution than the CD-system in the highest octave.

DVD-A does absolutely offer a much higher dynamic range than CD, but it is very questionable if SACD does.

SACD is in the high frequency range quite mediocre, even compared to a good CD-system one-bit DAC, and of course clearly inferior to a CD-player with a real multi-bit converter.

On the contrary, DVD-A is in theory 250 times better than the CD-system at all frequencies!

In today's reality though, it is hard to achieve such hyper-resolution, but maybe in the future? If the potential exists, recording and playback technology can evolve. Today the DVD-A resolution is about 16 times better than the CD-system and the bandwidth extends up to 100 kHz to be compared with 22,050 Hz for CD."

Now I became curious! This is the opposite of what I thought. I asked Öhman for a follow up ...

Niklas Ladberg: How did you come up with these conclusions?

Ing. Öhman: DSD (the coding technique used in SACD) is much better than CD in the low frequency range. The problems occur at higher frequencies. The noise level in the ultrasound register is more than 100 dB higher (-40dB under maximum output level, using narrow band analysis) when compared to DVD-A (-144dB under maximum output level, full spectrum noise).

Another way to describe the difference: The noise [power] from SACD is more than 20,000 million times higher than from DVD-A!

But maybe it is more relevant to know that this ultrasound noise from SACD is enough to warm up the tweeters voice coil with some detectable influence on reproduced sound. Besides, the ultrasonic may also affect the audible sound by down mixing in the air, at least at higher sound pressures.

A comparison with CD is harder because of the limited bandwidth of the CD-system. Signal to noise ratios in the range above 22,05 kHz can therefore not be determined, but noise level from CD can be as good as DVD-A – one can always use low pass filter! Then no ultrasound comes out from the CD-player.

The problem with SACD can be shown by theoretical calculations, measurements and by listening. I have done lots of all these three and every one of them points clearly in the same direction: SACD has not more resolution than CD above 10kHz. Our early estimations some years ago have now been confirmed by measurements and listening, made both by us and others. Today, many studies have been done, for example by Stereophile who has tested SACD players several times and confirmed our estimations.

NL: What does Stereophile say about the limitations of SACD?

IÖ: Oh, not much in the written tests. Stereophile is an advertisement dependant paper and they are a bit careful to say anything negative about anything, but they present their measurements, often done by the sharp editor in chief John Atkinsson. His comments are very informative! Although very careful, any equally sharp reader can decipher Atkinson's opinion from the text.

Stereophile also performed a large subjective listening test between some existing high performance recording formats and SACD was even considered to be inferior to PCM 16bit, 176,4 kHz.

Anyway, one should not take too much notice of "how many" said this or that. What is important is how things really are, not how many people believe this or that. Deciding truths by voting is seldom a good way. It can only show what people believe. Some peoples' beliefs might of course be right, but too often correct information and relevant listening experiences drowns in common misconceptions, preconceived notions and the media background noise.

This happens especially easy in this particular case since DVD-A and SACD never have been compared under equal conditions.

NL: I read in a technical paper about SACD. Sony writes:

> ***"The vast majority of A/D-converters used in PCM recording for conventional CD are 1-bit converters with high sampling frequency"***

Then it must be better to keep the signal in DSD-format (SACD) than convert the signal to PCM (CD/DVD-A)?

IÖ: Yes, that is correct, but is it relevant question? The CD-system seems worse than it is when used with a one-bit converter. The problem is not the conversion from a one-bit converter to PCM but the one-bit converter itself!

Besides, the one-bit converters used in CD-players usually have higher resolution than DSD, which only samples 64 times faster than CD-system sample rate (i.e. DSD sampling rate = 2.8 MHz). The low sampling rate in DSD is used because of the systems ineffective coding and lack of storage space. By packing the information it becomes a bit more effective but it still is ineffective compared to PCM.

One-bit converters for CD-players often use sampling rates between 11 and 50 MHz. The best one-bit converter probably is JVC's PEM-DD and it is much better than DSD.

This said with reservation, I might have missed some even better one-bit technology than PEM-DD. But as far as I know this is the technology that comes closest to true multi bit technology in resolution.

NL: I presumed that SACD uses DSD-technique for recording and mastering, but that turned out to be wrong:

> *Press release:* ***"A big surprise at the AES was the confirmation by Sony that DSD technique, used in SACD, uses multi-bit PCM during recording and mastering processes and that only uses one-bit technique as it applies to consumer playback systems. Jim Johnston of AT&T Research speculated that DSD and DVD-A data streams might be able to co-exist if output from different points within the same microprocessor."***

Apparently they use PCM for recording and mastering, even for SACD. Now the advantage of no conversion between formats suddenly disappears.

IÖ: Several documents show this is the case. Sony/ Philips has even officially recommended using PCM when editing the recorded material. I think it is a wise recommendation, because every manoeuvre in the PCM-domain is straightforward, easy to make and will not degrade the quality if performed with high enough resolution. Only the DSD-problems remain!

Therefore DVD-A is a purer and more straightforward system. No conversion between different formats and 144dB resolution at all frequencies up to 100kHz. Could it be better? Well of course it can, it can always be better, but DVD-A is good enough. DVD-A is what the CD-system should have been from the beginning!

NL: Back to the Sony technical paper:

> ***"On the playback side, most CD-players utilise one-bit D/A-converter to convert digital signals back to analogue."***

Is this correct? Don't most CD-players utilise multi bit-converters?

IÖ: Though some fine multi bit CD-players exists, unfortunately most CD-players today utilise one-bit converters. It is probably a price question. A so-called 24 bit one-bit converter (working with one bit technology inside but accept 24 bit input) costs about 2-4 dollars including 2 channels and digital filter.

A real 24 bit converter with 96 kHz sampling frequency and 8 times over sampling costs

about 10-15 dollars per channel – without the digital filter. For two channels and digital filter it ends up to approximately 40 dollars.

So multi bit technology is ten times more expensive as one-bit technology. Most manufacturers find it easy to choose ... Especially since Hi Fi-magazines and Hi Fi-stores never take a stand and point out the difference.

NL: From Sony again:

> ***"Although the bit numbers is just 1/16 of that used for the CD-format, the sampling frequency is 64 times higher. DSD can accommodate more than 4 times the information as the current format."***

4 times more information could not be wrong? Or is the depth of bits more important?

IÖ: The number of the bits is much more important. Actually the DSD-system is theoretically less dense in information than the CD-system. Even when the data is packed (as it is on an SACD) it is still not much better.

When Sony declares that DSD-format can store 4 times the data, they probably mean that there is 4 times the space on the SACD-disc compared to CDs. But since the DSD-coding is so ineffective, the real information is considerable lower.

The resolution/ information doubles when you double the sampling frequency (it is possible to be more specific, but for this example it is enough). But to double the resolution using PCM, you only have to add one more bit. If you go from 1 to 16 bits (adding 15 bits which use approximately 15 times more storage space), the resolution increases 65,536 times (from one step to 65,536 steps).

There is also another essential difference; the increase in resolution you achieve from raising the sampling frequency will be frequency dependant. A one-bit system will therefore have high resolution at low frequencies (where the information theoretically is low) and have low resolution at high frequencies (where the information theoretically is high).

By the use of noise shaping of high order, it is possible to increase the resolution at "quite high frequencies" at the expense of resolution at very high frequencies, but *only* for static, non transient signals. Transient signals will have poor resolution in a one-bit system. If the signal does not endure for a long enough time, the error will not be minimised by the noise shaper of the one bit system.

That's why you can read in documents from Burr Brown (who manufactures both one-bit and multi-bit converters) that you should use multi-bit converters for "*waveform synthesis applications requiring very low distortion and noise*". They have not written this for nothing.

A one-bit converter (i.e. the DSD system) cannot regenerate a short pulse with stringent form. It will change form from moment to moment. Every identical recorded pulse will show up with a new form.

One can always discuss the audibility of such behaviour, and if it is audible, one can discuss how much it disturbs. Objectively good reproduction is not important for everyone. So if the presentation of music is changed in some way, some people might see this as a minor problem, others think it is more serious.

It is beyond discussion that the lower resolution of one-bit systems is a problem under circumstances more precision demanding than audio. When you need both super precision and stability, when you need to know that a generated waveform looks like it is suppose to, nothing but multi-bit converters will do.

My entirely personal opinion/experience is that audio actually demands very high precision, and that the reproduction suffers from the lack of precision from one-bit converters and the DSD-system.

Perhaps I even prefer the old 13-bit PCM-based Denon system from the 70s. It was not a super high-resolution system, but it was as stable as it was consequent! I have a lot of these recordings (reissued on CD) and they actually sound fabulous!

NL: From Sony's paper ...

> ***"In general, the quantisation noise floor resulting from PCM is flat, according to the number of bits. With Delta Sigma modulation the noise floor is subjected to noise shaping. Because the DSD method uses a high sampling frequency, the quantisation noise is shifted to a higher frequency range. This reduces the amount of noise in the audible range for humans, which is relatively low."***

Is this correct?

IÖ: Yes, the first part, but one should be careful not to underestimate the hearing, as Sony does – or you end up with systems like SACD! Of course we can only hear single tones up to 20kHz but this does not implicate that we can allow any kind of noise

pollution above this frequency. Multiple tones in the ultrasound can create clearly audible phenomenon at higher sound pressures.

It can be discussed if we can tolerate the ultrasound noise generated from SACD. The noise from SACD just above 100kHz is higher in level than most of the treble in the audible range, at least when listening to the majority of acoustical music. It can also be discussed if DSD uses a "high" sampling rate. But apart from that: Yes, without noise shaping it will not work at all and that would be a lot worse. Now it is only a little bit worse than CD in the highest treble.

But why introduce a new super high-resolution system, that is "a little bit worse" than CD? Of course there are advantages when compared to CD also, but the drawbacks of SACD/DSD are completely unnecessary. Shouldn't a new system be better than CD in all aspects?

The noise level in the range above 100kHz is –40dB under maximum signal level (and is thus even visible on an oscilloscope!). The noise is in fact much higher than any possible music signal in the same frequency range. This can be compared to DVD-A where the noise level is –144dB in the whole audible range and also in the ultrasound range.

In reality these figures have not yet been reached in commercially available DVD-audio players, but the potential for future improvements is there. Today, the best DVD-A players reach a signal to noise ratio of about 120dB (0 – 100 kHz). The figure will be even higher when measured with small band analysis. This is far from 144 dB, but still very good.

I think it is a little embarrassing that no good DVD-A recordings have been released so far. They have all been of inferior quality. This has of course not made it easier for ordinary people to make a relevant opinion about the differences in the systems. The SACD recordings from Jan-Erik Persson at Opus3 are far superior everything released from DVD-A.

Now, when looking at "recordings in holistic view", the debate of the storing media/ system is much less important than the skills of the recording engineer – a good recording stored on compact cassette is far superior to a bad recording stored on DVD-A. But every debate has its time and place. It is not relevant to point out more important issues when we are about to choose a new storage media.

This time the storage media (SACD versus DVD-A) is the question, and it can very easily be distorted by comparing the systems on different recordings. The better

recording will always win.

The music is even more important of course, than the recording, but it is entirely subjective – not much to debate about. As an example; my taste in music is disputably deplorable. At least according to some people!

NL: I continue to read from the press release:

> *http://tinpan.fortunecity.commarrfield/216/pag_eng/oct2000.htm)*
>
> ***"The watermarking issue is having a big impact in the 109th AES convention at L.A (for the moment only at professional level, but in a future also at a domestically level). A big and increasing number of professionals have pointed out the fact that the actual methods of watermarking introduce sound quality degradation into the supposed high quality recording formats. While the watermarking promotion groups are pushing hard these methods, a group of very well known audio firms (Chesky records etc) and well-known engineers (Tony Faulkner of GRP) are very disappointed of what they consider an audible degradation of the sound quality. The first tests conducted recently in UK and USA (btw, not as impartial as supposed to be due to the use of ultra low quality recordings) using DVD-audio with the watermarking, have shown the fact of this degradation."***

Is the watermarking of SACD sound degrading?

IÖ: Well that depends on which watermarking is used! The so called "psychoacoustical water markings" are all sound degrading in various degree and are not associated with any specific storage system.

You can use it with any system of your choice, but it is unclear to what extent they are used in respective system. As far as I know, no released SACD recordings have used this type of watermarking, but I can be wrong.

The only thing I know for sure is that lots of CD-recordings have this kind of watermarking. But watermarking is a separate problem. It has nothing to do with the either of the recording and storage systems. You can choose not to use it. But the problems built in the DSD-system cannot be excluded!

Just to avoid misunderstandings I want to make a reminder that SACD has higher resolution below 5-10kHz than the CD-system. Exactly where the limit is, where each system (CD or SACD) is better, depends on if you are looking at a static or dynamic signal.

At frequencies below 100-600Hz the SACD-system could theoretically be even better

than DVD-Audio, but in reality this is not important. We are talking about so small flaws, far below the hearing threshold, so they can be disregarded. Any specific player however, can be very bad at low frequencies, but not due to the system if SACD or DVD-A is used.

Anyway, in the practical life, it seems like DVD-A wins over SACD at all frequencies, cleaner sound, lower noise, and a completely stable system, free from potential noise shaping algorithmic oscillations.

NL: How close to the potential resolution do we come with the player existing today?

IÖ: If we look at the players you can buy today, then the resolution is about 15-30 times lower than the full potential of DVD-Audio. That means 8-16 times better than the CD-system. A big improvement, but still a lot more is possible.

All these comparisons between systems made in "times" are based on pure technical specifications. If you want to have "worst case" audibility, take the logarithm and multiply it by 20, then you get the figures in dB. "15 times better" corresponds to a 23.5 dB lower noise level.

If these figures should be of any use for you, you must have a feeling for "how it sounds when an error signal changes X dB below the music signal". If the changes of the error signal take place below the threshold of hearing, naturally you will hear no difference at all.

Also, different errors have different audibility. Some are easier to detect. Others are more difficult. Some faults/errors are therefore more tolerable even though they are higher in level.

Who said it is supposed to be easy to form relevant opinions?

Relating technical performance to the experienced sound quality should be done with utmost care. My experience is that the quality improvements coming from using more bits are valuable only for high dynamic music material. An 11-bit recording will do fine for extremely compressed "radio-sound".

How people experience sounds in the ultra sound region seems to be quite individual. The most sensitive listeners seem to manage music better with no ultrasounds/ overtones at all, than with music plus ultrasound pollution. Probably no one minds a nicely reproduced ultra sound range (like in DVD-A), but when using such a very capable system, there is no guarantee that recording engineers manage to keep the

ultra sound "unpolluted".

When it comes to comparing CD, DVD-A and SACD against each another, it becomes much easier, because at optimum implementation, all error signals in these systems are in the form of noise. But you have to look at every frequency register alone, since the noise character of each format displays different spectral behaviour. I mention this again as I might not have been clear enough earlier in the text.

NL: Thank you for your time and for answering my questions!

IÖ: Hope I could straighten out some of your question marks.

Editor's Note

This article is reproduced with the kind permission of Niklas Ladberg and Ing. Öhman, with translation from Swedish into English by Marten Kihlberg. No changes have been made to the text other than by way of translation, and the actual words used herein are as accurate as translation allows.

The editorial "content" has been limited to the creation the web page (in standard ESP format) from the original article, a small number of grammatical corrections and the addition of one word [power], when referring to the noise level differences between SACD and DVD-A (i.e. 20,000 million is the power difference, not voltage difference).

I would like to thank Niklas Ladberg, Ing Öhman and Marten Kihlberg for making this information available.

Main Index

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Niklas Ladberg, Ing. Öhman, Marten Kihlberg and Rod Elliott, and is Copyright © 2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The authors and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from all parties named above.

Page created 27 Aug 2002. All rights reserved

The Audio Pages

Elliott Sound Products **Compression in Audio Recordings**

Copyright © 2001 - Rod Elliott (ESP)
Page Updated 05 Mar 2005

Articles Index

Main Index

Contents

- Introduction
- Compression Versus Limiting
- Additional Features
- Why Use Compression
- Contra Indications
- Peak to Average Ratio & Dynamic Range
- Typical Specifications
- What To Listen For
- What About Lossy Data Compression?
- Even Worse Things That Can Happen
- References

Introduction

The term compression has several meanings in audio - there is lossy data compression (e.g. MP3), lossless data compression (a wave file compressed using a Zip program), and level compression and/or limiting. In this article, I concentrate on the last form, although a few words about MP3 are in order (later).

I recently read a wonderful article by a mastering engineer by the name of Bob Katz (see references, below). Bob was adamant that many producers, engineers and musicians have joined a new race to see who's CD can be the "hottest" (i.e. loudest). There is a mistaken belief that this makes the sound more exciting - it doesn't - it makes it boring, and very tedious to listen to ...

Just like commercials on TV, which are compressed to within an inch of their lives. Does anyone find the sound satisfying? I'd be very surprised to hear someone (other than the producer) say yes - they are annoying, seem to be much louder than the programme you were enjoying, and cause a great many people to hit the mute button on the remote the instant they start (I'm one of them :-)

Compression Versus Limiting

So what is the difference between these two effects? Depending on your outlook, either not much or a great deal. A limiter is usually set to a fixed threshold, and any signal that attempts to exceed the threshold is pulled back (attenuated) by exactly that amount needed to maintain the predetermined level. If the input gain is set way high, then all signals below the threshold (including noise) are boosted - in the extreme case so everything is the same volume (again including noise!). Limiters are "hard" compressors - the absolute level is fixed, and the compression ratio may be as high as 100:1 or more. This means that the input signal must increase by 100 "units" to make the output increase by one unit. Many limiters claim that the ultimate compression ratio is infinity, however this is probably an over estimation of the true figure.

A compressor uses much the same (or at least similar) circuitry as a limiter. While some compressors boost the level of signals below a preset threshold by a predetermined amount and reduce the level of signals above the same threshold, this type of compressor is most commonly used in noise reduction systems - an expander is used at the playback end to return the original dynamic range.

The majority of compressors use a threshold setting (like a limiter), and reduce the gain progressively once this is exceeded. Compression ratios of perhaps 2:1 are common, so the output will rise (or fall) by one unit for every two units of input change. A 50dB dynamic range (above threshold) is therefore reduced to 25dB (from softest to loudest signal). Unlike limiting, the compression threshold is typically set lower than the peak level - the actual threshold level could be anything from +8dB to -40dB, depending on the effect desired.

For example, a guitar (after suitable amplification) may produce transients of perhaps 5V peak, but yield an average level of only 500mV. This represents 10:1 or 10dB peak to average ratio. The peaks are produced as the pick strikes (or releases if you prefer) the strings, and the average is predominantly the normal decay of the note before the next is played. The VU meter (I refer here to real ones, not the stupid things you so often see that bear no resemblance to a proper VU meter) gives a good indication of the average level, and therefore the perceived loudness (VU = Volume Unit).

A PPM (Peak Programme Meter) shows the peak levels - no surprise there. Some meters, mainly electronic versions, provide both indications. A bar shows the average (VU) level, and a dot that "sticks" at some higher level shows the peak amplitude.

By using compression, the same guitar may have the maximum level reduced to perhaps 1V - the average level will now be higher as well (softer sounds are amplified, loud ones attenuated). Peak to average ratio may be reduced to 6dB or less, and the note will seem to just hang on forever ... well not quite, but you get the idea.

This sort of compression is common on percussion, strings, vocals - in fact almost anything. It is appropriate if (and only if) it provides the *sound* the artist wants - when compression is used just to make something sound louder, then it is better to just turn up the volume. This way, the original dynamics are preserved. Incorrectly used, compressor/limiters will flatten the sound, and remove the life and soul of the music.

Additional Features

Compressor/limiters are usually fairly complex electronically. Since they already have a voltage controlled amplifier (VCA) circuit that must be of the highest quality to satisfy audio professionals, this can be used for other things as well, with little real increase in cost or complexity (they are already complex, so a little more won't hurt :-)

A common addition to these audio tools is a noise gate. This is provided with compressor/limiters, and is used to gate (or switch off) any signal below a preset minimum. Noise gates are used to remove unwanted low level signals, but are sometimes used to mess up the sound completely by removing the ambience. Better a little noise and a complete sound than a quietly decaying ambience that suddenly just stops. Used properly, a noise gate can seem to eliminate background hiss completely, while letting the signal through (the hiss is still there, but you can't hear it when the signal is present). Used improperly, the initial parts of sounds are cut off, and the natural decay is not present. This is (fortunately) rare in pro studios.

One final feature offered in many units is a "de-esser". The sibilants ("ssss" sounds) in vocals are often over emphasised by close microphone placement, mic characteristics, the vocalist, or equalisation - and often a combination of these. This can be very unpleasant, so the de-esser does exactly what its name implies - it reduces the sibilant sounds by an amount that the recording engineer can set according to need (or taste)

Why Use Compression?

Compressors and limiters are used in music for a multitude of reasons. The first (and should be the only) reason is for the sound. Used properly, a compressor - or more correctly a limiter - will place an absolute cap on the maximum level that can be passed. This is invaluable for preventing a large PA system from distorting, or making certain that the ADC (Analogue to Digital Converter) does not clip (exceed the maximum conversion voltage). Digital distortion is extremely unpleasant, and is to be avoided, as with all forms of hard clipping.

There are many other reasons to use compression or limiting. Many instruments do not have the sustain that the musician desires, and this can be corrected by using a compressor to extend the note. As the signal fades, the compressor increases its gain, so the note lasts longer.

Another reason is to restrict the dynamic range. Movie soundtracks are a prime example. If the maximum level of a car bomb exploding or a shotgun fired at close range were to be reproduced, and all conversations were at the normal level, no-one in the theatre would hear anything that was said, or would be deafened instantly by the explosions. By reducing the dynamic range, both can be accommodated at levels that are appropriate, but limited to an acceptable maximum and minimum loudness.

By contrast, many trailers and theatre advertisements are heavily compressed - they have a consistent loudness that is greater than that of the main feature. This technique can work for a limited period (it gets your attention), but becomes very tiring very quickly.

The over use of compression results in a flat, lifeless reproduction. In his article, Bob Katz refers to "wimpy loud sound", and at some stage we've all heard it. You put on a CD, and it is LOUD, so you turn it down, so the loud parts don't leave you loudspeaker cones on the floor. You wait, you listen, you wait some more ... it never happens! There are no loud sections! There are no quiet sections. Everything is at the same volume from beginning to end, and the result is indeed wimpy. Certainly the CD is louder than others

you own, but it is the same volume from start to finish and leaves you as flat as the sound.

How to make music bereft of life - compress it until it bleeds (to death). No hi-fi, regardless of cost or sophistication can make rubbish like that sound good, since there is virtually nothing that can be done. An expander (essentially the opposite of a compressor) may restore some vestige of what the artist intended, but compression is not easily undone unless you can obtain the exact reverse of the original settings - this is both difficult and time consuming to even attempt, assuming that you have the equipment in the first place (very few hi-fi systems incorporate an expander, so most of us are well and truly screwed).

Contra Indications

Compression is commonly used in the final mix, and this is where things can go seriously wrong - everything is at the same volume, peak to average ratio is minimal, and the resulting sound is *almost always* worse than it was before the compression was applied. Used correctly, a small amount of compression may be useful with some musical styles, but it is completely unsuited to others. I have several CDs that sound "exciting" at first, but the sameness of having a constant barrage of sound at the same level becomes extremely fatiguing in only a short time. On some, I can hear the compressor/limiter acting ("breathing" or "pumping" are terms commonly used for this effect), which means that it has been over used, and the CD is then relegated to the "don't bother listening to this" pile. Most unfortunately, this pile is getting bigger, and many of the modern CDs are worse than older ones because of the stupid, unnecessary and pointless game of one-upmanship by the record companies, all trying to get the "hottest" CD on the block.

I don't want it to be "hot", I want it to sound the way it should, with real dynamics, soft and loud passages, and things that make me jump! Fortunately, I am not alone, but unfortunately, record companies are still producing material for people with crap systems - "Make it sound good on a crap system - we'll sell more". This is rubbish - people with only a boom box don't care that much, otherwise they would strive for something better. Those among us with good or excellent systems should not have to listen to something that was mixed on a pair of near field monitors with the quality of a transistor radio, and compressed so heavily that it has lost all of the dynamics that make music what it is - or should be!

If the CD is a little quieter than expected, then I simply turn up the volume - even without

a remote control, this is hardly an arduous task. Better that than have a whole pile of "hot" CDs that I can't bear to listen to because they have had all their life removed by an over zealous compressor-head.

Peak to Average Ratio & Dynamic Range

All music has a peak to average ratio (and dynamic range - see below), since there are peaks and dips in the level (even when heavily compressed), and the average level must be lower than the peaks. The trick is to know what the peak to average ratio should be. It is commonly quoted as being between 10dB and 20dB (a power ratio of between 10:1 and 100:1). By this reasoning, music with a 10dB P-A ratio will need perhaps 50W to handle the peaks, but will provide an average power of only 5W. This is typical of a lot of music, and even some orchestral music will be at this ratio without any compression (relatively uncommon, but Bob Katz has experienced exactly this).

A ratio of 20dB is at the other extreme, so the same 50W amplifier will only produce an average power of 0.5W - this is where the use of high powered amplifiers for hi-fi is important - by the time the average power is high enough, the peak power is massive. A quick example ...

You want to listen to music at 90dB (SPL). Your speakers are rated at (say for convenience) 90dB/m/W, so with two of them, the effective sound pressure (at 1 metre) is 93dB SPL with 1W into each channel. You (the listener) are some distance away, so the level may be 3 to 12dB lower at the listening position. We shall assume 6dB as a reasonable guess for a typical listening room (although it may be considerably more than that, depending on room treatment, furnishings, etc.).

For 1W per channel, your SPL will be about 87dB SPL, so to get the extra 3dB, the power must be doubled to 2W per channel. If you have music with a peak to average ratio of 20dB, you will need 200W per channel to reproduce the music without distortion - assuming that you have that much power, the peak SPL will be in the order of 110dB ...

(90dB + 20dB P-A ratio = seriously loud).

Compression is your friend! Such a high P-A ratio will cause most high end systems (and their owner's ears) grief at high levels, so some degree of compression will make the reproduction less arduous on your system (and a lot less likely to frighten the cat to the point where it perches tightly on your head :-) The magic is to find the ratio that keeps the ratio to a reasonable figure (and there are no absolutes here!), while

preserving the soul of the music. It can be done, and I have many CDs and vinyl albums that do it very well (Bob will also tell anyone who cares to listen how to do it well, too).

It can also be done very badly, and so many new releases do just that. Mind you, a lot of old releases were just as bad - this is not a new phenomenon, but has been happening ever since the compressor was first invented - or at least used in anger.

Dynamic Range

Is there a difference between peak to average ratio and dynamic range? The answer depends entirely how long the averaging period is. Generally, the two are considered separate, but with enough compression they become equivalent. The dynamic range of a piece of music is the difference between the softest passage and the loudest - it takes little imagination to realise that if it is sufficiently heavily compressed there will be no difference at all.

With a good mix, and a compressor/limiter that is correctly adjusted, the difference between the two will be less than in "real life", but great enough to create excitement - this is where experience and careful adjustment come in. The range of sounds we can hear (and will be assaulted by) is enormous, from the faint rustling of leaves on a very quiet night through to jack-hammers or jet planes (and certain motorcycles!). To expect to reproduce this range from a home hi-fi or theatre system is generally not possible and is undesirable anyway.

The difference between the two is blurred - there is an almost infinite grey-scale, rather than any black and white distinction between the two, so I shall leave it at that.

Typical Specifications

A discussion of compressor/limiters would be incomplete without a brief explanation of the features and controls typically offered. A typical unit may offer the following (adapted from real specs for a typical unit) ...

Max. Input Level:	10V RMS (+22dBu)
Dynamic Range:	118dB
Signal to Noise Ratio:	>100dB
Headroom:	18dB <0.05% at +4dBu, with 6dB compression
Distortion:	<0.05% at +4dBu, with 6dB compression

Limiting Threshold:	-40dBu to +20dBu
Attack Time:	0.1ms - 200ms
Release Time:	50ms - 3s
Compression Ratio:	1:1 - 100:1 with selectable hard or soft compression knee

Some of these terms are self-explanatory, while others may need a little more information.

- **Maximum Input Level**
  This is the maximum signal (in Volts RMS and/or dBm or dBu **). If this is exceeded, the input stage will distort due to clipping. This indicates that a unit with the above specification cannot be used for a direct speaker feed (few can, and a Direct Injection box is nearly always needed).

- **Dynamic Range**
  The range of the quietest sound (signal) to the loudest. The spec above is misleading, since it ranges from the noise floor to the maximum input level. The effective dynamic range is limited by the minimum signal to noise ratio the engineer will accept, so subtracting (say) 50dB from the above would be reasonable.

- **Signal to Noise Ratio**
  The level of noise referred to a specified output level. Again, the output level is not specified above, so we don't know if the actual noise floor is better than -100dBu or -78dBu, since the reference input or output level was not defined. Many manufacturers use A-Weighting for these measurements, which is also misleading for most applications. A-Weighting should be used only where the sound source is some distance from the listener (possibly hundreds of metres), or is expected never to exceed about 70dB SPL near field.

- **Headroom**
  This can be tricky, since there are many different ways to describe it. Most commonly, headroom is the difference (in dB) between the typical or specified output level, and the maximum the unit can provide without distortion.

- **Distortion**
  Another can of worms! Distortion measurements are not really meaningful if the level and frequency are not stated. It is even harder with anything that uses a VCA, since their distortion can vary considerably depending on the degree of amplification or attenuation. Distortion (probably) should be worst case - the frequency, amplitude and amount of gain that produces the most distortion. This would probably look bad in the specs, so is not used.

- **Limiting Threshold**
  This is the range of signal level where the unit can be set to function. Below the threshold, a limiter does nothing - it just passes the signal straight through (with maybe a little fixed gain or loss, as appropriate). Above the threshold, signals are attenuated by an amount determined by the compression ratio (see below).

- **Attack Time**
  A measure of how long it takes before the unit changes the gain when a signal is applied. This may be very fast to prevent *anything* from exceeding the threshold, or deliberately slowed down to allow a percussive attack to the instrument. A relatively slow attack might be used with drums to increase the apparent dynamics, whilst actually reducing them.

- **Release Time**
  The release time determines how long the gain takes to return to "normal" after it has been reduced. A short release squashes the dynamics of the sound completely, and a very long release time ensures that all material remains at a constant fixed peak level, while still allowing normal variations. The overall dynamics (of a musical piece) are still compressed, since the soft passages will allow the unit to apply the maximum gain. This is all controllable.

- **Compression Ratio**
  As discussed above, this can be low (1:1 means no compression), almost infinite, or anywhere in between. An infinite (or high) ratio is a limiter, anything between 1:1 and about 10:1 is a compressor, although some would argue (correctly) that 10:1 is really limiting. Some units also offer a "soft" or "hard" knee (the threshold). A soft knee means that the onset of compression or limiting is gradual, ranging over a few dB, while a hard knee means that once the threshold is reached, the limiting action is immediate, with no gradual onset.

** dBm is a reference level based on 1mW into 600 ohms. This represents a voltage of about 775mV. dBu is based in a reference level of 1V RMS (also sometimes seen as dBV).

What To Listen For

An excellent way to hear compressor/limiters in action is an outside broadcast on TV. While the presenter is speaking, the level is constant, with very little variation - even when they are off-axis from the microphone. When there is a break in the commentary, the background noise can be heard to increase at a (relatively) fixed rate, until it is as loud as the presenter's voice, or someone starts speaking again. It is then instantly reduced to where it was before.

As you get used to what to listen for, you will hear many CDs where the level of the backing track falls when the singer (or someone else) starts making their noises - this (to many audio professionals) is quite normal, but it is not - it is a typical case of over application of compression in the final mix. When an additional sound is added to those already present, it is *supposed* to get louder - this is called dynamics (or even "micro-dynamics" - a reduced scale version of the real thing).

These effects are especially noticeable with commercials on radio or TV - listen for them

so the sound can be identified. Should you purchase a CD that does the same, complain to the record company - they have ruined your music!

What About Lossy Data Compression?

MP3 - love it or hate it, it is here (probably) to stay. As can be determined from a multitude of sources, MPEG Layer 3 (or MP3 for short) discards information that theory (and a lot of experimental testing) indicated would be inaudible. It uses a well known characteristic of our hearing called "masking", where it is known (and can be proven) that certain frequencies and levels are completely inaudible when accompanied by another signal at a higher level. The points where masking take effect are beyond the scope of this article, but differ according to relative levels and frequency, and the frequency band itself. An MP3 encoder breaks the signals into sub-bands using filters, and each is treated differently according to a set of rules built into the encoder signal processor.

While it is generally considered that a high bit rate (128kb/s or above) MP3 track is of "near CD quality", many people will dispute this vehemently. My own experiments and listening tests indicate that imaging is poor, and the precise placement of instruments and vocals is missing. Some instruments - especially the harpsichord (so I hear) sound completely different when encoded, almost regardless of the bit rate. A good test is to "rip" some pink noise (preferably generated by an analogue source), and compare the difference.

There should be no difference - or at least it should be inaudible, but this is not the case! At 320kb/s the difference is barely audible - one has to listen carefully to hear it, or figure out just *what* to listen for ... but there is a difference, and it also shows up very clearly on the analyser of Winamp. The peaks are flattened, so the dynamic range (or peak to average ratio) is degraded, and the sound of the noise lacks "life" compared to the original recording.

If we can hear a difference with noise, why would music be any different? It isn't! Ok, noise has a relatively constant bandwidth (DC to daylight in the extreme), and excites all frequencies more or less simultaneously. Well, so does a lot of music, albeit for short periods at a time.

Will my comments here make MP3 go away? Of course not, and nor should it go away, because it is a useful way to archive recordings, or provide people (who insist on *not* hearing approaching traffic while they run or cycle) a convenient medium for portable

sound.

Even Worse Things That Can Happen

On a final note (pardon the pun :-) a reader recently sent me an out-take of a CD. It was clipped! Not just compressed and limited to the maximum (that too), but with actual clipping - flat tops on some peaks. I asked him to check his setup very carefully to ensure that the record level was not set too high, and he assured me that he had at least 3dB of headroom above the peak CD level.

At some stage, I shall check some of the CDs I have that annoy me because of the constant loudness to see if they have the same problem. Probably not, since the clipped CD was from an "indie" (independent producer) so would not have had the controls in place one would expect from an established mastering house.

Be that as it may, there is not really much point in setting up the "ultimate" hi-fi system, with headroom to spare and almost zero distortions of any kind, only to have the music CD pre-distorted, compressed, limited and bent so far out of shape that it is no longer useful for anything other than a coaster.

References

- Compression - A most excellent article by Bob Katz, on the perils of using compression or limiting on the final mix. As Bob says - leave this to the mastering engineer, and if you don't like the result - complain!
- Some examples of specifications were extracted from Alesis and Behringer information - no material from either manufacturer has been directly used or copied - merely representative information on performance and functionality.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 21 Dec 2001./ Updated 05 Mar 2005

Elliott Sound Products **Earthing Your Hi-Fi - Tricks and Techniques**

Copyright (c) 1999 - Rod Elliott (ESP)
Page Created 30 Dec 1999

Articles Index

Main Index

Introduction

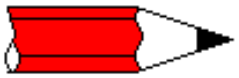
It is not uncommon to see hi-fi equipment with the earth (ground) disconnected from one piece of equipment or another, usually to prevent a hum loop which ruins the listening experience. However, there is nothing quite like being electrocuted to *really* ruin the experience should something go wrong!

Electrocution

First, a quick word on electrocution. It is not fun, and electricity kills a great many people worldwide every year. A current of 50mA (barely enough to make a low wattage lamp even glow) is sufficient to send your heart into a state called "ventricular fibrillation", where the heart muscles are all working out of synchronisation with each other. Little or no blood is pumped, and you will die within about 3 minutes unless help is immediately at hand.

Sometimes (but less often), your heart will simply stop. If this happens, it is possible that with external heart massage that it might re-start, and occasionally it might even re-start by itself - rare, but it can happen. The result of fatal electrocution is that you will no longer be able to enjoy the hi-fi that you have spent so much time and money putting together, and all other earthly activities are similarly curtailed :-(

This article has been prompted by many e-mails I have received asking about hum, earthing and what should be done to ensure that equipment is safe, and does not hum. There are other causes of hum in a sound system other than electrical (safety) earthing issues, but I will contain this particular article to the basic issues of safety and eliminating hum loops while maintaining a high degree of safety.



The regulations regarding safety earth connections vary from one country to the next, and I do not have the details for each case. This article is general, and if unsure, you should consult the appropriate electrical supply authority in your country to obtain the rules that apply to you.

I have used the terminology I am familiar with - the live conductor is called the 'active' and the return conductor is called 'neutral'. The safety conductor is called 'earth' or sometimes 'earth ground' (particularly in the US). These terms differ, so make sure that you know what they are called where you live.

How The Safety Earth Works

The basic idea of an electrical safety earth (or ground) are pretty much the same everywhere. The case (chassis) of the equipment (and except for special situations, the internal electronics) is connected to an earth pin on the mains outlet. This is then connected through the house wiring and switchboard to an electrically solid earth point, which is commonly a (copper) water pipe, or a stake buried deep into the ground.

Should a fault develop within the equipment that causes the active (live) conductor to come into contact with the chassis, the fault current will flow to earth, and the equipment or main switchboard fuse or circuit breaker will blow. This protects the user from electric shock, bypassing the dangerous current directly to earth, rather than through the body of the unsuspecting poor bastard who just touched it. If this experience does not kill, it will invariably enrich the vocabulary, but regrettably the new words will be useless in polite company.

Earth leakage circuit breakers (RCD - residual current detectors - see below) measure the current in the active and neutral conductors. If these differ by more than a few milliamps, the circuit is disconnected. The principle is simple - if the current in the two wires differs, some of it must be going somewhere that is undesirable, so the supply is interrupted almost instantly. While these are mandatory in some countries (or under some circumstances), it is best not to rely on any advanced technique, but provide a system that is intrinsically safe - this is extremely difficult in reality.

There are exceptions to the basic earthed equipment method of protection. Some equipment is designated 'Double Insulated', and usually has a symbol of two concentric squares that indicates that the equipment is double insulated, and that an earth connection is not needed (or in some cases *must* not be used). The common plug-pack (wall-wart) power supply is nearly always double insulated, and such equipment has reinforced insulation, designed to ensure that it is not possible for the live AC connection to connect to the secondary electronics in any event - including a complete meltdown. The electrical safety tests to verify that a product meets the Double Insulation standards are rigorous and expensive, and are very difficult to meet with high powered equipment, and even more so when the equipment has a metal case. Nearly all power amplifiers (for example) are *not* double insulated, and require an earth connection.

Colour Codes

The common colour codes for mains wiring are shown in Table 1, below. The Active (or Line) is the

'live' conductor, and carries the full AC supply voltage. The Neutral conductor is not live, but is intended to be the return path for all current in the active lead. Although this has caused great confusion to a great many people, it is sensible and logical (I will not go into the reasons here). The safety earth (or ground) conductor is intended to provide protection against electrocution, and where fitted, must not be disconnected.

Conductor	IEC	US	Alternative
Active (Line)	Brown	Black	Red
Neutral	Blue	White	Black
Earth	Green/Yellow	Green	Green

Table 1 - Common Mains Colour Codes

These colour codes are not standardised, and some variations may be found in different countries. The column headed 'Alternative' refers to an old code that used to be used in Australia and several other countries prior to the IEC codes being adopted. The one common theme of these codes is that they have been designed so that colour-blind people will not get the wires mixed up. The use of green with yellow stripes for the earth makes this even more secure. I have no information on the history of the determination of the colours used, but it is of little consequence since we can't change it.

Earth Loops

Figure 1 shows a typical connection of two hi-fi components, and includes the house wiring and main earthing point. As can be seen, there is a loop (indicated by the dotted line), which includes the interconnect cable, power leads, and a small part of the house wiring. Such loops are a major cause of hum in systems, and it is not uncommon for people to remove an earth wire from one or the other mains connectors to break the loop and stop the hum. The situation is much worse if different wall outlets are used for different parts of the sound system. In this case, the loop may extend all the way back to the main switchboard, making it longer, and more likely to have a significant voltage between individual earth connections.

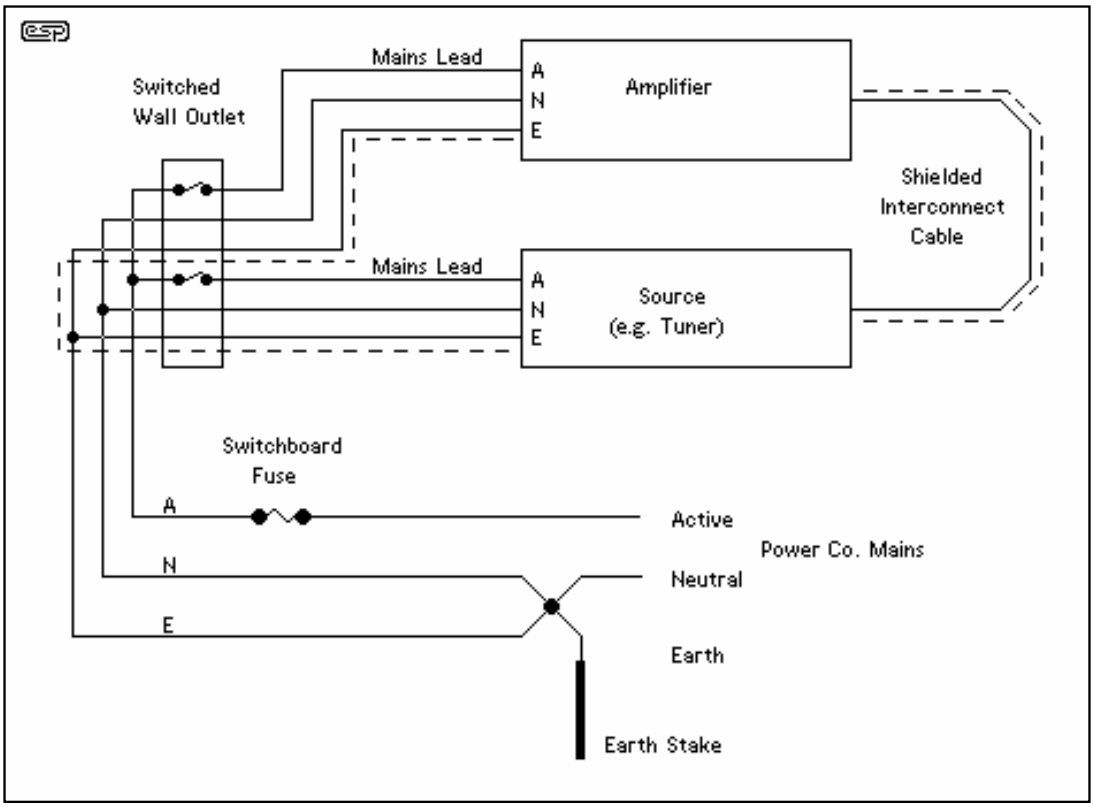


Figure 1 - The Formation of an Earth Loop

Also note that the neutral (return) conductor is attached to earth at the main switchboard. This is standard in Australia and many other countries, but might not be the case where you live. Check with an electrician who can tell you how this is done (if you really want to know).

What happens if the amplifier develops an electrical fault that allows the live AC conductor to come into contact with the chassis? The current will flow from the chassis, through the earth connection, and the fuse will blow in the switchboard or in the equipment if a fuse is fitted.

Should the safety earth be disconnected from the power amp (for example), if a fault occurs in the amp, the only earth return is now via the interconnects (assuming that the source is earthed). These interconnects are not designed to withstand the fault current that can occur with a major electrical fault, and may disintegrate before the fuse. You now have a live chassis on the amplifier - just waiting for someone to touch it and die!

Residual Current Detectors

Many new installations use an 'Earth Leakage' or 'Residual Current' detector, a device that will disconnect the AC supply if the current flowing in the active (live) conductor is not exactly matched by that in the neutral. Any imbalance means that current is going somewhere it should not be, and the device will trip.

These safety circuit breakers are very fast acting, and have saved many lives since their introduction. The 50mA that will kill you is detected by the breaker, and the power is disconnected - fast! Most of these type of breakers will operate on as little as 20mA, so you are not only protected against major faults, but also against excessive AC leakage caused by faulty insulation or moisture.

This does not mean that you can now go around disconnecting earth connections to stop hum - the safety devices that may be fitted to your house wiring are designed to trigger on a fault *before* you find it the hard way. In many countries, it is illegal to tamper with electrical (mains) wiring unless you are licensed - but in all countries, if it can be proven that you disconnected an earth that allowed a fault to kill someone else, *you are liable, and may be subject to criminal charges* ! Is that scary?

Main Earth Connection

For those who build amps (as is the case with many of the readers of these pages), a common question is "How should I connect the mains safety earth to the chassis?". As I stated above, the regulations change from one country to the next, but the principles are the same. Figure 2 shows a view of the basic connection, which is very safe. The lug used must be an approved earth lug (or one that meets any standards that exist where you live).

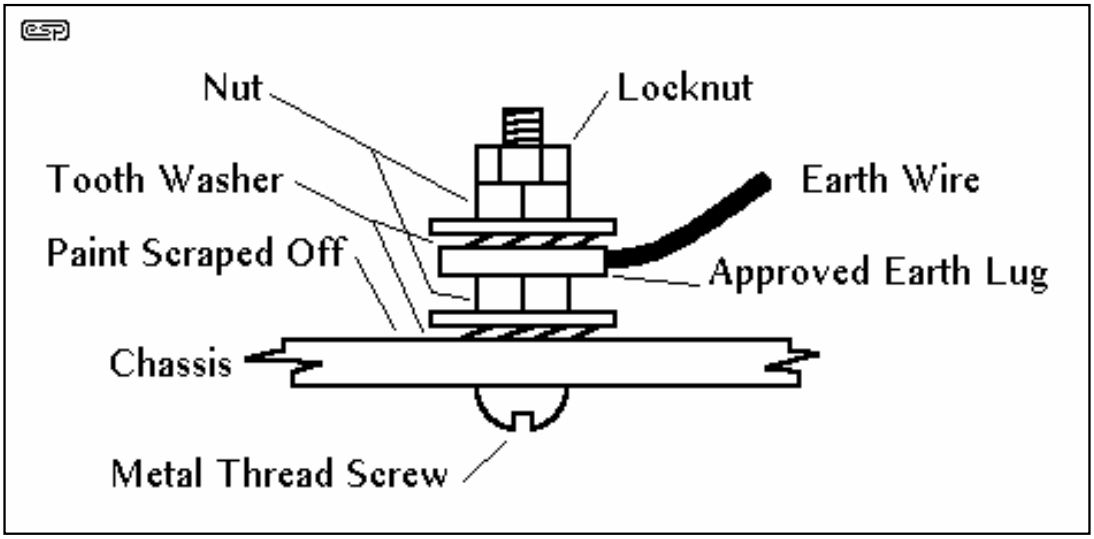


Figure 2 - Safe Method Of Connecting The Safety Earth

Any paint (or anodising, in an aluminium chassis) must be scraped away to expose bare metal, and the tooth washer ensures that there is a good 'bite' into the metal itself. The use of two nuts is strongly recommended, since the second one acts as a locknut, and prevents the first nut from loosening. The flat washers shown are optional, but highly recommended. They may be a requirement in some countries.

Do not use the earth connection as mounting for any other panel or component - it must be dedicated to the task of providing a safety earth point. If a component mounting bolt is used, at some stage it may be disconnected by a service (or other) person, which means that the apparatus is unsafe until everything is (hopefully) put back where it belongs - this does not always happen.

Make sure that the electrical connection between metal panels is also very well made. Some chassis are available in a kit form, and when screwed together, may not make good electrical contact with each other. Should the mains come in contact with a panel that has a flaky connection with the one that is earthed properly, the same potential for disaster is still present. All exposed metal must be properly and securely earthed.

The internal electronics of an amplifier should also be earthed, but now we have the problem of the hum loop again. There are two possibilities here ...

- Don't earth the internal electronics, or use a simple 'loop breaker' circuit to allow the case to act as a shield for radio frequency interference, but no solid connection is made (this is a common approach). This provides protection should there be a failure from the incoming mains to chassis, bit provides none at all if the transformer were to develop a fault between primary and secondary windings. Such faults are uncommon, but they can (and do) occur.
- Use a high current loop breaker circuit, ensuring that even major fault currents will be bypassed to the safety earth conductor. Such a circuit was described as a part of the 100W Guitar Amp project, but is shown again below. Be warned that this circuit (while safe) may not be legal where you live.

Use Of Loop Breaker Circuits

While very effective (and safe), as mentioned above such a circuit might not be legal where you live. If this is the case and hum is causing you grief, the use of balanced interconnects will solve the problem - but at some cost, and will require balancing circuitry at each end of all the interconnects. This is the approach taken in all professional equipment, and is highly effective, allowing all safety earth connections to remain exactly as they are to prevent electrocution of the artists or road crew. Suitable circuits for home (or professional) use are shown in the projects section.

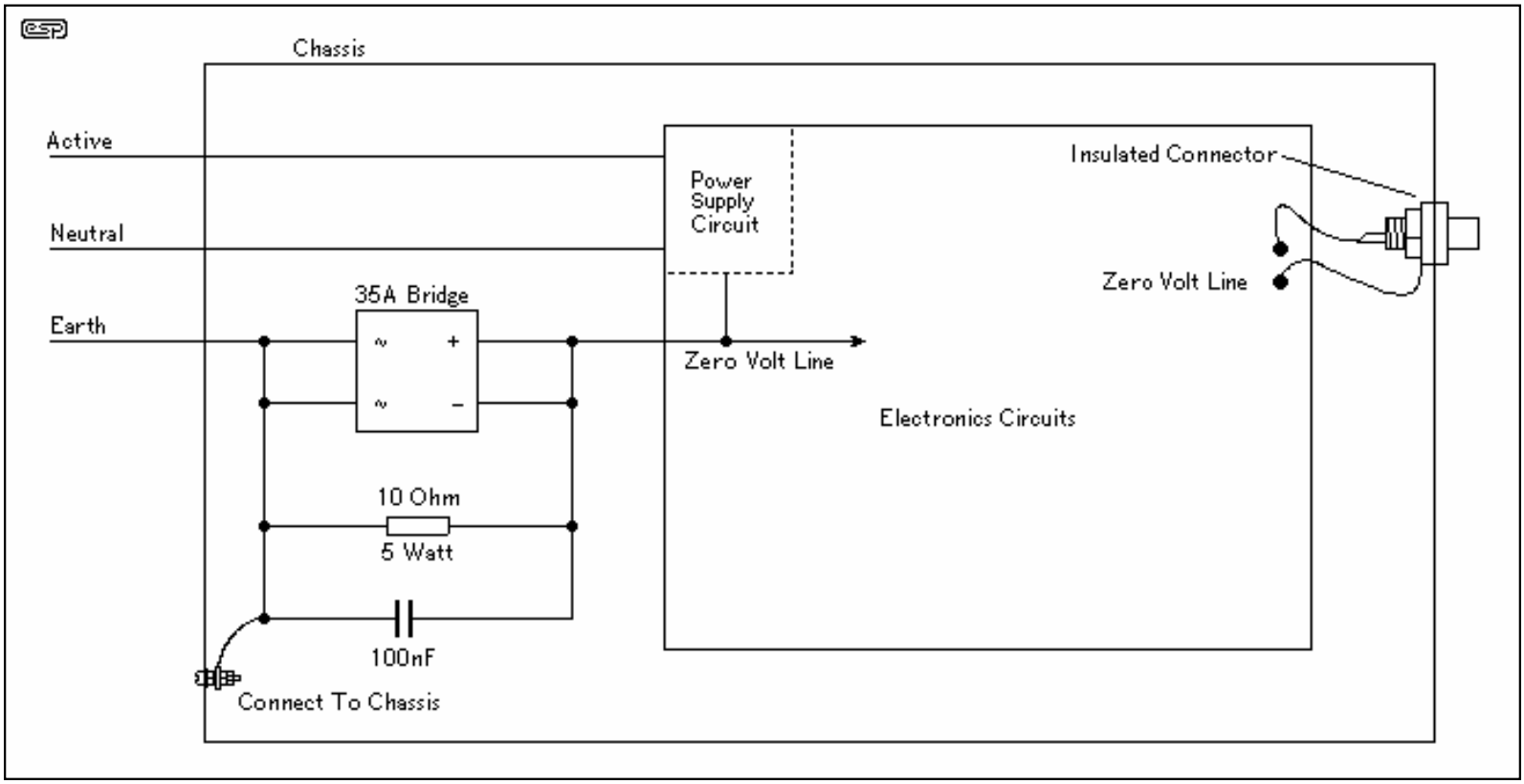


Figure 3 - A High Current Safety Loop Breaker Circuit

I have simply shown all internal electronics as a box, with the only connection to the loop breaker being the zero Volt line. This is most commonly taken directly from the centre tap of the main amplifier filter capacitors, but should always be connected to a point where there is high current wiring back to the transformer. It is the transformer that provides isolation from the mains, and the possibility of an internal transformer fault must be catered for.

The only exception is if a double insulated mains transformer is used, but these are rare. Should the transformer be of "conventional" construction (not a toroidal), then the transformer body - the steel core - must be connected to chassis directly. *Do not use any loop breaker circuit to isolate the transformer core, as it is unnecessary and dangerous to do so*.

The loop breaker works by adding a resistance in the earth return circuit. This reduces circulating loop currents to a very small value, and thus breaks the loop. The capacitor in parallel ensures that the electronics are connected to the chassis for radio frequency signals, and helps to prevent radio frequency interference. Finally, the diode bridge provides the path for fault currents. The use of a large chassis mounting (35A) type is suggested, since this will be able to handle the possibly very high fault currents that may occur without becoming open circuit. Note the way the bridge is wired, with the two AC terminals shorted, and the two DC terminals shorted. Other connection possibilities are dangerous, and must be avoided.

In the event of a major fault, one (or more) of the diodes in the bridge will possibly fail. Semiconductors (nearly) always fail as short circuit, and only become open circuited if the fault current continues and 'blows' the interconnecting wires. High current bridge rectifiers have very solid conductors throughout, and open circuit diodes are very rare (I have never seen a high power bridge go open circuit - so far at least). Use of the bridge means that there are two diodes in parallel for fault current of either polarity, so the likelihood of failure (to protect) is very small indeed.

When a loop breaker is used, it is vitally important that all input and output connectors are insulated from the case. If not, they will instantly defeat the loop breaker by providing a direct connection from the zero Volt point to chassis, and no benefit is obtained. (Electricity has an annoying - but perfectly logical - tendency to travel along the path of least resistance, and a direct short circuit will always have less resistance than the loop breaker.)

It is not uncommon to have a voltage of 1 or 2 Volts between the earth connections of power outlets that are wired separately back to the switchboard. This small voltage, with a total resistance of perhaps 0.5 Ohm, will cause a loop current of 2 to 5 Amps, all of which flows in the shield of the interconnect. This is sufficient to cause a voltage difference across the interconnect, which the amplifier cannot differentiate from the wanted signal. By breaking the loop with the 10 Ohm resistor, the current is now less than 200mA, and the voltage across the interconnect will be very much smaller, reducing the hum to the point where it should no longer be audible.

An earth loop will typically inject a 50Hz or 60Hz hum into the signal - if you have a 100Hz or 120Hz hum (which often has a hard edge to the sound), you have done something wrong in the wiring of the power supply, and the technique above will not help.

Mains Filters

In some cases, a transformer may be fitted with an electrostatic shield, which are lamentably uncommon in hi-fi transformers. Where provided, these too must be connected directly to the main earth point, and not via the loop breaker. Again, this is unnecessary and potentially dangerous.

The purpose of the electrostatic shield is to intercept (and earth) any interference on the incoming

mains. It does this by preventing any signal from being capacitively coupled from the primary to secondary windings, so the only form of coupling in the transformer is via the magnetic field in the transformer core. Most mains transformers have relatively poor high frequency response and this helps to further reduce interfering signals.

This can dramatically reduce extraneous noises (clicks, pops, whirring sounds, etc.) that might get into the system via the house and supply company wiring. This has great potential to pick up noise, since there may be 50 to 100km (or more) of cable involved between your amplifier and the generating plant.

In some cases, a mains filter might be fitted to amplifiers or other equipment (such as specially designed mains leads or 'black boxes') to reduce any interference. Where fitted, if an earth connection is provided, it must be connected to the safety earth and chassis - never to the amp's zero volt line. Typical filters will use Metal Oxide Varistors (MOVs) to cut off any high voltage spikes, and a capacitor and inductor network to filter out anything that is not at the mains frequency.

A true 50Hz (or 60Hz) tuned filter will be a large unit indeed, so most line filters only work at frequencies above a few kiloHertz. This is generally enough to get rid of most interference, since a well designed power supply should be able to filter out the majority of noise from the mains. Mains filters usually use the mains earth as a reference, so it must be present for the units to work correctly. Not using the safety earth as a reference is extremely dangerous, since the filters have capacitors which can (and do) become short circuited if a high voltage spike manages to get through and punctures the insulation.

Conclusion

Electrical safety cannot be over emphasised. Hum is damn annoying, and everyone wants it gone. There is no good reason to sacrifice one for the other, since safety and hum-free operation can peacefully co-exist with care and the right techniques.

As I have said several times, make sure that you find out the legal requirements in your country, and don't do anything that places you at risk - either from electrocution or legal liability. Neither is likely to be a pleasant experience.

Where the mains is noisy (apparently a common occurrence in the US), use of a dedicated mains filter is useful to prevent mains noise from entering the system. This will generally be unnecessary if the supply is well designed (especially if an electrostatic shield is used on the transformer), but this is often the exception, rather than the rule.

The use of 'specialty' mains leads (unless fitted with a proper filter which will be in the form of a box in line with the cable) is unlikely to solve the problem - regardless of claims made by the manufacturers or reviewers (see The Truth About Cables, Interconnects and Audio in General for my comments on these - this article made a lot of audiophiles very unhappy, but advertising hype does *not* negate the laws of physics).

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 30 Dec 1999

The Audio Pages

Elliott Sound Products **DIY Heatsinks**

Copyright © 2006 - Rod Elliott (ESP) (See note)
Published 11 February 2005

Articles Index

Main Index

Contents

- 1 Introduction
- 2 Description
- 3 Heatsink Assembly
- 4 Heat Spreader
- 5 Final Assembly
- 6 Conclusion

Note: The basis for this article was originally written by John Inlow, and was available on his website up until 2002 (when the entire site disappeared). Drawings have all been substantially revised (and colour added to make them clearer), and are based on John's originals. The photo in Figure 1 is a cropped and cleaned version of John's original picture. Parts of this document may still retain original copyright. The text has been almost completely re-written.

1 Introduction

Build your own heatsinks? Absolutely - this page shows you how to build a heavy duty heatsink and chassis, suitable for Class-A or high power Class-AB amplifiers. While there is no claim that the end result will be cheap (nor are store-bought heatsinks), the performance can be as good or better than anything you can buy.

At only 20% efficiency for a Class-A amplifier, the wasted heat is enormous! Out of 100 watts of input power, only about 20 of those Watts are available as useful sound. The rest of the power (80 Watts) is dissipated as heat, and must be removed - all the while keeping the transistor junctions at a safe operating temperature. It takes big heat sinks to remove the excess heat from the power transistors. It can also be unbelievably difficult to locate affordable, large heatsink stock (and this seems to be worse in the US). So, this page describes a solution to the problem.

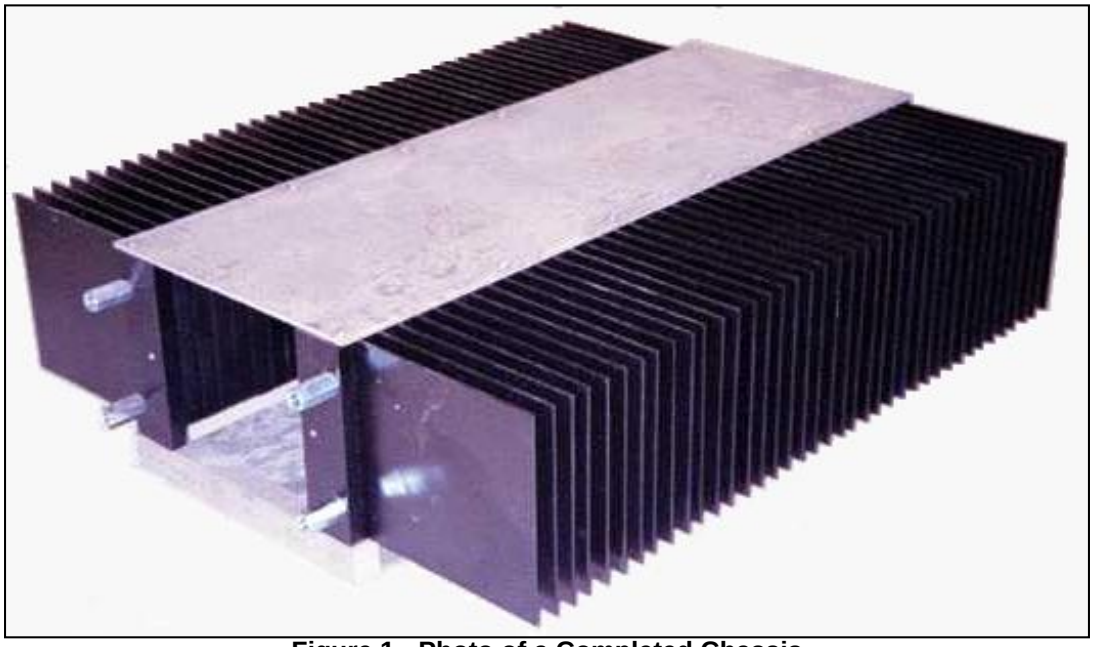


Figure 1 - Photo of a Completed Chassis

The pictured chassis (John Inlow's photo) will easily dissipate well over 160 Watts of wasted power - all as heat. If this meets your requirements, get ready to head for your workshop in readiness to make a significant amount of mess. The design is totally adaptable, so any size of heatsink can be produced, designed to fit any opening size or application. The chassis, although time consuming, is easily reproduced by anyone with basic construction skills. Use a drop saw (also known as a 'chop' saw in the US) with a carbide tipped blade designed to cut soft metals.

The saws and blades are readily available from most hardware stores, and are relatively inexpensive - actually, one can purchase the power saw for less than the blade (a very odd situation indeed). Be sure to use plenty of cutting fluid designed for machining aluminium. Denatured alcohol (methylated spirit) is excellent, but is highly flammable. On the positive side, there is no oily residue that can spontaneously combust if incorrectly stored. Take ***extreme*** care with all cutting fluids, as there are risks with all of them - especially anything that works well with aluminium.

---

2 Construction

Before going into great detail, it is important to look at the following drawings so you can see what is involved. There is no doubt that you will need some fairly serious tools to be able to tackle something of this nature. Apart from the drop saw and metal-cutting blade mentioned above, you will need a drill press - a power hand drill could be used, but the results will be disappointing or very, very time consuming (or both). You also need taps (for cutting threads, not whatever you might have been thinking). A decent sized work area is also needed, and be aware that you will create a vast amount of aluminium chips, powder and swarf. The kitchen table is not recommended.

You also need the usual array of hand tools and miscellany - a hacksaw, files, drill bits, G-cramps, and various bits of scrap material that can be used to construct drilling and assembly jigs, etc., etc.

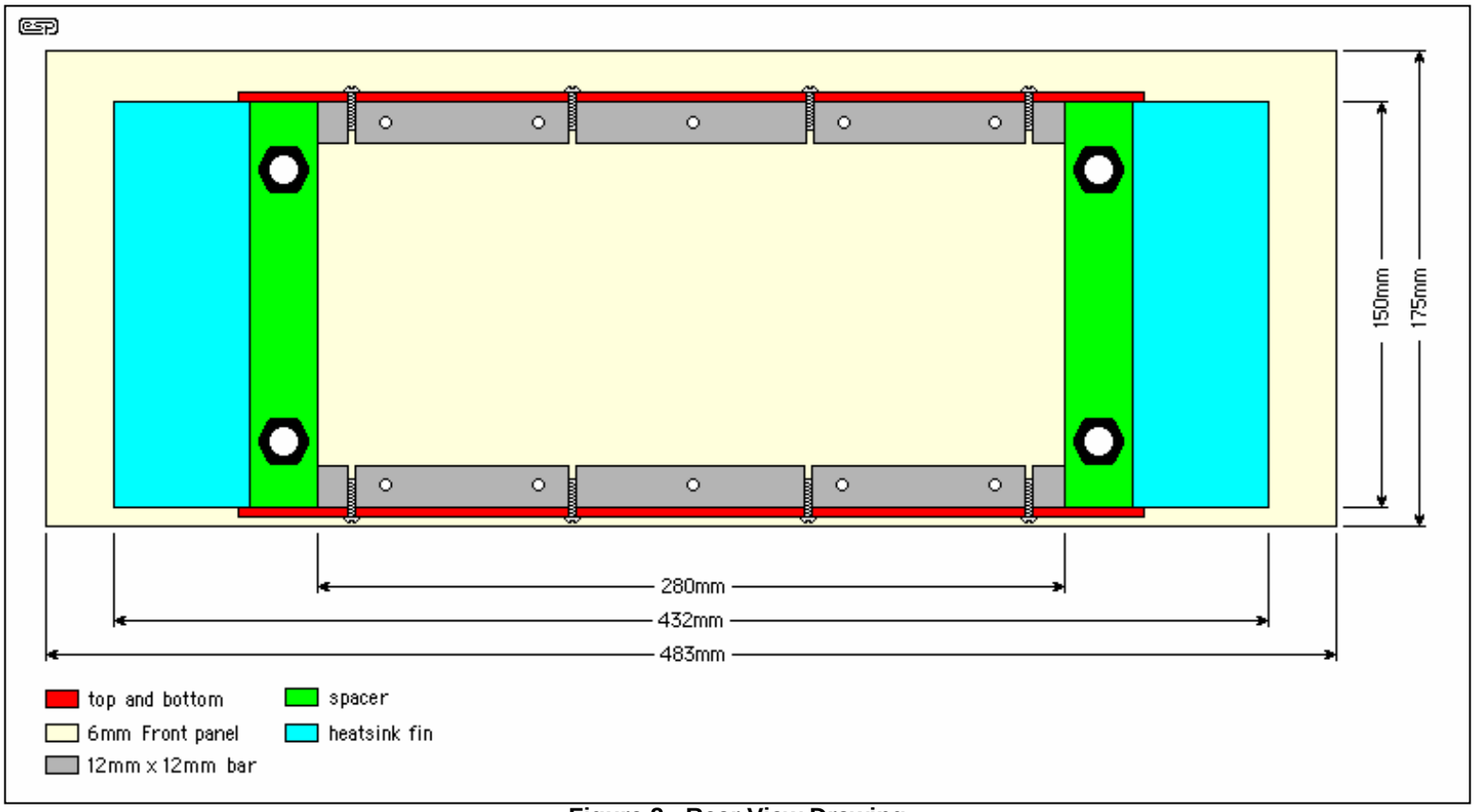


Figure 2 - Rear View Drawing

Looking from the rear, we see the two heatsinks, top and bottom covers, as well as the front panel. If you insist on using imperial measurements, then divide all measurements shown by 25.4 to obtain inches. The dimensions are not critical for the most part - adapt to suit your specific application.

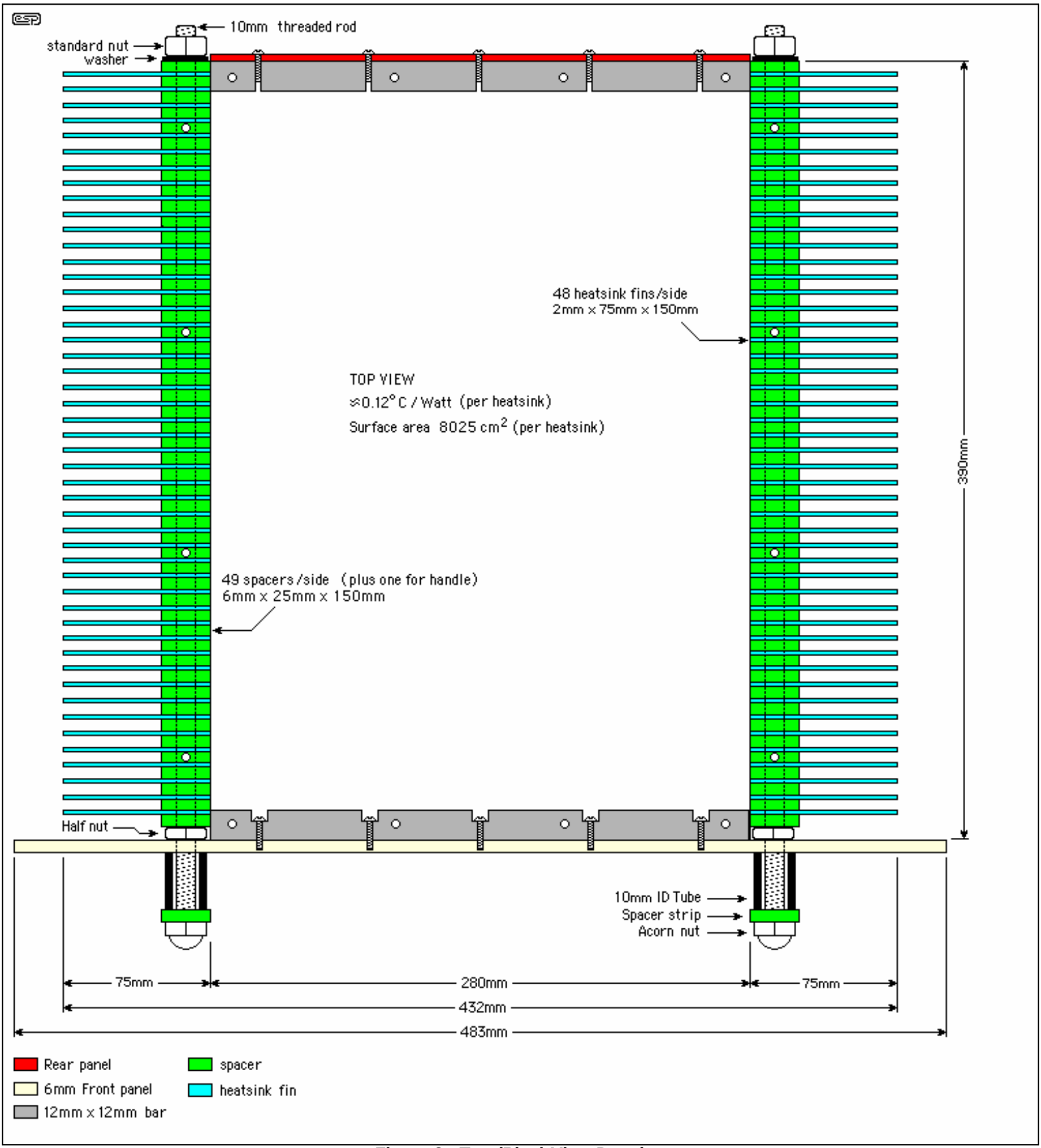


Figure 3 - Top (Plan) View Drawing

The plan view gives you the rest of the picture. Again, adapt the dimensions to suit your application. The heatsink sections can be as large (or as small) as you need, limited by available funds and your patience (as always).

It is suggested that you use 6 x 25mm (or 0.25" x1.0") bars for the spacers and 2 x 75mm (0.08" x 3.0") bar or cut

sheet for the heat dissipation fins. The length depends on your intended height - all drawings here are based on a 150mm (6") heatsink height. Starting with a spacer section on the outside, the two sections are repeated (spacer-fin-spacer-fin ... spacer) until you have the length you need. To achieve the desired depth for the chassis, the plates are bolted together onto four (two for each heatsink) 10mm (3/8") threaded rods. Drilling the holes is a tricky and time consuming job - they must be aligned perfectly. The difficulty of the process is reduced if you create a jig to position the material as you drill the holes. Although it adds time to the whole process, partially pre-drilling each section with a centre drill (a special drill bit with a thick shank and small stubby tip) will ensure that the drill bit does not wander, causing holes to be off-centre. It is very important to clean up the holes after the drilling is completed. Using either a larger drill bit or a de-burring tool, chamfer the holes to remove the jagged edges that form when drilling. Don't even think about avoiding this step - the plates will not touch one another evenly so heat transfer will suffer, and your assembled heatsink will never be straight.

If you use 10mm threaded rod, the holes should be drilled to 12mm. This allows for the occasional hole that is slightly off centre, and also lets you align the assembly perfectly before it is finally clamped up tight.

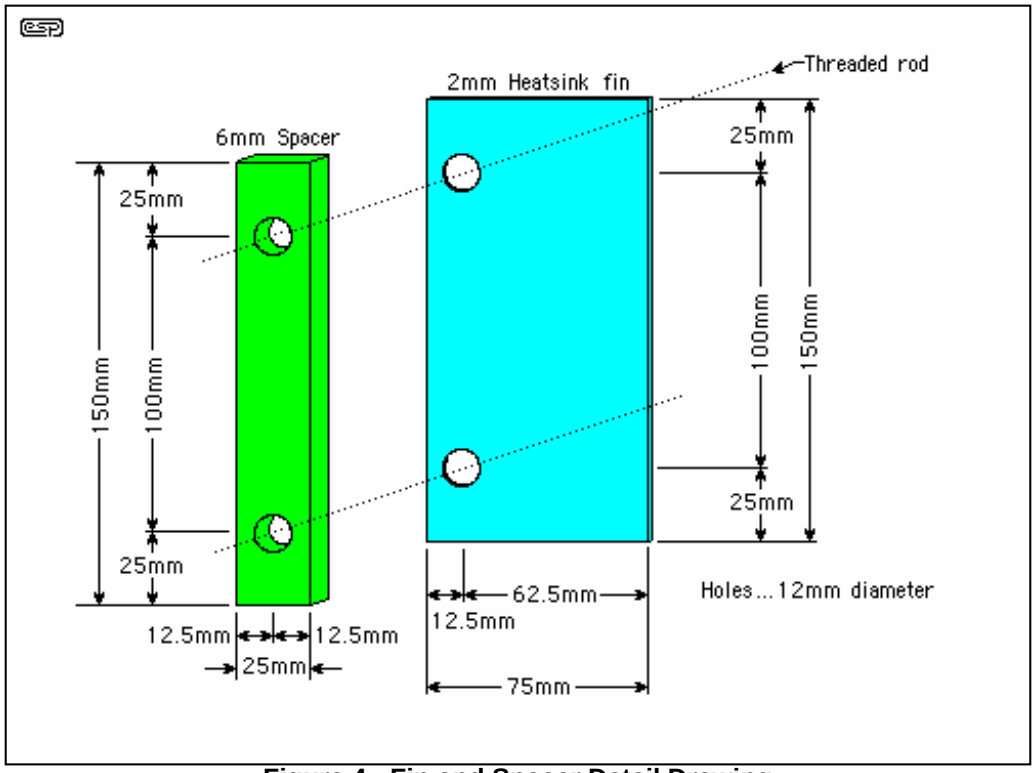


Figure 4 - Fin and Spacer Detail Drawing

Keep repeating the above pattern until you achieve the desired depth for your chassis. The same procedure is used for both sides of the chassis. The threaded rods must be long enough to pass through all of the spacers and heatsink fins, and still have sufficient left over for nuts. Optionally, the rods may be extended further to create a mounting for handles at the front of the chassis.

Make sure that all surfaces are flat after drilling. De-burr all edges before assembly to get the best possible contact between the spacers and fins.

---

3 Heatsink Assembly

When everything is machined, cleaned and ready to assemble, begin by screwing a half nut onto one end (to become the front) of the four threaded rods. Make sure that the distance from the end of the rod to the face of the half-nut is correct for your handle sections (allow for front panel, tube (or long nut if you prefer), handle section and acorn nut). Then slide on the first spacer, followed by a fin,then a spacer (etc.) in alternating fashion. When

completed, screw a full nut onto the remaining end (rear) of the rods - this should only be finger tight at this stage! You need to get the distance at the front dead right if you are planning on adding handles - the final acorn nut does not have much depth. Feel free to add washers at the front as well as the back, but make sure they are allowed for in the length of the threaded rod.

Alignment is critical. Because your heatsink is made from many separate pieces of aluminium, it is possible for it to assume many different and entertaining shapes, within the constraints of the threaded rod. None of the potentially entertaining shapes is useful - you need flat and square. Period.

To achieve this, you'll need a flat surface with a piece of scrap angle the same length as the heatsink assembly screwed to one edge. You also need a small hammer, a square, a scrap piece of aluminium and a piece of timber. First, lay the assembly onto the surface, with fins pointing upwards. Slide it across so that it contacts the angle. The alignment is by nature repetitive - each step will need to be repeated - possibly several times ...

- Using the scrap timber, gently tap the edge opposite the angle (tap towards the angle) until the heatsink lies dead straight along the angle. All fins and spacers should be level on both sides.
- Use the square to ensure that the ends are at a perfect right angle to the base board *and* the angle section. Adjust as needed.
- Tap the fins and spacers down (use the scrap aluminium to reach between the fins to get to the spacers). The idea is to ensure that every fin and every spacer is flat against your alignment surfaces.
- Tighten the rear nuts slightly, and repeat the above process.

Once completed and the nuts are firm, you can clamp a piece of wood on top of the assembly, thus clamping the heatsink to your flat surface. Be careful that you don't disturb any of your alignment during this process. Now the rear nuts can be tightened fully. They should be tight, but not ridiculously so - a stripped nut or threaded rod is not a bonus at this stage.

Upon removal from the assembly jig, the heatsink should be flat and square, requiring the minimum effort to obtain excellent thermal contact between the reinforcing bar or mounting plate (see below). If the nuts are tight enough, firm effort on your part to bend or twist the completed assembly will result in nothing more than heatsink imprints on your hands - the heatsink itself should remain nice and flat, with no bending or warping. If it *does* warp, you will have to repeat the final step, after loosening the nuts just enough to allow the assembly to be made flat once again. Although this is a frustrating step if things move, it is highly recommended - you will find out for yourself how rigid the heatsink is, and whether (or not) you need to add a reinforcing bar (or perhaps just tighten the nuts a little more).

If it is possible to distort the assembly (or it comes out pre-warped), then you must use reinforcing bars screwed to the back of the heatsink (all drilling and tapping must be into the spacer strips *only*. Alternatively, a full-length flat plate will also provide reinforcement, but at somewhat greater expense. These options are discussed below. Note that one or the other needs to be used anyway - at issue is how long it must be to achieve rigidity and good thermal contact.

---

4 Heat Spreader

After assembly and prior to fitting the heat spreader (either a reinforcing bar and/or flat spreader plate), it is advantageous to mill the rear of the heatsink to present the flattest possible surface. Given that few hobbyists have access to a milling machine, an alternative is to carefully file the entire rear surface - this will be an extremely tedious job, but will improve performance. A linisher (belt sander) can also be used, but the surface must be finished with a fine grit so it is as smooth as possible. The photo in Figure 6B shows the rear surface of a small heatsink I made, and the linishing marks are visible. I deliberately did not complete the job so the anomalies could be seen clearly.

Although the surface shown is very flat (better than 25um / 50mm), the surface finish will almost certainly not be good enough for direct mounting of semiconductors. This makes the heat spreader mandatory. You can check

surface flatness by trapping a thin hair (for example) at various places on the surface with the edge of a steel rule. It should be not possible to pull the hair from beneath the edge of the rule at *any* position under the heat spreader location.

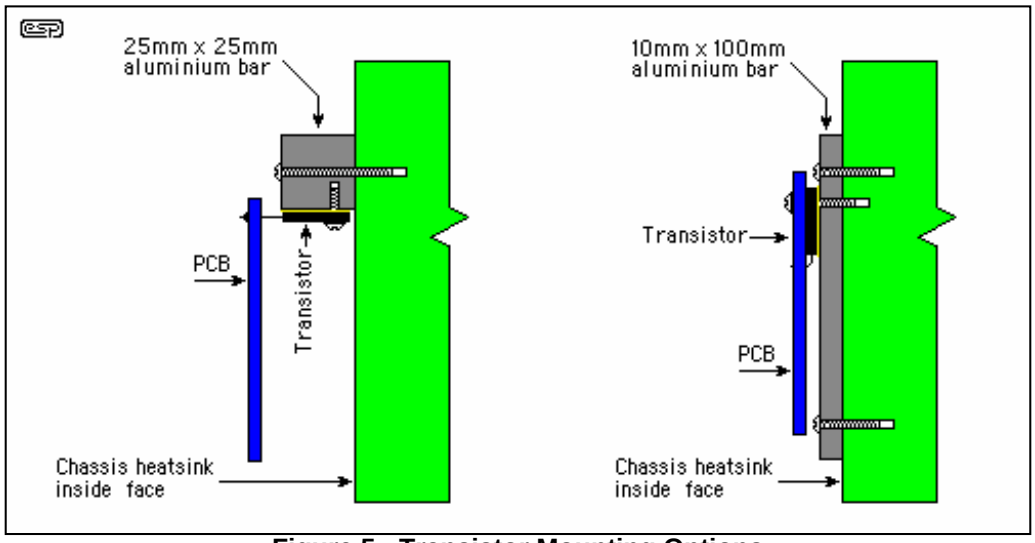


Figure 5 - Transistor Mounting Options

Before assembly of the heat spreader to the heatsink itself, apply heatsink compound between the spreader and the heatsink back. Apply a thin layer, and check that the two surfaces mate well by pressing the bar onto the heatsink. Remove it, and check that the heatsink compound is evenly distributed and shows signs of full contact. This will be immediately obvious upon inspection.

The transistors may be mounted directly to the reinforcing bar as shown. Normal transistor mounting procedures apply. Alternatively, attach a section of flat bar as shown in the right-hand drawing. This method produces a nice flat surface, ideal for mounting boards where the transistors (or MOSFETs) are under the PCB.

It is imperative that all holes for the bar or plate line up with the centres of spacer sections - drilling and tapping such that a hole is part way between a fin and a spacer will cause deformation of the assembly, resulting in potentially dramatic loss of performance.

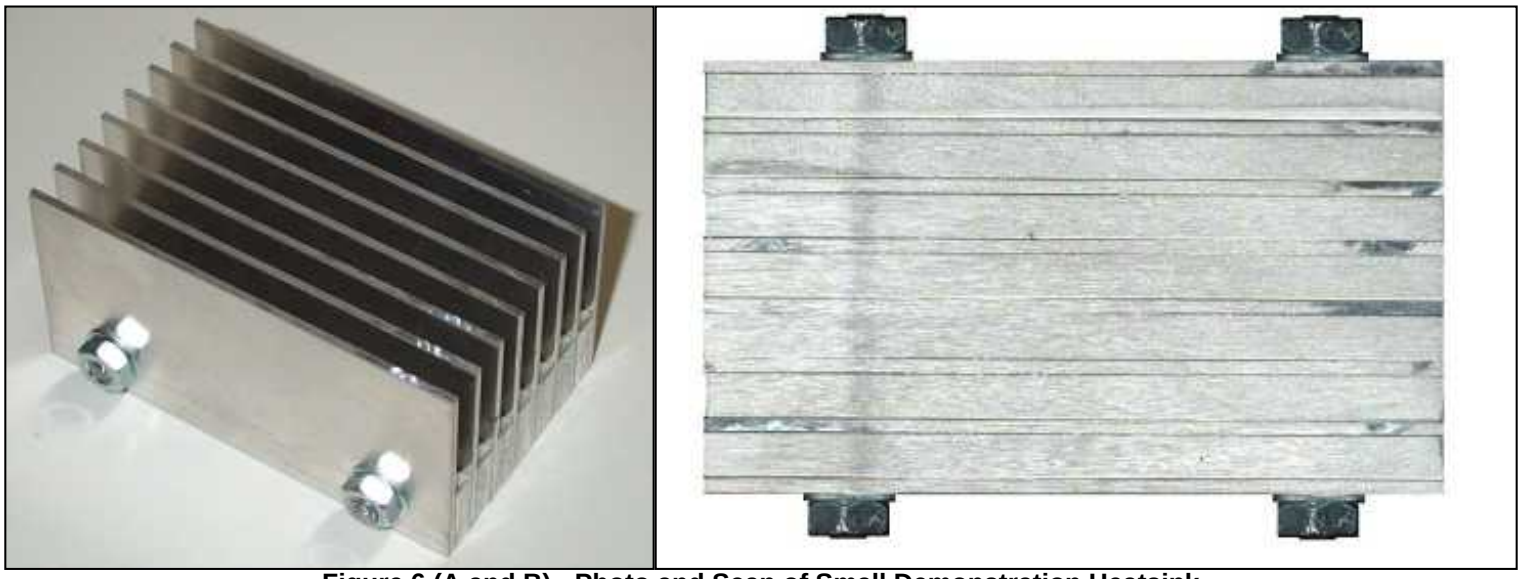


Figure 6 (A and B) - Photo and Scan of Small Demonstration Heatsink

The above photos are of a small demonstration heatsink I built. The left side (A) shows the general construction, while B shows the surface finish on the underside (the latter was scanned to get the best image of the surface). I

deliberately didn't complete the machining process so you could see the aberrations that you *will* get when the heatsink is assembled. The dark areas are actually the shiny (not sanded) areas of the fins. This was despite my following the instructions listed above, so you will also have the same problem. Attempting to remove every imperfection is futile unless you have access to a milling machine - it will be too boring and frustrating for words. The last 5% of the surface could easily take 90% or more of the total construction time.

This is the reason for using a heat spreader - it distributes heat over a much larger surface area than you will get with any transistor, but the surface still needs some basic attention or heat transfer between spreader and heatsink will be badly affected. This will lead to an excessive temperature on the spreader, and an even higher temperature for the transistors. This machining is the most important part of the exercise!

I used a linisher (essentially an upside-down large belt sander), but careful filing will also work very well. Yes, it will be tedious and hard work, but the results will be worth your efforts.

You will also note that I didn't follow the procedure, in that I have fins (rather than spacers) at each end of the heatsink. Feel free to break the rules too, provided you work out exactly what you need. The whole idea is that this process allows you to make a heatsink that exactly fits your needs. In case you were wondering, the overall size of the heatsink pictured in Figure 6 is 100mm (h) x 60mm (w) x 52mm (d). There are 8 fins, each is 32mm deep, giving a thermal resistance of about 1.16°C / W (if black enamelled) or 2.0°C / W (polished aluminium, as pictured).

---

5 Final Assembly

Once the heatsink sections are drilled and tapped to accept the top and bottom covers, the heat spreader has been drilled and tapped for the transistors, and no further work needs to be done in that area, you may start the final assembly. The bottom needs to be drilled for external feet and for all internal hardware that will be attached to it. As shown in Figure 3, the heatsinks (as well as top and bottom covers) are drilled to accept mounting screws. The heatsink holes are tapped to accept the size screws you will use, as will the heat spreader.

The front panel simply attaches to the two heatsink sections as shown in Figure 3, and the top and bottom panels attach to the heatsink and front and rear panels. Although no additional mounting for the rear panel to heatsinks is shown, this can be added if desired. You can use 12mm square section, or a piece of angle if you prefer. The extra is not really needed, but the rear panel will flex a little when the top is not in place without it.

If you find that you need the reinforcing bar(s), first assemble the heatsink as described above, then drill and tap the holes for the reinforcing bar (in both heatsink spacers and bar). Assemble the heatsink as described, carefully turn it over and file, sand or otherwise machine the rear surface so it is as flat as possible. Then, attach the reinforcing bar, and screw down lightly. You will quickly see if there is any misalignment, and this must be corrected before you permanently attach the bar to the heatsink. In general, it is expected that if the assembly process is carried out carefully, the heatsinks will be very rigid indeed - almost as if they were one solid piece of aluminium.

The unit as described is *very* heavy, so file and sand all edges and outside corners of the ribs to reduce the risk of being cut. It will be much easier to assemble if you drill and tap all the heatsink holes for mounting the reinforcing bars and / or heat spreader before final finishing.

Paint the finished heatsink with a spray can of flat (matte) black enamel. It should be a self-etching type designed for aluminium finishing for best results. Don't be tempted to try for a perfect finish between the fins, as you will end up applying too much paint which will reduce performance. You could get all the sections black anodised before assembly, but that would be an expensive option because of the number of pieces. Don't even contemplate pre-finishing the heatsink components - any paint between fins and spacers will degrade performance dramatically, and may cause the heatsink to bend or twist when it is assembled.

The front panel arrangement may look rather suspect, with the 12mm bar attached using screws into blind holes in

the panel. With a 6mm panel, this allows 5mm hole depth, and about 4mm of this can be threaded. Having done exactly this on many occasions, I can assure the reader that it works perfectly. You do need to be extremely careful to ensure that the holes don't go all the way through, but other than that, the process is straightforward. Tapping blind holes is a cow, but you can cheat and use self-threading screws. The latter must be square-ended - conventional tapered self-tapping screws will not work! If you are concerned, use a good epoxy resin (24 hour setting type - *not* 5 minute) as well as the screws for a permanent bond.

---

6 Conclusion

As described, each heatsink section will have a thermal resistance of around 0.12°C / W. This is assuming a middle-of-the-road value for emissivity of about 0.8 - typical of a flat black surface finish you will be able to apply at home. This is a very good figure, but it will be degraded if the heat spreader is too small, or makes ordinary (as opposed to excellent) thermal contact with the heatsink, etc. The distribution of transistors will also have a bearing on the thermal performance.

However, even if we were to assume the worst case and the thermal resistance is effectively doubled, it is still very good at 0.24°C / W. The 'typical' figure (0.12) was calculated using the ESP heatsink calculator program (see the downloads section of the ESP website). Heatsink temperature was taken as 60°C, at an ambient temperature of 25° C. Even using the worst case figure, this means that when dissipating 80W (as discussed at the beginning of this article), the heatsink temperature will stabilise at around 20°C above ambient (45°C at 25°C ambient) ... and that's *worst case!*

Because of the large thermal mass (as well as actual mass), this heatsink will take a significant time to reach full operating temperature. Each heatsink will weigh in at about 5.9kg for aluminium alone, and even the baby one I made (see Figure 6) weighs 480 grams (admittedly, that's with the steel threaded rod, washers and nuts). Both heatsinks will weigh about 12kg (over 26 pounds in the old measurements), and you have yet to add the panels, transformer(s) and other components.

The basis for this article was originally written by John Inlow, and was available on his website up until 2002. Drawings have all been substantially revised (and colour added to make them clearer), and are based on John's originals, as is the photo in Figure 1.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Parts of this page (in particular Figure 1 and parts of the redrawn diagrams) may also be the intellectual property of John Inlow.

Page created and copyright © 05 Dec 2005./ Published 11 Feb 2006

Elliott Sound Products **Impedance**

Copyright © 1999 Rod Elliott
Last Updated 24 Jan 2005

Main Index

Articles Index

Contents

- Introduction
- Amplifier Impedance
- Speaker Leads
  - Power Loss
  - Damping Factor
  - Speaker Impedance and Protection
  - Speaker Protection Systems
- Impedance Matching
  - Radio Frequencies
  - Telecommunications
- Balanced Vs. Unbalanced
- Hi-Fi Super Cables
- Conclusion

Introduction

We hear so much about damping factor, the effect of speaker leads (and how much better *this* lead sounds compared to an "ordinary" lead), and how amplifiers should have output impedances of micro-Ohms to prevent "flabby" bass and so on. But what does it all really mean?

Before an informed judgement can be made, we need to look at some of the real factors involved. There are a multitude of impedances involved in a typical amplifier to loudspeaker connection, most of them having a vastly more profound effect than the amp's output impedance alone.

For example, I have demonstrated a modified version of the 60W amp, which used current drive output. This version exhibits a damping factor of 0, since it has an output impedance of about 200 Ohms - the most common reaction was "Wow, that sounds just like a valve amp!".

Not really true, since valve amps actually have a damping factor of at least 1.5 or so, and exhibit other characteristics which are very difficult to emulate with transistors or MOSFETs, but this was the overwhelming critique of the amp's sound.

Indeed, my very own (bi-amped) hi-fi uses an amp for the bass-mid with a designed output impedance of about 4 Ohms. This provides a useful extension of the bottom end (I'm using sealed enclosures), without excessive peaking at resonance (there is a bit, and the correct impedance is probably 2 Ohms, but I just haven't gotten around to making the modification yet). To some, this is absolute nonsense - how can you have tight bass with no damping? Easy, damp the enclosure properly, and don't expect someone else (the amp) to do it for you.

For the purpose of this article, I shall assume an amplifier of 100W into 8 ohms nominal load, which is typical enough for a reasonable system.

Amplifier Impedance

It all starts with the amplifier. An amplifier will always have a defined (and measurable) output impedance, and this will vary depending on frequency. In some cases during measurement, it may appear to change with load impedance too, but that is most likely caused by protection circuitry. In many cases, measuring the output impedance is extremely difficult, but the majority of amps will give sensible results.

With commercial products, the output impedance is often quoted, but rarely with and further information. To be meaningful, the impedance should be quoted at at specific frequency, or preferably at a range of frequencies. It would be helpful if the power level were to be stated as well, but this is not often the case. If you really want to know, you will have to measure it yourself.

Although relatively easy to do, there is some risk involved, so never undertake this process unless you know exactly what you are doing, and accept the risk that you may damage the amp if you make a mistake. You will need a resistor of known accuracy, with a resistance equal to the amplifier's stated load impedance. It must be capable of dissipating either the amplifier's full output power, or a level of power that you will not exceed. A 10W wire wound resistor is recommended, preferably non-inductive for broad band accuracy.

For this exercise, we shall assume the amp referred to in the introduction - 100W into 8 Ohms. An 8.2Ω 10W resistor will be fine for the basic test. The amplifier output voltage should be kept below 9V RMS at all times (9V and 8.2Ω is 9.8W - within the resistor's rating). Measure the resistor's actual value - do not use the stated value, as this could have an error of up to 10%. For the exercise, I will assume the resistor is exactly 8.2 Ohms.

Set a signal generator to the desired test frequency, and apply the signal to the amp input. Adjust the level until you have a convenient voltage below the 9V maximum. Measure the voltage carefully. Call this V1 (let's assume 5.00V)

Now, connect the load resistor directly to the amp terminals, and measure the voltage again. Do this accurately and quickly - the resistor will get hot quite quickly if the voltage is anywhere near the 9V maximum. Call the loaded voltage V2. It should be less than V1. (Assume 4.88V)

Now, you can calculate the amp's impedance (I will do this the long way, as it's far easier to remember) ...

I = V2 / R = 4.88 / 8.2 = 0.595A

The amplifier is most easily represented as a perfect voltage source having zero Ohms impedance, with a series impedance that we see as its actual output impedance. Knowing the voltage drop in the amp's internals (Vd) and current (I), we can calculate the output impedance (Zo) ...

Vd = V1 - V2 = 5.00 - 4.88 = 0.12V
Zo = Vd / I = 0.12 / 0.595 = 0.2Ω, or 200 milliohms

The so-called 'damping factor' is simply the speaker impedance (Z_L) divided by the amp's output impedance ...

DF = Z_L / Zo = 8 / 0.2 = 40

This test will usually show highest damping factor (lowest output impedance) at low frequencies, with the impedance rising from around 500Hz and upwards. The actual frequency depends a lot on the amplifier's topology - it may be as low as 100Hz, or as high as several kHz.

In some cases (rare, but it can happen), you will find that the voltage *increases* slightly when the load is applied. This condition is usually an indicator of bad internal wiring practice, and the amp has negative impedance.

Since this is an inherently unstable state it should be avoided, although it is remotely possible that some misguided soul thought it was a good idea. It isn't. With the exception of special-purpose amps where there is a need to solve a specific problem, negative impedance is generally something to be avoided. See the article (Variable Impedance Amplifiers) that covers this exact topic for full details.

Speaker Leads

After the amplifier, we have the speaker lead itself. This has become (for reasons which I must confess I find entirely obscure) a subject of great controversy. There is actually nothing controversial about a piece of wire, despite the manufacturer's assertions to the contrary.

The purpose of the speaker lead is to carry the output current from the amp to the loudspeaker, preferably with as little power loss as possible. This implies that the lead should be capable of carrying currents of perhaps 5 amps or so for a distance of typically 2 metres. This is hardly a difficult task - the mains lead to a 2400W electric heater carries a continuous current of 10 Amps (assuming 240V operation), and usually manages this with little loss. The house wiring manages to carry this current from the switchboard through to the power outlet for far greater distances, still with relatively little loss. Then, of course, there's the cable run between the switchboard and the power station - ????

Power Loss

Some of the more expensive speaker leads (costing 10's of dollars per metre) are capable of carrying

currents far greater than can ever be drawn by a loudspeaker connected to our "typical" amplifier.

At the very worst, a nominal 8 Ohm speaker system may have an impedance at some frequencies (usually at or near the crossover frequencies) of perhaps 4 Ohms - some may be lower than this, and I would suggest that in such cases the designer should return to the drawing board forthwith, since a truly monumental error has been made.

So, let us assume that the truly monumental error has been made, and look at the current the amp might need to supply. To make life simple (for the time being) we will look at the maximum possible power on a continuous basis - this will never happen, but it is useful as a comparative figure. The figures below are rounded, and assume that the power supply is actually capable of providing the currents quoted - this is highly unlikely.

Impedance	Power	Voltage	Current
8.0 Ohms	100 W	28 V	3.5 A
4.0 Ohms	200 W	28 V	7.0 A
2.0 Ohms	400 W	28 V	14 A
1.0 Ohm	800 W	28 V	28 A

These figures take things to extremes - an 8 Ohm loudspeaker which falls to 1 Ohm would have to be the most ill-conceived possibility imaginable, but we will stay with this, since it is useful to demonstrate the point, and there is probably more than one of them out there already.

Now, let us imagine a speaker cable with a resistance of 0.1 Ohm - for a 2 metre cable, this will be the so-called "speaker" cable that is a very thin figure-8, and rated at 5A (low voltage only).

Even at the lowest impedance of 1 Ohm, where at one isolated frequency the poor amplifier will be attempting to supply a maximum current of 28A - but will it? The likelihood of this one frequency being driven to full power is extraordinarily remote, but we will use it anyway.

So, now we have 28V across 1.1 Ohm, so the current is reduced slightly, to about 25A, so 2.5V will be lost across the cable at this frequency. The amplifier is only attempting to deliver a mere 650W to the load, with the cable dissipating the rest as heat. A 0.9dB drop in level is nothing compared to the agony the amplifier will be suffering, with the real possibility of second breakdown in the output devices (which does not always result in the instantaneous destruction of the transistors) - can you imagine what that would sound like?

A small amount of resistance is inevitable - there is no super-conducting speaker cable available as yet, so you are always going to have a small loss. It will require very thin wire and a very low speaker impedance before you even lose 1dB, so I believe we can dispose of that myth (note that this is a very simplistic look at the issue - there is a *lot* more involved). For a more in-depth discussion of this topic, see the various cable articles.

Damping Factor

Having established that resistance within reason is not an issue as regards power loss, there is damping

factor to be considered. Quite apart from the fact that "damping factor" is a rather ill-conceived term, we need to look at the reality of what the amplifier is capable of controlling within the speaker cabinet.

According to the proponents of "damping factor should ideally be infinity +/- 3dB", the amplifier is in total control of the speaker, provided its impedance is low enough. What utter rubbish. The loudspeaker is an electro-mechanical device, and its sensitivity to external impedance is easily tested. Indeed, if the amp were in total control, with the amplifier connected (and turned on), you would not be able to move the cone at all by tapping it.

Try this experiment. Disconnect the speaker leads from the speaker, and tap the cone lightly with your fingers - listen very carefully to the character of the "thump" of the speaker. Now, connect a short piece of stout wire directly between the speaker terminals - this will have a resistance so low as to be considered 0 Ohms.

Tap the cone again, the same way as before. Can you hear a difference? I would expect that you can, because if not the speakers are either so well damped internally that the amplifier's contribution is irrelevant, or there is so much internal resistance that no amplifier can save them.

Now, remove the short, stout wire, and replace it with a 4 metre length of the cruddiest bit of wire you can find, bell wire or telephone cable is ideal, but thin fencing wire will do equally well. Tap the speaker cone again, and listen carefully to see if there is any difference between the sound with the "dead short" versus the piece of rubbish you just connected, which might typically have a resistance of 0.5 Ohm (maybe more).

Could you hear any difference? If you have access to an AC millivoltmeter, you should try to measure a voltage as you tap the speaker cone. You will typically find that very little change will be evident in the character of the sound between the dead short and the bell wire, and only a small voltage (a few millivolts) will be indicated.

This is a rather extreme test, and it is quite possible that a very slight difference in "tonality" will be observed. It is equally possible that you will not be able to hear any difference whatsoever, despite the fact that damping factor has been reduced to a small fraction of its former value with the dead short circuit created by the short wire.

Even the best amplifier in the world will have some impedance, as will the most expensive cable. These can actually be "removed" completely by building an amplifier with a negative output impedance, but having tried this approach, I can assure the reader that a horn compression driver is the only speaker I have ever tested which sounded better with negative impedance. All other speakers, whether horn loaded, sealed or vented box sound universally disgusting with negative impedance.

Negative impedance is exactly what it sounds like - the amplifier is modified so that the voltage output rises with increasing load. This is surprisingly simple to implement, but speakers seem to hate it - most seem to prefer a small but measurable amount of positive impedance. This is the equivalent of using a cheap speaker lead, but has other implications - there is no loss of power due to the resistance of the lead, and high frequencies are far better since there is none of the capacitive loss incurred by the really cheap speaker lead.

Have you ever wondered what it is about valve (vacuum tube) amps that has many audio enthusiasts drooling over the latest - usually extremely expensive - offering from this manufacturer or the other?

They will wax lyrical about the extended bass, and the sweetness of the highs etc. etc. One of the things about valve amps is their relatively high output impedance - this may be up to about 6 Ohms for an amp without any feedback, and will rarely be less than about an Ohm (8 Ohm output selection is assumed). The damping factor is obviously grossly inferior to that of a transistor amp, but the sound quality is generally considered more musical, or "sweeter" by a great many enthusiasts.

There are many other factors involved than simply the output impedance, but if this were so important (or as important as some would have you believe) then valve amplifiers would be universally condemned for their poor low frequency performance - phrases such as "woolly", "muddy" and "over-emphasised" spring readily to mind. Instead, we can read reports where reviewers have praised the low frequency performance, claiming an extra 1/2 octave or so of bass extension. This is in spite of the fact that very few valve amplifiers can actually make it down to 20Hz without rolloff and output transformer saturation distortion starting to be readily observable.

Well before transformer saturation distortion and other undesirable effects make themselves known, the output impedance rises. This is at the very frequencies where damping factor is supposedly most important. Yet these amplifiers continue to receive rave reviews from listeners - we must conclude that the damping factor cannot be as important as is so often claimed. Could it be that *all* owners of valve amplifiers have tin ears, and couldn't tell the difference between a symphony orchestra and a bag of cat litter? This is possible, I suppose, but I do believe that it is somewhat unlikely !

I shall not go so far as to say that the myth is disposed of, but I believe that an extremely low amplifier output impedance is not as important as many people think it is, and that in some cases a small (preferably controlled) amount of deliberately introduced impedance is useful for correcting the characteristics of a loudspeaker driver whose Q is lower than desirable for the enclosure design used. Indeed, I have made such modifications to equipment - raising the output impedance of the amp so that a studio monitor driver could be matched more exactly to the enclosure, and this was at the request of the speaker designer.

Please see the article Effects Of Source Impedance on Loudspeakers for more information on this topic.

For the Thiele-Small Parameter Set ...
This can be a useful trick if the total Q of a loudspeaker is too low (for example, because of a very low Qms - mechanical Q). Such a speaker is normally not suited to a vented box, but by raising the total Q (Qts) by means of increased output impedance from the amp (thus raising Qes), a very satisfactory result may be had.

An experimental plot of a Dynaudio 24W-100 driver shows that an extra full octave is obtainable by raising Qes to 0.9 (from the quoted 0.45). Admittedly, the box required becomes somewhat large, but this has always been the price we pay for extended bass anyway. A similar test on another loudspeaker with a very low Qts showed identical results, with a full extra octave obtainable at the low end, simply by raising the Qes of the driver. The results are shown in Figures 1 and 2.

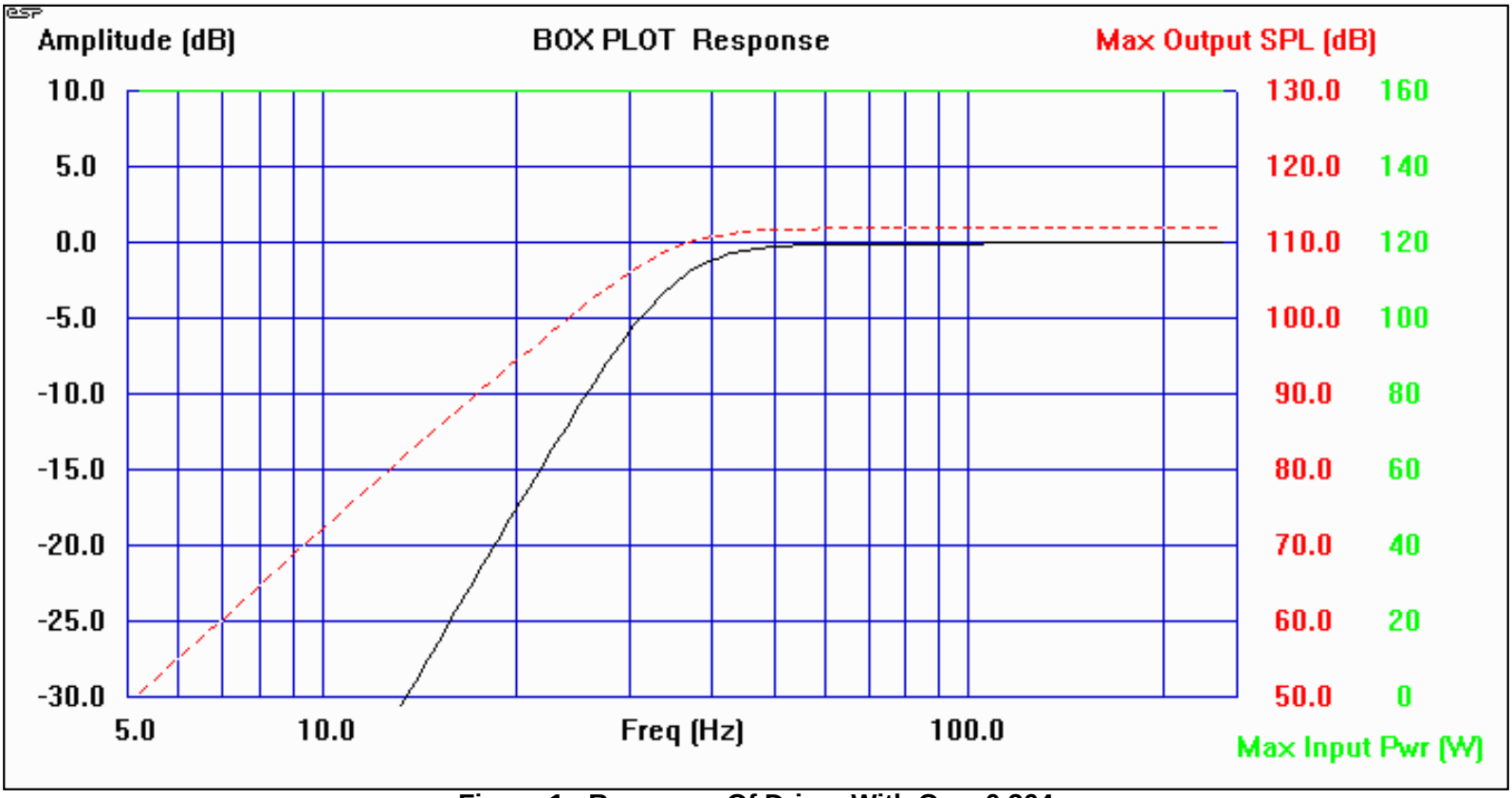


Figure 1 - Response Of Driver With Qes=0.364

As can clearly be seen, the lower -3dB frequency is about 35Hz which although acceptable, but can be improved. Box size is 61 Litres. By raising Qts (using the expedient of increasing Qes with a defined output impedance), we can improve matters in the low end department ....

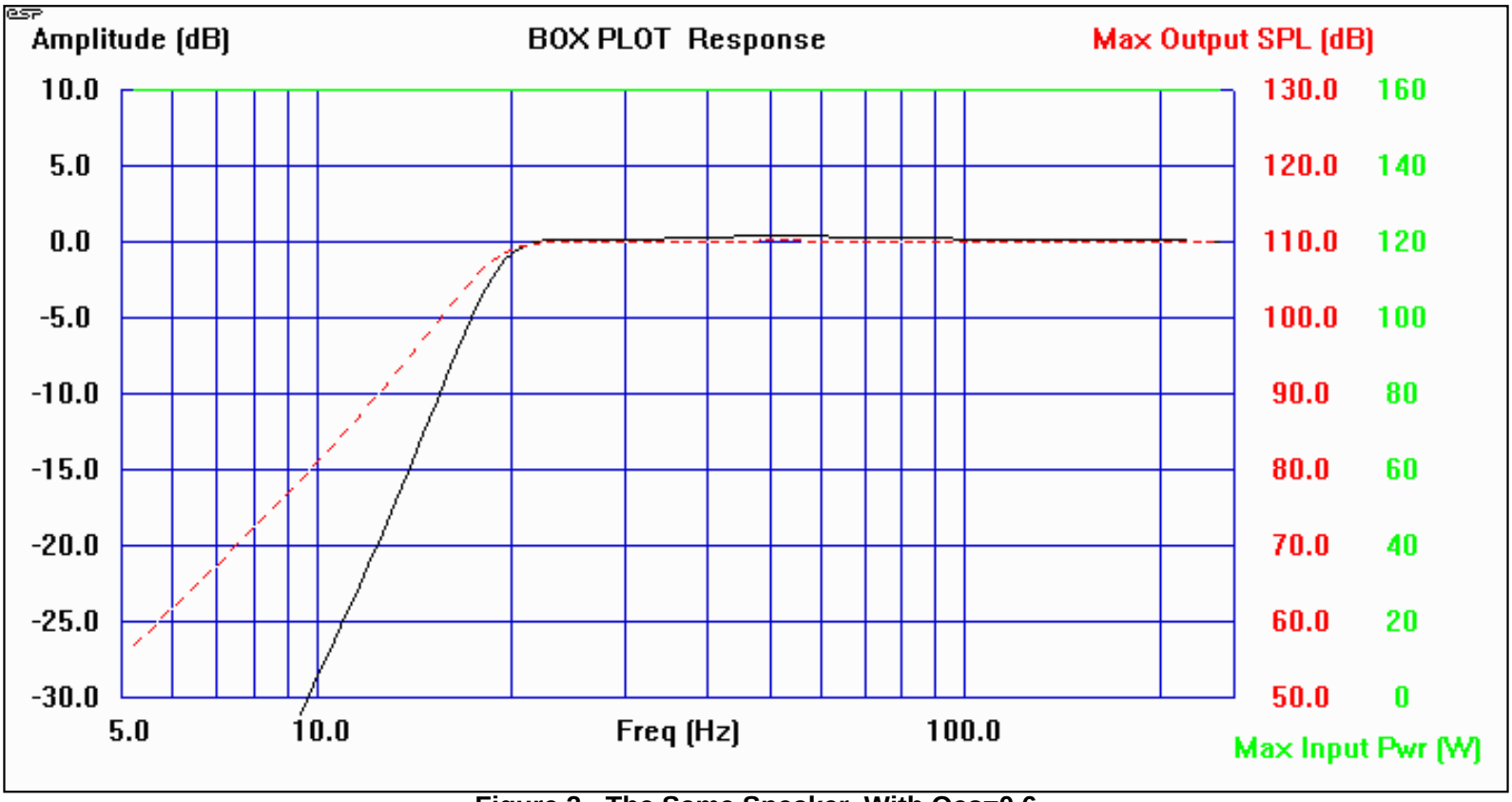


Figure 2 - The Same Speaker, With Qes=0.6

The lower -3dB frequency is now about 18Hz, which represents an extra octave of bottom end. Box size is now 271 litres (just a tad on the large side), but the principle is sound nonetheless. It is worth noting that these two plots are optimised for the given parameters, with no further fiddling with the parameters. BoxPlot (a very useful shareware program) was simply given the details, and made an optimised calculation from the details provided.

Modifying the output impedance is remarkably easy to do in any amp, but naturally the impedance selected must exactly match that required to increase Qes by the desired amount. If it is greater (or less) than optimum, then the box alignment is no longer valid, and bottom end response is unpredictable (at best). So too with boxes which have been designed with a zero Ohm source, since this will never be achieved in practice.

Speaker Impedance And Protection

Speaker systems are rated for a nominal impedance, which will be typically 8 Ohms for hi-fi systems. Many of these claimed impedances are very misleading, since the actual impedance varies widely with frequency. There will be two impedance peaks at low frequencies, corresponding to the driver and enclosure resonance (the upper peak), with the vent and enclosure resonance providing the lower peak. Sealed enclosures will exhibit only the speaker/cabinet resonance peak, since there is no vent used.

Then there are impedance dips, where the actual impedance may fall to 1/2 (or less) the rated impedance. These are nearly always caused by the crossover networks, and can impose a significant reactive load on the power amplifier. It is very difficult to design the perfect passive crossover network, which is a very good reason to use a bi-amplified system with a good quality electronic crossover network, whose characteristics are far more easily controlled than any passive design. (See Bi-Amplification - Not Quite Magic (But Close))

The reactive load imposed on the amplifier by a speaker load causes far higher dissipation in the output transistors than the simple resistive load generally assumed during testing. At the extreme end, consider a load which is completely reactive (i.e. inductive or capacitive). The voltage and current are 90 degrees out of phase with each other, and no power is consumed by the load - even though there is voltage and current present (and measurable). Assuming a voltage of 20V and a current of 2A, the actual power is zero, so the amplifier must dissipate not only the normal internal losses inherent in all power amplifier designs, but the 40 Volt/Amps reflected back from the reactive load. (Volt/Amps - or VA - is roughly equivalent to Watts - but only when the load is resistive, implying that work is performed).

In reality, the reactance is always accompanied by some resistance, so the amount of power converted into work (moving the loudspeaker cone to create sound) will always be non-zero. An additional quantity of the supplied voltage and current are converted into heat (another form of work) due to resistive losses in the voice coil and crossover network. The reactive (also known in electrical engineering as the imaginary) component is reflected back into the output of the amplifier, where it must be absorbed and converted into heat.

It is the reflected power from the loudspeakers which is responsible for a great many amplifier failures. Because of the low efficiencies of most modern speaker systems, more power is needed from the amplifier. This means that the amp will have to dissipate more reflected power and this can lead to overheating (or internal "hot-spot" localised heating) which leads to the destruction of the output transistors. Some amplifier protection systems are sufficiently sophisticated that they can prevent this form of damage completely, but

will generally provide an additional side-effect - the deterioration of sound quality. This is often noticed as a "grainy" or similarly described quality to the sound, and is difficult to eliminate when protection is used. A certain IC power amplifier I have tested has a very comprehensive protection circuit, which seems to work very well. However, as the amp reaches the point of clipping, the distortion component is multiplied tenfold by the protection circuit, with the result that what should be completely inaudible distortion becomes very audible indeed.

It will come as no surprise that I am not a fan of protection circuits in amplifiers, for this very reason. A little care on behalf of the listener to ensure that the amp has sufficient power for the highest listening level desired (plus a bit more for safety), together with being sensible and not shorting the speaker leads, means that protection circuits can be dispensed with in a hi-fi system. This is not the case with lo-fi (Ghetto-blasters and the like), but for true high fidelity sound systems active (or "real-time") protection should be avoided.

Speaker Protection Systems

Speaker protection circuits come in a variety of flavours, the most common being fuses, "poly-switches" and DC offset cutout relays. Looking at each in turn is useful.

Fuses - A fuse is the simplest protection device for speakers, but is not very effective. The fuse is usually capable of protecting a speaker, but is quite incapable of protecting the amp - the transistors will almost invariably blow far quicker than any fuse. The fuse will prevent the fault current from the amp from blowing (or setting fire to) the speakers, provided that it is correctly rated. Basic protection is all that is really offered.

Poly-Switches - These are basically thermistors (thermal resistors) which are normally low impedance, but when their temperature reaches a preset value they go into a high impedance state - thus limiting the current to a safe value. Because they are essentially a non-linear device, and have some resistance even when not activated, I cannot recommend that they be used for hi-fi applications. Their internal resistance adversely affects damping factor (which may be a "so-what"), but the fact that they are non-linear means that some degree of distortion must be introduced by their inclusion. The degree of added distortion is bound to be dependent upon how close to their limits they are being run, but in no case could this be considered negligible. When we are striving for amplifiers with less than 0.01% distortion, it takes very little non-linearity to raise that by an order of magnitude.

DC Offset Relay - This is one item that has no bad habits, provided that the relay contacts are capable of handling the output current of the amplifier without overheating. This is not as trivial as you might think, since few small (and/or reasonably priced) relays have a suitably high current rating, and many do not have sufficient contact separation to break the arc created by 50V DC being dumped into a speaker load.

With some of the relays I have seen used, one can almost guarantee that the relay will be destroyed if it were to be called upon to do its duty. However, the basic idea is sound, and introduces no distortion or other undesirable effects to the final signal. With a little extra circuitry, the relay can also be called upon to open under other fault conditions (such as excessive output current), and the only characteristic will be that the sound cuts out completely while the fault is maintained. This is a better solution than trying to provide active protection with its resultant sonic degradation.

Impedance Matching

If you are into RF (radio frequency) design, or perhaps telecommunications, this becomes an important topic. For hi-fi audio, it is a meaningless concept and will actually cause an increase in noise and distortion.

It has often been claimed that a 600 Ohm microphone should be matched to a 600 Ohm input for best performance. This is simply wrong, and microphone manufacturers specifications will support me on this.

Imagine a 600 Ohm microphone with an open-circuit output voltage of 5mV. If the mic preamp has an input impedance of 600 Ohms, the microphone output is reduced by 6dB to 2.5mV because of the simple voltage divider created. It helps to use the engineering "model" for a signal source of any kind, which is basically a "perfect" (meaning zero Ohms impedance) voltage generator, with a resistance + inductance or capacitance (sometimes in combination) in series with the output.

If the output is loaded, then the available voltage from the source drops, and that in turn means that more amplification is needed to obtain the final voltage needed. If the output is reduced by 6dB, this means that an additional 6dB of gain is required to compensate - therefore the circuit will have more noise.

The ideal for a microphone is to use a high impedance input, but this creates other problems, so a compromise is needed. Typically, a good mic preamp (for microphones of up to 600 Ohms) will have an input impedance of between 1.2k and 2k Ohms. This causes far less loading, and does not cause any problems for the microphone.

Generally, it is desirable that the output impedance is low, and the destination impedance high, and this is the case with the majority of modern equipment. Preamps usually have an output impedance of less than 1k Ohm, and power amps will have an input impedance of at least 10k Ohms, but more commonly 22k or 47k.

So why is impedance matching important for RF and telecommunications? The reasons are completely different, as we shall see.

Radio Frequencies: When an RF voltage and current are transmitted along a wire, the impedance of the cable itself becomes significant, and for any distance that is "significant" - which is to say any distance greater than about 0.1 of the signal's wavelength - matching is necessary. The wavelength is calculated from the speed of light (3 x 10 ^ 8 m/s, or 300,000km/s) multiplied by the "velocity factor" of the cable. This varies from about 0.7 up to 0.9, meaning that a signal travels more slowly in a cable than in free air or space.

Wavelength = C / f (where f = frequency and C = velocity)

A 1Mhz signal travelling in a typical coaxial cable (velocity factor of 0.8) will have a wavelength of ...

Wavelength = ((3 x 10 ^ 8) x 0.8) / 1 x 10 ^ 6 = 240m

Based on this, any attempt to transport a 1MHz signal further than about 24m will start to cause problems unless the send and receive impedances are properly matched - not only to each other, but to the cable as well.

In the hi-fi audio world, this is not an issue, since this is 50 times the highest frequency we can hear, and

few instruments create appreciable harmonics above 20kHz anyway. In theory, we could send an audio signal 12km without having to worry about impedance matching.

Telecommunications: Impedance matching is a requirement in telecommunications networks, but not for any of the reasons you might think. It is actually rare for an analogue phone line to run more than about 5km from the exchange (C.O.) to the user's location.

The impedance matching required is to enable the hybrid - a circuit that allows simultaneous transmit and receive on a single pair without interference - to function properly. If the impedances are not properly matched, you will hear too much of your own voice when you speak, and the far end speech will be too soft.

Because of the distances involved, the telecommunications network is balanced, to prevent noise.

Balanced Vs. Unbalanced

The conventional hi-fi connection is unbalanced. One conductor is earthed (the shield), and the other carries the signal. Except that this is rubbish. The shield also carries the signal, since without a return path, we have an "air-gap" problem, and no signal will pass - other than general noise and maybe a tiny bit of the desired signal due to capacitive coupling between the two circuits.

This unbalanced signal is actually fine for short distances - few hi-fi interconnects will be longer then a metre or so. Provided the source impedance is low, there will be very little noise introduced, and the signal will pass unscathed from one piece of equipment to the next.

However, if 10mV of noise were to be picked up along the way, then this is added directly to the wanted signal, adding hum or other undesirable noises to your music.

In contrast, a balanced connection uses two wires, one carrying the normal signal, and the other an inverted (but otherwise identical) version of the signal. At the other end, the two signals are recombined. Any noise that was picked up along the way will be seen to have the same polarity on both wires, and is not "seen" by the receiving equipment. Only signals that are in "anti-phase" (of opposite polarity) are picked up.

This noise on both wires is called common-mode noise, and a properly balanced circuit will amplify the wanted signal, but reject common-mode noise by as much as 60dB (i.e. 1/1000th of the noise gets through). This is known as common-mode rejection ratio, and is a figure quoted by all opamp manufacturers for the opamp inputs.

If the same 10mV of noise were picked up by a balanced cable, only 10uV of unwanted signal will be present at the output of the amplifier if the common mode rejection is 60dB. To look at it in a different way, the signal leads will have to "collect" 10 Volts of noise to allow a 10mV noise signal to pass.

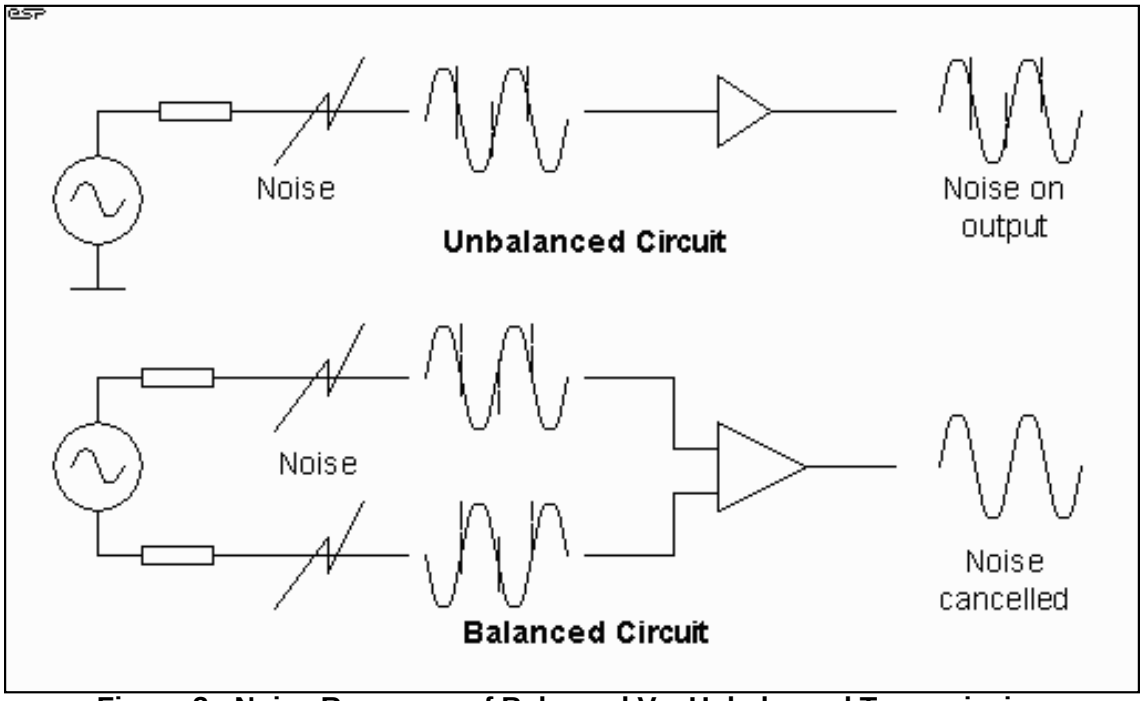


Figure 3 - Noise Response of Balanced Vs. Unbalanced Transmission

Figure 3 shows how the noise is cancelled in a balanced circuit. The noise pulses are applied to both leads - it is important that the leads are twisted together, to ensure that any noise is picked up by both. Because the signal is 180 degrees out of phase and the noise is in phase, only the out of phase signal is allowed through the amp. Any in-phase signal (common mode) is rejected, cancelling the noise signal almost completely.

In professional audio work, balanced leads are used almost exclusively. These are always shielded as an additional protection against noise.

Hi-Fi Super Cables

We have all seen advertisements for audio interconnects costing obscene amounts of money, and some of them actually do seem to sound better than "ordinary" interconnects. There is no "magic" here, these are often no more than a "pseudo balanced" cable, where the shield is connected at one end only, and the signal is carried by a pair of wires within the shield.

This is hardly worth all the money the hi-fi shops ask, since you can make them yourself, with good quality cable and connectors available from any electronics dealer.

As for so-called "directional" cable, this is utter rubbish. The only thing that is directional is the choice of which end the shield should be earthed at - send or receive. For cables having the shield connected at only one end, the answer is usually the sending end (e.g. the preamp, when connected to a power amp). It is worth pointing out that shielded cables should *always* have the shield connected at both ends, or noise reduction is imparied.

How can a cable carrying an AC signal be directional? There are some proponents of the oxygen free copper concept who will claim that if there is oxygen in the copper, it will be as copper oxide, and that copper oxide is a semiconductor, semiconductors rectify, and will therefore introduce distortion. Prove it to

me!

I defy anyone to produce concrete proof in any form that a cable is capable of introducing distortion - at any level. Even if we assume that there is some validity in the copper-oxide rectifier theory, all semiconductors have their forward conduction voltage (e.g. 0.65V for silicon) - I don't know what it is for copper oxide, but even if it were as little as 100mV (highly unlikely!), this would require that there was a 100mV or more difference between adjacent conductors for rectification to occur. Since the loss along a 1m length of signal interconnect cable will be a very small fraction of this, I (and many others) do not consider that this is in any way possible.

So, if there is no validity at all in spending $200 for an interconnect cable, why do people say how great they sound? Easy. If you had just spent that much and could hear no difference, would you be game to admit it to anyone? No, you will be tempted to try to convince yourself that you *must* be able to hear *some* difference, otherwise you just wasted $190, since a $10 cable would work just as well.

As a matter of interest, you might also want to have a look at my editorial page, which has some challenging things to say about $3000 mains leads (and no, that is not a misprint!).

Conclusions

It is quite possibly obvious that I do not believe any of the hype about cables. My own system uses perfectly ordinary cable throughout, and I have no desire to change this in any way whatsoever.

I am utterly unconvinced by any claim that a cable can make my hi-fi sound any different - never mind better, just different. I have never been able to measure distortion in a cable, and I know of no-one who has.

Certainly, some "super" speaker cables (because of high inductance or capacitance) can make a power amp unstable or even oscillate, but this is invariably a bad thing, so I don't want any of that stuff at any price.

One final point - gold-plated connectors. They do not conduct any better than any other type (worse than some), and they do not introduce (or remove) a "sound". Gold is used because it does not oxidise, so the connections don't have to be jiggled about every so often to remove crackles or other mechanically induced noises - no more and no less than this. (Although I have seen a claim that solder joints to gold can simply separate for want of anything better to do at the time. I have yet to see this proven, but it is possible.)

Main Index

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Update Info: 28 Jan 2000 - Corrected some errors and typos./ Updated page layout 28 Nov 1999./ 24 Jan 05 - Added amp impedance info.

The Audio Pages

Elliott Sound Products **Microphones**

Copyright © 2006 - Rod Elliott (ESP)
Page Created 10 Jan 2006

Articles Index

Main Index

Contents

- 1.0 - Introduction
- 2.0 - Basic Microphone Characteristics
  - 2.1 - Directional Patterns
  - 2.2 - Transducer Types
  - 2.3 - Specifications
- 3.0 - Sub-Types of Microphones
- 4.0 - Element Construction
- 5.0 - Powering Schemes
- 6.0 - General Usage Guidelines
- 7.0 - Conclusion
- 8.0 - References

1.0 - Introduction

Microphones are often poorly understood, and this article seeks to provide some basic details about the various types, how they work, and the interfacing of the microphone with a suitable preamplifier. While it would be tempting to explain mic techniques, proper usage, etc., these are not topics that will be covered. There are several reasons, but the main one is that there are so many possibilities that it is impossible to cover them all.

Instead, the focus will be on the microphone basics - how each type works, along with its advantages and disadvantages. The following brief summary is a warm-up for the real thing - and even though it looks like a lot to cover, there are (and will remain) several omissions. For example, carbon microphones will not be covered because they are no longer used in new equipment, and 'esoteric' microphones (such as the so-called shotgun mic) will not be explained in any detail.

With microphones, the terms Directional, Cardioid, Omni-directional (or just omni), Hyper (or Super) Cardoid, etc. refer to the polar response, but these terms are sometimes loosely applied. The directivity of all microphones is frequency dependent, and becomes spherical (omni-directional) as the frequency decreases. There are exceptions, and these will be looked at as we progress.

The microphone, abbreviated to 'mic' or 'mike' is an essential part of the process of getting our music from the performers to our listening rooms. Mics are also used for sound reinforcement, ensuring we can hear everything at a concert (and also often ensuring that we can hear very little for hours afterwards). Correct microphone selection and placement during recording minimises the amount of equalisation that is needed, because the sound is already the way the producer intends. The choice is enormous, as the brief summary below indicates.

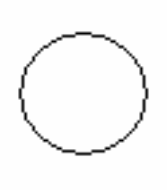
2.0 - Basic Microphone Characteristics

Although the number of different microphones looks daunting, they are all based on common parameters ... these are directional patterns and transducer types, and almost every microphone made is covered by the two listings below.

2.1 - Directional Patterns

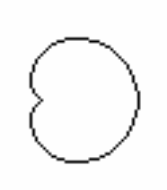
The directional characteristics of microphones are defined in the capsule (or capsules, in the case of dual capsule mics). Contrary to what some may claim, any type of microphone can be configured to have any of the listed configuration patterns. The directional characteristics are frequency dependent, and refer to the free field response - placing a microphone very close to any surface changes its directional characteristics, and they become unpredictable because of the almost infinite number of possibilities. Directional microphones are also called 'pressure gradient' mics, because their directional characteristics are created by means of varying pressure to the front and rear of the diaphragm (the pressure gradient).

Omni-directional ...



Pick up sound (more or less) equally from any direction. Omni-directional refers to the frequency response being essentially flat, regardless of the direction of the arriving sound waves. Omni mics often give fewer feedback problems compared to most cardioids, but this is highly dependent on correct usage. Omni mics have minimal proximity effect, and are (generally) better suited to instruments. These mics are not commonly used for live production because of limited understanding.

Cardioid ...



The most common directional pattern. These usually have a proximity affect that colours and enhances the bass end of vocals at close range. Different cardioid mics may suit male and female singers. Singers should own their microphones and be skilled in the techniques of using them, in the same way that musicians own their instruments. Cardioid mics are often misused for instruments, typically used in very close proximity to drum skins (among other misuses). Naturally, if this gives you a specific sound that you want, then it is no longer misuse.

Hyper-Cardioid ...

This is an exaggerated version of the cardioid mic, so it is more directional. A side-effect is that a small lobe is created at the rear of the microphone, so these mics must never be 'aimed' so that the rear lobe points towards a floor monitor (for example). Sometimes a distinction is made between 'super' and 'hyper' cardioid microphones, but other descriptions will consider them to be equivalent.

Figure-8 ...

The figure-8 mic picks up sound equally well from the front and back, but rejects sound coming from the sides (as well as top, bottom, etc.). The pattern can be looked at as an extreme form of hyper-cardioid, where the front and rear lobes are equal in amplitude and frequency response. Many dual element microphones combine an omni and figure 8 capsule to allow switchable directivity.

2.2 - Transducer Types

At the heart of every microphone is a transducer - simply a mechanism that converts one form of energy to another. The source (input) energy is sound, and the output is electricity. An electrical waveform is produced, which matches the acoustic input with as little modification as possible. All directional microphones must (by definition) alter the received sound to some extent. It is not possible to modify the directional characteristics without also altering the nature of the sound that is picked up. This is not necessarily 'bad', just different.

Likewise, many cartridge / capsule types (the actual transducer) have their own sound, whether real or imagined. This often influences the choice of microphone type for different tasks - for example, there are mics that are favoured for bass (kick) drums that may be deemed unsuited for anything else. This is not necessarily true, as experimentation can often demonstrate.

Condenser ...
More correctly called capacitor mics, these are generally considered to be the ultimate. They have exceptional detail, and can usually tolerate very high sound levels. Distortion is very low, because the diaphragm movement is so small (comparable to that of the human ear drum). Capacitor mics most commonly use a high DC voltage to polarise the 'plates' of the capacitor sensor, although some use the change in capacitance to modulate a radio frequency oscillator. The frequency modulated 'carrier' is then fed to a detector stage to be converted back to audio. Capacitor mics (of all types) require power - this may be supplied via the P48 (48V phantom feed) from a mixing desk, or may be an external power supply. Probably the most famous of all capacitor mics was the Neumann U47.

Dynamic ...
The dynamic mic uses a mechanism that is very similar to speakers. The majority are robust and can accept extreme sound levels, ideally suited for live productions. Most are cardioid, although omni-directional and hyper-cardioid types are also available. Dynamic mics are the most common of all types used in live work, and they are often used for studio recording as well. One of the best known is the venerable Shure SM58.

Electret ...
Also called 'Electret Condenser' or 'Electret Capacitor' microphones use a permanently 'charged' plastic membrane so that a high voltage polarising signal is not needed (as is

the case with true capacitor mics). Most are omni-directional, although cardioid inserts are also made. Like capacitor mics, an impedance conversion stage is essential because of the extremely high intrinsic impedance. While electrets can be used for stage work, they may distort at high sound pressure level (SPL). Many vocalists are capable of driving electret mics well into distortion. Temperature and humidity (such as from the breath of vocalists) can adversely affect them. Professional electret microphones are excellent for recording.

Ribbon ...
These are common in recording studios, but less so in live work because they are comparatively fragile. A very thin (usually aluminium) ribbon is suspended in an intense magnetic field, and generates a small current when it is moved by sound. Ribbon mics have extremely low impedance, typically (much) less than 1 Ohm. A transformer is used to raise the impedance (and output voltage) to a usable level. Although Ribbon mics have an inherent Figure-8 pattern, they are also available with cardioid or hyper-cardioid patterns. 'Planar ribbon' mics are a variation of the theme. These use a thin membrane with a planar (flat) coil deposited on the membrane.

Carbon ...
These microphones used to be very common - every old style telephone had one. The carbon mic has one major advantage over every other type - it has gain! Because the microphone element is made up of carbon granules, speech activating the diaphragm will compress and release the granules, changing the resistance significantly. Since power is needed by these mics, this is provided by the telephone line. The microphone gain is such that no additional amplification is needed to allow a normal phone call - even over a considerable distance. In the early days of telephony, this was essential to the operation of the 'phone network - so much so that if it were not for the carbon microphone, the telephone would never have been even useful (let alone gain acceptance) in those early days. Cheap and reliable amplification has made them redundant now.

As a short side-note, it is worthwhile mentioning that the telephone system uses a nominal 48V supply (see phantom feed, below). The influence of telephony on electronics as we know it is huge - so much so that the development of the phone system drove many of the inventions that we now take for granted. Have a look at the vast contribution of Bell Laboratories (which used to be an integral part of AT&T). Bell labs invented the transistor - the very cornerstone of every electronic product we use.

2.3 - Specifications

The output level of microphones should be rated in millivolts per Pascal (mV /Pa), although there are many variations. Other conventions used include dBm at 0.1 Pa (this will always be a negative number). All new microphones should be rated in mV / µPa, but the older standards persist in some countries.

1 Pascal = 10 micro-Bar = 94dB SPL
0.1 Pascal = 1 micro-Bar = 74dB SPL
1 dyne/cm2 = 0.1 Pascal = 1 µbar

There are also noise ratings (these vary widely, both in output noise and the way it is specified), output impedance, recommended load impedance, polar response, frequency response, etc. Frequency response claims are meaningless without a graph showing the actual response, and for directional mics this should also indicate the distance of the mic from the sound source. Cheap microphones are particularly bad in this respect, and it is not uncommon to see the frequency response stated as (for example) 50 - 20,000Hz. Because no limits are quoted (such as ±3dB) this is pointless - any microphone will react to that frequency range, but may be -20dB at the frequency extremes, with wide variations in between.

A proper graph showing the response at all frequencies will quickly show the actual response, although it is uncommon for any manufacturer to state the distance between mic and sound source, or the method used to take the measurement.

A polar response graph will also show the directivity at a number of different frequencies. As frequency decreases, the directional pattern commonly approaches omni-directional, although some mics maintain excellent directivity even at very low frequencies (the secret is in the rear chamber).

3.0 - Sub-Types of Microphones

There are some microphones that appear to be completely different from those described above. Not the case, as the essential characteristics and transducer types don't change, but more / different hardware or electronics are added to give additional functionality.

RF (Transmitting) Mics ...

These are available in many variations and professional types can be very expensive. A conventional microphone transducer having one of the directional characteristics listed

above is connected to a small radio frequency (RF) transmitter so the mic can be used without the need for cables. The transmitter for professional radio mics requires excellent frequency stability, and receivers are highly specialised to ensure no 'drop-outs' and maintain a good signal to noise ratio (SNR) at all times.

These mics require specialist knowledge and experience to use them correctly. Many have automatic limiting and compression that has to be managed carefully, because compression limits the dynamic expression of good singers, causing them to sound comparatively flat and lifeless.

PZM™ ...
The Pressure Zone Microphone™ (also known as a boundary mic) is a special application of the electret mic. A miniature electret sensor is mounted a small distance (typically less than 1mm) from (and facing towards) a flat plate. They are often used on the floor or walls, tables (for conferencing and the like), but can also be attached to large flat discs or plates. They have exceptional performance, and can effectively reduce reverberation if used carefully.

There are several variations on the basic technique, allowing for a single stereo mic unit, a cheap 'knock-off' made by Radio Shack (Tandy in Australia) called a boundary mic, but lacking the characteristics of a true PZM, and a few others.

Dummy Head ...
The dummy head mic technique yields extraordinary performance, but the recording can only give the full effect when listened to through headphones. Electret mic capsules are either embedded in a true dummy head (wig-carriers can be used ... meet Yorick below), or miniature capsules are worn in the ears of the sound recordist. When played back through headphones, the original sound field is essentially restored, and the listener hears the sound as if s/he were there.

The requirement for headphones has limited the appeal of the technique.

Shotgun ...
Shotgun mics are worthy of a complete article to themselves. Usually fairly long, they have an extreme directional pattern, and typically only pick up sound from the general vicinity directly in front of the mic. There are several ways to make shotgun microphones - techniques include a long 'barrel' with slots designed to create an interference pattern that rejects sound from the side, and multi-element designs with phase and amplitude balance between elements. An old method was to use multiple thin tubes of differing lengths, arranged so that the longest tube is in the centre, with smaller tubes

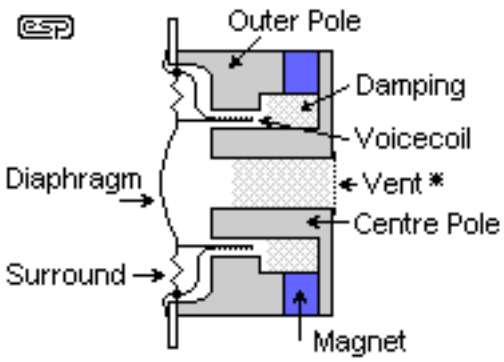
surrounding it.

Some shotgun mics use a combination of methods, as well as careful attention to the mic capsule's rear chamber. These mics are useful for location sound recording (for movies or TV), nature recordings, and anywhere else where very high discrimination is needed.

4.0 - Element Construction

The following is intended to give you an idea of the basic techniques used to make various microphone elements. These are the basic building blocks, and while some (such as dynamic mics) can be used with no additional circuitry, most others require some additional components to be useful.

Because the dynamic mic is one of the most prolific (or so it would appear to the uninitiated), it will be covered first.

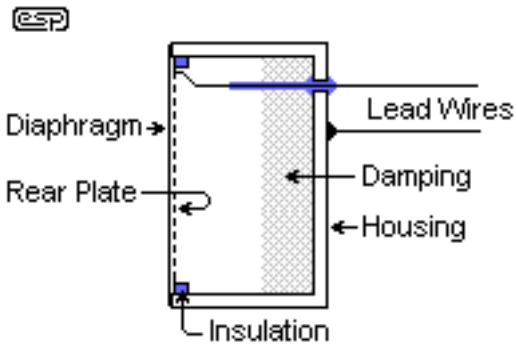


The general arrangement of a dynamic mic is shown to the left. A diaphragm is coupled to a voicecoil that is suspended in a strong magnetic field. As the diaphragm (and thence the coil) moves in sympathy with the arriving soundwaves, an electric current is generated. In a perfect microphone, the electrical current will be an exact replica of the acoustic signal, but in reality this is never the case.

The element (also known as a capsule) shown has a vented pole-piece (indicated with a *), and this is typically done to create the required directional characteristic. For an omni-directional dynamic microphone, the back would be sealed.

As you can see, this is very similar to the construction of a small speaker, and indeed, a speaker will work as a microphone (and a dynamic mic can also make noise). Naturally, the speaker and mic are each optimised for their intended application, and neither works particularly well when its role is reversed. 99% of intercom systems use the speaker as a microphone.

Typical dynamic microphones have an impedance of around 150 - 300 ohms, although some are higher or lower than that. While it may seem tempting to match the impedance of the microphone and preamplifier, this is ill advised, as it will reduce the signal level by 6dB, and thus reduce the signal to noise ratio.



A capacitor microphone is much simpler mechanically, but the material quality is critical for good performance. Because the capacitance is so small, the insulation resistance must be very high, as must the impedance of the following stage. It is not uncommon to find in excess of 100 Megohms input impedance for the impedance conversion stage.

While the capsule shown has damping material, this may not always be the case. The distance between the diaphragm and the rear of the housing can be made small enough so that no ill effects occur within the audio range.

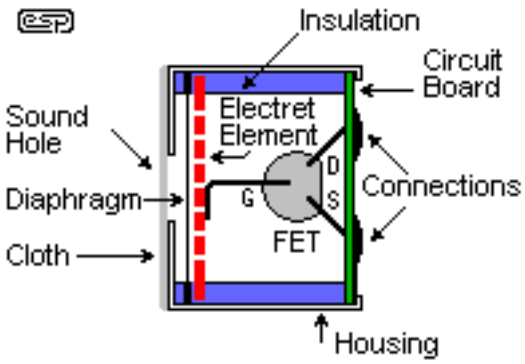
The backplate must be polarised so the microphone will work. While this may be as low as 48V, this may not be insufficient to allow a worthwhile signal level. Voltages up to 200V will be found in some examples. This places great constraints on the insulation, and means that such mics can be adversely affected by moisture.

In some cases, the microphone capsule may have two diaphragms, each spaced as close as possible to the backplate. This will create a microphone with a figure-8 directional pattern. The one shown is omni-directional - this may come as a surprise because sound coming from the rear of the mic is shielded from the diaphragm by the mic itself, but this only applies at very high frequencies. Many Neumann mics use a dual diaphragm capsule, and switch one diaphragm to change the directional characteristic from cardioid to omni-directional.

The diaphragm of capacitor mics must be conductive, and it is common to use metallised plastic film (Mylar is popular). The metallisation film must be protected from moisture, so may be on the inside of the capsule. In almost all cases, the insert will have a tiny bleed hole to allow the air pressure inside the housing to match that of the outside atmosphere. If this were not provided, the diaphragm to backplate spacing would vary with atmospheric pressure.

Electrically, a capacitor mic can be represented by a signal source in series with a capacitance equal to that of the capsule itself. As noted above, this will be very low. A typical capacitor microphone (such as the Neumann U47) has a capacitance of around 80pF (See References.)

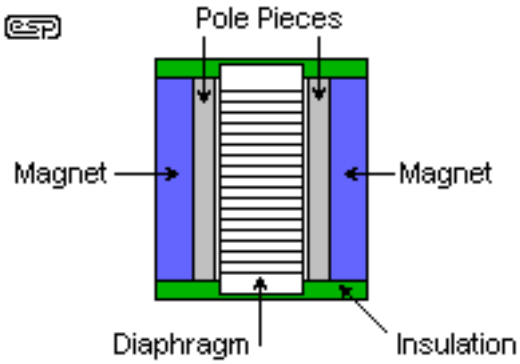
Historically, these mics have been known as 'condenser' mics for many years. 'Condenser' is the old term for a capacitor.



Electret microphones work using the same general principles as a capacitor mic. Instead of using a DC polarising voltage, the backplate is an electret material (this is a so-called 'back electret'). This material is a plastic that is subjected to an intense electrical field during processing. This causes the plastic material to retain a charge (more or less) permanently. The electret surface must be metallised to make it conductive. Some electret mics use the diaphragm as the electret element (and use a conventional backplate), and while this works very well, they do not have an indefinite life.

The FET shown is almost always included in the capsule itself. This is the impedance converter, and in most cases there is no resistor from the gate to common (ground, mic housing). This is one reason that electret mics can react badly to a sudden loud sound, and may lose sensitivity for a few seconds. The FET gate circuit relies on surface leakage alone to bias the FET correctly.

While I stated earlier that dynamic mics seem to be the most common, they are soundly (pun intended) beaten by electret microphones. All modern telephones use electret mics, including mobiles (AKA cell phones), answering machines, computer headsets, and virtually every piece of electronic equipment that needs to hear voice commands, noise, etc. The electret has been the most successful mic capsule ever developed - over 100 million are produced every year !



The ribbon mic has a special place in the heart of many a sound engineer. They have an inherent figure-8 pattern, although this is often modified to produce more 'conventional' patterns. Because the impedance of the ribbon is so low, all such microphones use a transformer to raise the impedance and output voltage to more usable levels. The transformer is almost always in the same housing as the microphone element itself.

Ribbon mics are often though to be fragile, and many of them are. There are others that are very robust indeed. Because ribbon mics use a relatively large diaphragm (much larger than most other mics), they can be very sensitive to air movement - even at subsonic frequencies.

However, high SPL does not usually bother a ribbon mic in the least. Provided the ribbon remains in the gap, almost nothing will cause a ribbon mic to distort - apart from the aforementioned air movement which *must* be avoided. Even apparently gentle air movement can distort the ribbon, which then must be replaced. 'Planar' ribbons are used by some manufacturers - a planar ribbon is not a ribbon in the true sense of the term, but uses a metallised coil printed on a thin plastic carrier. These are very rugged according to the literature.

Because of relatively low output level (even after the transformer), you need a very quiet preamp for ribbons. They have very low self noise, so preamp noise can easily exceed the microphone noise.

5.0 - Powering Schemes

For those microphones that require power, the most common option is phantom feed (P48). This uses a nominal 48V DC applied to both signal leads via a 6.81k resistor. A good example of a P48 powered microphone is described in Project 93, and Project 96 describes a 48V power supply and P48 distribution scheme. For the sake of completeness, Figure 5 shows the general arrangement of a 48V phantom feed system. Although the feed resistors are shown as 6.81k, 6.8k resistors can be used instead. It is recommended that they be matched to within 0.1% so common mode rejection is not compromised.

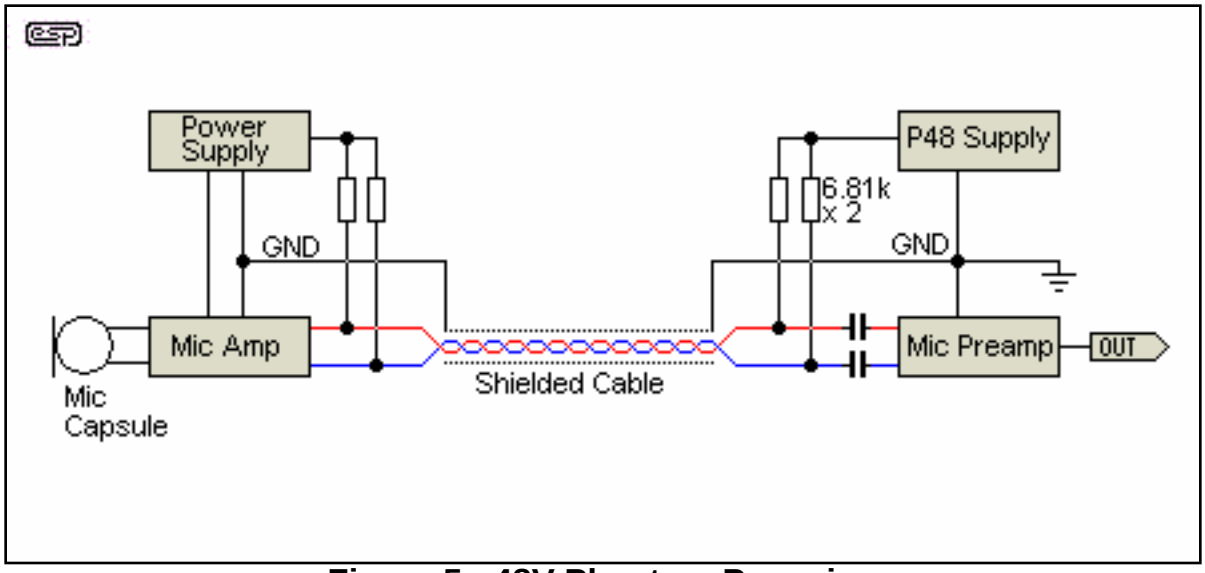


Figure 5 - 48V Phantom Powering

Although the phantom feed supply voltage has been standardised at 48V, there are many supplies that do not comply, with some operating at 30V or even less. While mics designed for P48 power might work with these low supplies, they may not. In general, phantom feed power supply *must* be able to supply 48V. The accepted voltage range for P48 is between 38V and 52V. A 'new' sub-standard has arisen, called P24 (20V - 26V), but this is (IMHO) a seriously retrograde step, creating potentially disastrous incompatibilities between competing standards.

Some time in the late 1960s, Neumann (of microphone fame) converted its valve (tube) capacitor microphones to solid-state. They decided upon a remote powering system that they called Phantom Power, and this was a trade mark of Neumann. Although other manufacturers originally avoided the trade mark (using terms such as 'simplex' instead), with time the term Phantom Power has become generic. DIN standard 45596 describes the powering of any device that uses the P48 phantom powering scheme.

Because phantom power is a common mode signal (it appears equally on both mic leads), plugging a balanced dynamic mic into a 'live' P48 powered mixer channel will not harm the microphone. The mic may make strange and/or loud and/or rude noises if the internal insulation is degraded (by age, saliva, beer, rum+coke, etc., etc.). In general, it is better to switch off the P48 supply unless it is needed.

Phantom powering is not the only way that power is supplied to microphones. Another standard is called T12 - as well as transverse feed, A-B powering, parallel powering, and occasionally by its full name ... 12V Tonader (it originated in Germany). It is not commonly found outside the film industry, and is totally incompatible with P48 powering. Adaptors can be fabricated, but require a transformer.

The T12 system uses 180 ohm feed resistors and a 12V supply, but the DC is not sent as a common mode signal like phantom feed. Referring to an XLR mic connector, the positive DC is applied on pin 2, negative on pin 3, and earth (ground) on pin 1. *However*, there is also a reverse version, with positive on pin 3 and negative on pin2. T12 powering will probably damage dynamic mics that are inadvertently connected while the T12 power is on.

Capacitor microphones using valves (tubes) will almost always require a special outboard power supply, and multi-pin connectors are common. Because of the current needed by the valve heater, the 2 - 4mA available from P48 is completely unsuitable. These power supplies will be specific to the microphone - as far as I know, there is/was no standard adopted by manufacturers, so each will be different.

6.0 - General Usage Guidelines

For live applications, the number of 'open' microphones (i.e. connected and picking up sound) should be kept to minimum. Unnecessary use of a large number of open mics creates excessive comb filter distortion. This reduces intelligibility and increases feedback problems. There are many recommendations that you may find - you may be advised to minimise the number of different microphones, for example. Exceptions are (directional overhead) for percussion and (dynamic high velocity) for bass drums. Placing any mic too close to an instrument, sound source or surface affects its response. This effect may be good or bad, depending on what you are trying to achieve.

There are many sites on the Net that give some general idea of what microphone to use where, but these are mainly a matter of opinion. Everyone who uses mics has different ideas on optimum placement and type. Some are reasonable, a few good, and a lot that are (IMO) just plain wrong. One thing that is almost never mentioned is that where you place a microphone may change its characteristics.

If a mic is placed very close to a surface (be it a wall, floor, drum skin or singer's face) it will no longer have the directional characteristics you purchased it with. Likewise, holding a mic in such a way that your hand cups the back of the mic ball will change directionality radically and unpredictably.

Something that is not well understood is just how much signal you can get from a microphone. A typical dynamic mic is easily capable of 0.5V RMS (500mV) when held close and singing (or in my case yelling) loudly. This may seem extreme, but look at the specification for the SM58 as an example. 1.85mV at 1 Pascal (94dB SPL), so 185mV

results at 114dB SPL - anyone can yell that loud at close range. You will get 500mV at just under 123dB SPL. While this may seem pretty extreme, many vocalists can achieve such levels at close range - good mic technique includes 'pulling back' from the mic when singing loudly, and getting in close for soft passages. This is a vocalist's natural compressor, but many singers don't have any mic technique at all (there seems to be an increasing number that don't have any singing technique either, but that's a different matter).

At these levels, you can completely forget using electret mics, as they will just distort badly. Because sensitivity is much higher than a typical dynamic mic, the mic may attempt to produce perhaps 3-5V RMS at the same SPL (123dB), and this is not possible with standard electret capsules. This is especially the case if it is powered by a 1.5V battery! Such mics are very common (and very useless for most applications).

7.0 - Conclusion

This article has only scratched the surface, but is a good starting place. Although there are a great many variations, the details above cover the majority of microphones in general use.

As an experiment, I was recently forced (i.e. it was something I'd been planning to do for well over a year) to build Yorick (as in "Alas poor Yorick, I knew him well" - Shakespeare). Yorick is a dummy head microphone, and details are available in Project 112 so you can build your own version. Tests are very encouraging, with an amazing ability to locate the sound source.

Yorick - My Dummy Head Microphone System

Please note that any resemblance between 'Yorick' and a certain well-known entertainer (who seems rather fond of cosmetic surgery) is completely coincidental .

Although you can purchase a Neumann dummy head mic already made, I suspect the price will be a fairly strong deterrent. There are other methods of achieving much the same result, but there is something rather nice about having a 'real' head rather than a plastic or MDF disc with mics on each side (the hair is optional of course). Each capsule uses a P93 mic amplifier board. In my case, I already had a suitable preamp that is multi-purpose, but the P93 mic amp is the easiest way to build the unit.

7.0 - References

1. Phantom Powering for Condenser Microphones - http://www.uneeda-audio.com/phantom/index.htm
2. Shure SM58 Data Sheet - http://www.shure.com/pdf/specsheets/spec_wiredmics/sm58.pdf
3. Microphone History - http://history.acusd.edu/gen/recording/microphones1.html
4. Neumann U47 Schematic - http://www.vintageking.com/site/files/images/u47.gif

---

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2006. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 10 Jan 2006

The Audio Pages

Elliott Sound Products **Why Do Tweeters Blow When Amplifiers Distort?**

Copyright © 2001 - Rod Elliott (ESP)
Page Created 28 July 2001

Articles Index

Main Index

Contents

- Introduction
- Power Distribution
- Overdrive Conditions
- Example Power Waveforms
- Excessive Compression

Introduction

A vexing question, regularly asked and rarely answered properly - Why Do Tweeters Blow When Amplifiers Distort? The answers are actually quite simple, but the common misconception is that the distortion creates harmonics, and the additional harmonic content destroys the tweeter.

Not really - woofers and midrange drivers can also blow from a distorted amp, and this is unlikely to have anything to do with harmonics. Certainly, there are additional harmonics generated, and they will be at relatively high levels, but rarely high enough to cause more than relatively minor stress to the tweeters (in particular).

It is also worth looking at the article Amplifier Clipping, as this provides some additional

information that is not covered below. Although it usually does not affect tweeters, the effects on low frequency drivers can be very harmful.

Because of the nature of music and the over-use of compression, it is not only tweeters that are at risk. Woofers can also be damaged, due simply to excessive continuous power.

Power Distribution

A great part of the mystery is uncovered when we look at two aspects of music - the average versus peak power, and the energy distribution of typical music material.

It is commonly accepted (and quite valid) that music has a peak to average ratio of about 10-20dB. This means that if the signal is being amplified by a typical 100W amplifier, the amp's power rating limits the absolute maximum power to 100W (give or take a little). Since this is the peak, the average must be somewhat lower, and we will assume 10dB for the sake of convenience. Average power is therefore 10W or less at the onset of clipping.

This is not dynamic range per se, but it is most certainly a part of the overall dynamic range of the music signal. The term 'dynamic range' usually refers to the very quietest up to the very loudest passages in a given piece of music. In some cases, there is no variation whatsoever - it starts loud, is loud in the middle, and (just to be different) finishes ... loud. The peak to average level may also be compressed, but it is difficult to reduce it to less than 10dB without it becoming flat and lifeless. If done incorrectly, it can simply become a jumbled mess with no intelligibility whatsoever (and no, I'm not going to take this to its logical conclusion and denounce various styles that may be classified as music to only a select few)

Most speakers are rated for a continuous power and an instantaneous power - the voice coil and to a lesser degree the suspension can withstand short bursts at much higher powers without damage. This does ***not*** imply that such power will be reproduced cleanly, and it will almost certainly be with a large increase in distortion. The peak power rating defines the maximum *transient* power the loudspeaker can handle without suffering electrical or mechanical (stress induced) damage.

Nearly all tweeters are rated to 'system power', and this will usually be quoted relative to a specific crossover frequency. An hypothetical tweeter may be rated at 100W system power when crossed over at 3,000Hz. The power that it can withstand is not 100W! Not

at any frequency or for any duration.

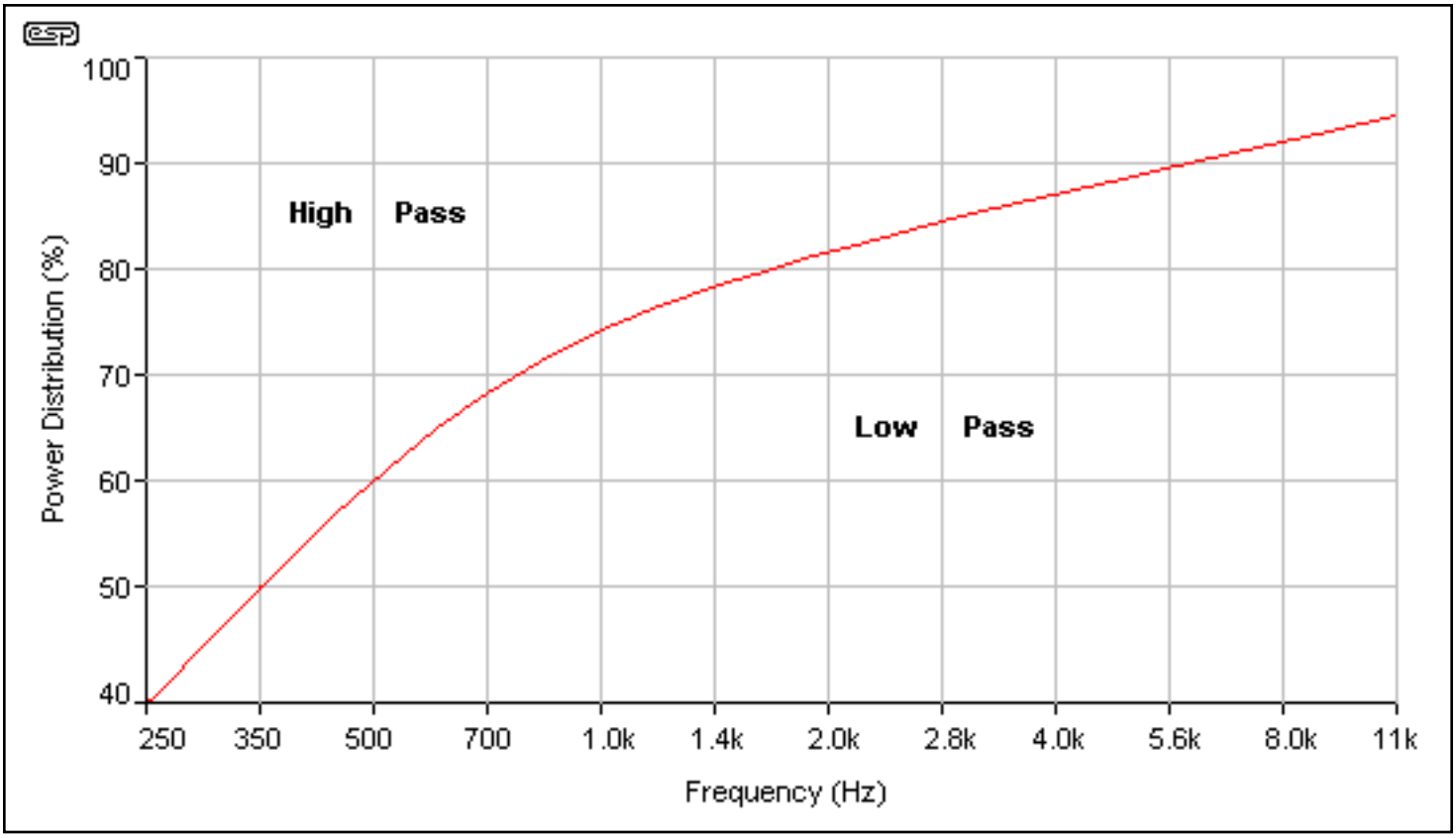


Figure 1 - Power Distribution Chart

The above power distribution table is approximate (as must be the case), and applies for 'typical' music - whatever that may be. If we look at the case for a crossover frequency of 3kHz, we can see that 85% of the power is in the low frequency spectrum, and only 15% in the high frequencies above 3kHz. It is not difficult to deduce from this that the peak power to the tweeter will be in the order of 15W at full power from the amplifier, with the average at about 1.5W

This is the way the system was designed to be used, and as long as the power amp does not clip, all is well (well, almost - read on).

Overdrive Conditions

When an amplifier is overdriven, the sound becomes distorted. This manifests itself in many ways, but the two we are interested in are the generation of harmonics, and the reduction of dynamic range - both the true dynamic range and the peak to average ratio. Let's assume that the amp is overdriven by a mere 3dB, so the average level is now 20W, and the peaks are clipping. With many systems (or listeners), this will be virtually inaudible. Careful listening will uncover the fact that there is distortion present, and there

is a definite reduction of intelligibility.

The speakers - both tweeters *and* woofers, are now being asked to absorb twice the power that would be normally obtainable, and the power is more constant - the signal is compressed by the power amp. Add to this the additional harmonics generated by the clipping waveform, and the tweeter may actually be getting up to 3 times the continuous power that was available before clipping. Peak power remains the same, since it is limited by the amplifier's power supply voltage.

Now, let's overdrive the amp by 10dB. The amp is delivering in excess of 100W, since it is reproducing square waves much of the time. The woofer will be subjected to perhaps a continuous 100W of power, and around 15W continuous will be available to the poor tweeter. Of this, probably less than 1% will be converted into sound (1% represents an efficiency of about 92dB/W/m). Ferrofluid helps, but virtually no hi-fi tweeter can withstand that sort of continuous power for any duration.

The tweeter was never designed for that! Just look at a 10W wirewound resistor for example. It is big and chunky, and made from a ceramic material that is designed to handle *a lot* of heat. Run one at 10W to find out just how much heat you will get. There is very little airflow around the tweeter voice coil, and the heat has nowhere to go. The result is that the voice coil will quickly overheat, and the adhesive that bonds the coil to its former, the former itself, and even the enamel insulation on the coil will be damaged. The result (naturally) is a dead tweeter.

As for the woofer - unless it is designed to take 100W or more continuous sinewave power, it will also overheat and eventually die. It takes a lot longer, because there is airflow around the voice coil, and the coil is bigger and has greater thermal inertia, but die it must if the abuse is maintained.

Example Power Waveforms

The following diagrams illustrate the above. Due to a complete inability on my part to create a power waveform that conformed to the examples I used before, the scale is somewhat different. In this case, the peak to average ratio is 30dB. Figure 2 shows the power over an arbitrary amount of time - it may be seconds or minutes - the result is much the same.

Note that this is not the signal waveform that you might see on an oscilloscope. This is a power vs. time waveform, and is not meant to represent anything specific - it is merely

an example.

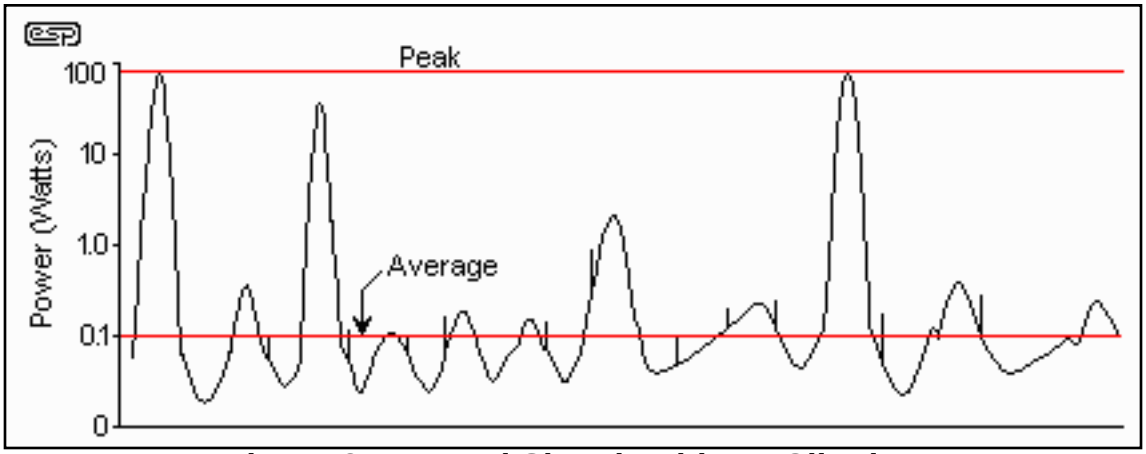


Figure 2 - Normal Signal, Without Clipping

This shows how the peak and average power are very different. Average power is 100mW, and the peak power is just under 100W, representing a 30dB ratio (approx). This can be quite typical of classical music, although there are obviously many other styles that can achieve similar ratios. It must be said that lovers of classical music are the least likely to clip their amplifiers except for short periods.

Any music with a smaller peak to average ratio simply makes the following examples happen sooner, with a smaller increase in overall volume control setting.

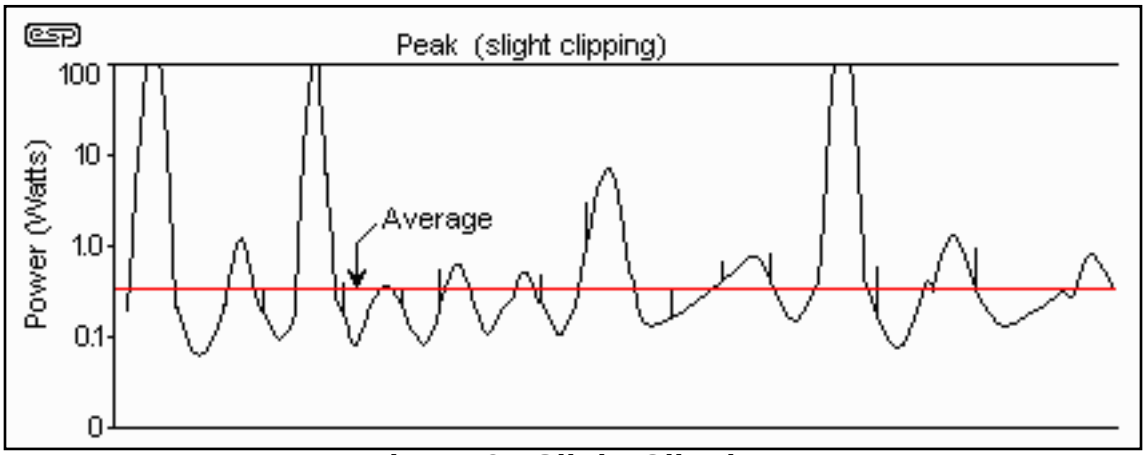


Figure 3 - Slight Clipping

Should the volume be advanced by 5dB, the peaks cannot exceed the power supply voltage, so they are clipped. This will be virtually inaudible with some music styles, but the average power is now about 300mW - an increase of almost 5dB over the previous case.

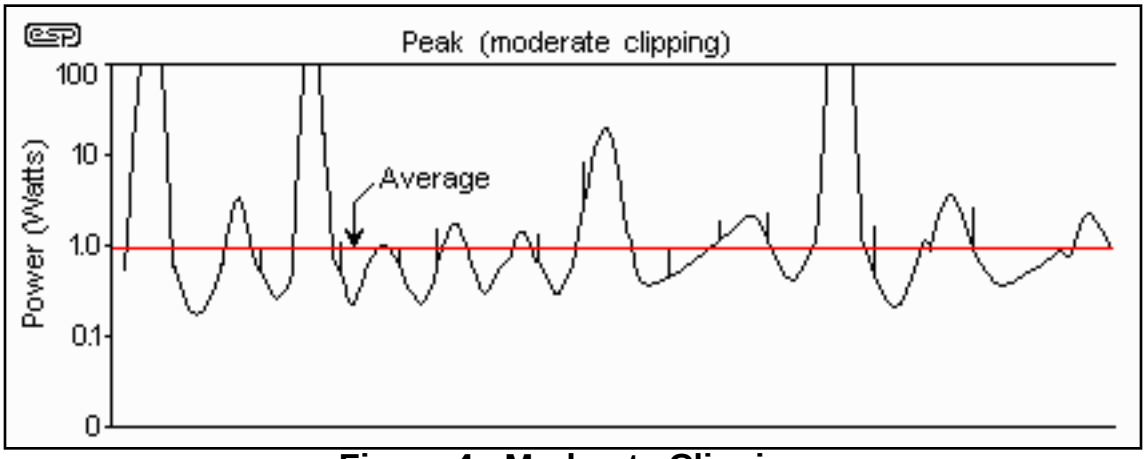


Figure 4 - Moderate Clipping

Should the volume be advanced by 10dB from the original, the peaks are clipped even more. This will be audible with most music, but is not at all uncommon. Parties are probably the worst offenders, although I have heard many high power (10kW or more) PA systems driven to at least this level of distortion. Average power is now 1 Watt, an increase of 10dB from the first example. The tweeter will probably be able to withstand this quite easily, but will no longer be linear (not that it matters any more, since there is so much amplifier distortion).

Bear in mind that most "ordinary" (non-audiophile) people will not be aware that there is anything wrong at this point! Liberal quantities of alcohol enriched fluids will ensure that ears remain unresponsive to the assault. Given the number of times I have heard about peoples' speakers "blowing up" during (or after - allegedly) a party, we can safely conclude that the requirement for more noise is far greater than any need for quality.

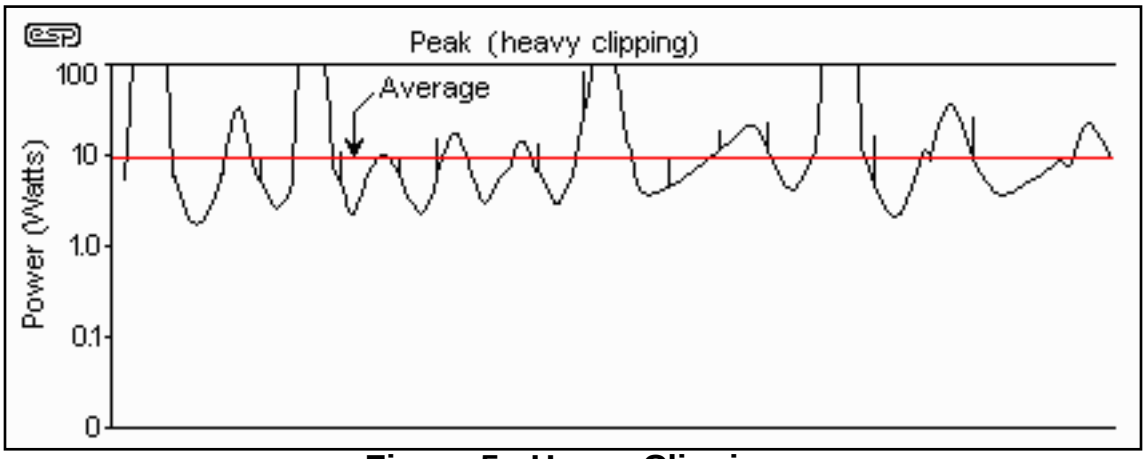


Figure 5 - Heavy Clipping

This is the bad guy! The average power is shown at 10W, but in reality is slightly higher because of the heavily clipped waveform. Few tweeters will last long with sustained power at this level - they are simply not designed for it. Midrange speakers and woofers will probably be pushed beyond their limits. Remember too, that not only is the "in band"

signal increased, but so is the out of band signal. The tweeter must now handle 10 times the normal power dissipation at all frequencies below the crossover frequency. This alone can cause damage, without the heating effects of such a high sustained power level.

Is the last example extreme? Not at all - I have actually understated the reality. When any system is driven so hard that the distortion is obvious to even untrained ears, it is reproducing close to square waves much of the time. A 100W amplifier driven to full power with a square wave will produce 200W, and when driven into heavy clipping with a music signal can easily achieve well over 100W average continuous power.

Conclusion ? ---- Exit tweeters (and/or midrange drivers and/or woofers).

Excessive Compression

All the above is valid when an amp clips, but what about speaker drivers that blow for no apparent reason, and with no clipping at all? Well, the people who create the final mix for CD (and those who actually transfer the final mastered tracks to CD format) have been using compressors for a long time - almost as long as music has been transferred to mass media in fact. What is different now, is that many of them seem to use their compressors with increasing vigour. The result is music that is compressed so heavily that the 10dB peak to average ratio we used in earlier examples no longer applies. Each individual track is compressed, the final mix is compressed, and it may be compressed yet again before it is transferred to CD. For reasons that I cannot fathom, it seems that every new CD that is released has to be 'hotter' (louder) than the last. Guess what people - there is a limit, *and we have reached it !*

Towards this end (and to make matters worse), you will find some CDs that are pre-clipped, with the peaks of the occasional loud signal neatly cut off at the top and bottom. With a peak to average ratio of perhaps 6dB in the worst cases, we need to examine the effects on the loudspeaker drivers.

Using the same 100W amp as above, it was already established that the average power would be around 10W at the onset of clipping. Although this is not a trivial amount of heat, most speakers will be able to dissipate it without too much difficulty. As the peak to average ratio is reduced by compression, the effect is exactly the same as clipping the amp, except we don't get the harsh distortion.

With a 6dB peak to average ratio, the poor loudspeaker now has to cope with an

average power of 50W as the peaks reach clipping level. Unless you are dealing with serious professional drivers that are designed for continuous high power, 50W becomes a lot of heat to dispose of. Should you be using a 200W amp, that increases to 100W - continuous! If you think 100W is trivial, try holding a 100W resistor that is dissipating its rated power. Actually, don't! You *will* burn yourself quite badly. Remember that over 99% of all power that goes into a domestic loudspeaker is dissipated as heat, with less than 1% being converted into sound.

As you can imagine, it is entirely possible to destroy drivers without ever letting the amp clip. Speaker systems rated for 200-500W almost always refer to peak power, and the average (or continuous) power rating is in reality as little as 10% of that claimed. This is perfectly alright for most people in a domestic environment, where very high SPL isn't needed. If you like to listen really loud, then you will almost certainly have problems with drivers failing. Perhaps the loudspeaker manufacturers haven't quite got the message that a great deal of music is compressed much harder than they anticipated.

The above applies to tweeters, midrange drivers and woofers equally - all can be damaged by excessive sustained power. Regardless of claims, there are actually very few loudspeakers that can withstand 100W of continuous average power for any length of time. To be able to do so requires large diameter voicecoils (preferably with aluminium formers to aid heat removal), vented pole-pieces, finned magnet assemblies and somewhere for the heat to go - preferably not just into the cabinet where it can't escape. This is not a trivial exercise, and professional drivers that can withstand that sort of abuse are very expensive indeed.

Finally, the CD (and other digital formats) allow the recording engineer/ producer to have as much bass as they want, plus some more for good measure. This can extend to very low frequencies - below audibility in fact. Tuned (vented) loudspeaker enclosures (usually) provide better bass extension than sealed types, but if there is any material recorded (and amplified) that is below the box tuning frequency, this will cause massive woofer cone excursion because the rear of the cone is unloaded below the vent tuning frequency. You won't hear the deep bass, but the cones will move very energetically indeed, and can easily reach physical excursion limits. Loudspeaker damage is guaranteed if this is done at high power levels, and can include buckled and/or detached voicecoils, broken lead-in wires and torn suspensions.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 28 Jul 2001

Elliott Sound Products **What *Is* Hi-Fi**

Copyright © 2001 - Val Pazin (for ESP)
Page Created 05 May 2001

Return to Articles Index

Back to Home Page

Contents

- Introduction
- About Hi-Fi
- Equalisation
- Hearing Things
- What They Really Need to Know

Introduction

Hi-Fi - Where did it originate?

Its origin comes from the basic need to communicate. The methods of communication have become less sophisticated (i.e. more blatant) in recent years, and this same claim has probably been made by every generation for hundreds of years. There are only a few of the more subtle methods left, and music is most assuredly one of them. Naturally, the better the reproduction, the greater the enjoyment, although it must be said that a degree of education is needed to separate the good, the bad, and the indifferent. The downright ugly will not be discussed here.

As we scurry along with our lives, along the way we become excited by elements that we have not paid much attention to, largely because we are distracted by the fact that one's wife is pregnant, the bank approved the huge mortgage, pressures of work, etc. As these pressures

ease, we become more interested in the subtle aspects of music reproduction, where before it may simply have been used as something to dance to, or as background noise.

(At this stage I'm hoping my spelling is above average as I'm polishing off a Scotch & coke, and yes it is *very* dangerous to drink coke on its own!).

> This article is a contribution from Val, and is subjected to my standard editing policy only - i.e. a small change here, some extra text there, and the occasional editorial comment, as shown here in reduced font and indented. The views held are those of the author, and do not necessarily represent or contradict the opinions of Rod Elliott (ESP).
>
> For overseas readers who may be unaccustomed to it, the (Ed.) means editor, just in case you were wondering. No? Never mind :-) (Ed.)

About Hi-Fi

Hi-Fi has evolved into a fine art as technology has improved the quality of every aspect of audio. As is usual there are the two steps forward one step back, which occurs in many areas of technology.

For example, when we first made the transition from vinyl records to CDs, there was something very wrong with the CD sound. This was partly due to the fact that test measurements were designed to pick flaws on the existing (vinyl) system. The flaws that were overlooked cause serious distortions, therefore reducing the confidence, enjoyment and interest of the general population.

Currently we are heading into a golden era for digital sound due to higher resolutions of the digital converters and more attention is being paid to the often neglected analogue stage. I have for quite some time been improving CD players for a moderate cost and achieving a good performance for price ratio. Some owners that I know have spent AU$1,200 on machines that originally cost AU$1,000 by importing clock upgrades and souped up analogue sections. A case of too much money and not enough brains, or a lot of over engineering, maybe both.

I have had better results using quality opamps and very low ESR (Equivalent Series Resistance) capacitors for between AU$180 and AU$320, including minor repairs 60% of the time. To simplify things, the analogue stage is trying to cope with digesting a 44.1kHz square wave residual as per the original specification, and over 340 khz with single bit converters. With SACD this will be over 2 megahertz. Small amounts of this type of signal destroy the performance of the opamps and as machines gradually deteriorate (sometimes as early as 18 months.) the supply lines become infected by odd order noise. Having experimented with over 50 types of opamps, I have developed a bias for a select few. (Not necessarily the fastest

ones as they tend to oscillate, but only until you attach test equipment or remove the lid :-)

I have just returned from installing a Yamaha AV receiver and a couple of Hifi VCRs at a friendly customer's place. He greatly appreciated my prompt attention to earn some money while providing him the ability to enjoy his DVD collection, refrain from exercising and to keep him company for a short while.

Having known Rudolph for some time, we have become good friends and have a trusting relationship. That is, I suggest items and he buys what I recommend.

We came upon this relationship due to the fact that I pointed out that he was responsible not in choosing his own equipment, but choosing someone with the confidence to do it for him.

He may easily have ended on the hifi (HELL) heap, as did many of my customers, except for the fact that I got him early. Many people in an approach to obtain a great deal, take on a huge burden on deciding what are the important criteria when purchasing equipment they know little about. At this stage I know that the three most important things in a stereo system are ...

1. the recorded material,
2. the speakers,
3. the room they are performing in.

It has taken me 12 years to realise this and helps me to find people who also know what they are about in audio. Hopefully it will only take me a few essays to explain this simplification.

People who do mixes for bands pick this up, as do most women but we poor audiophiles are side-tracked by brands such as Krell, Gryphon, and their gold plated friends (plated so as not to expose the crap underneath, perhaps). A woman won't sit for very long if the atmosphere is poor, nor will a person with little or no understanding of acoustics. If something is irritating them they just leave.

That's right folks, let your wife or girlfriend listen to her favourite music through your stereo and if she stays for longer than 15 minutes you're on a winner, you either got lucky or have an engineering degree in acoustics, electronics and have been fortunate to make quick decisions based on your hearing.

Equalisation

EQ - there you go, I have just sinned according to many a purist. I have fitted many graphic equalisers in homes, cars, and PA systems and every time, yes ***every time*** there is an improvement. This also means I know how to use the tool correctly.

I agree with my customers, silver cables do make a difference and so do pointy feet, Teflon, temperature, humidity and many other factors but the holy trinity (recorded material, speakers, room) still account for 90% of it. Sometimes more but never less!

Oh, by the way my audiophile customers who loathed the thought of EQ'ing their system now have them permanently connected, and occasionally use the bypass switch to hear what they had before (YUK). As was said on many an occasion; "Its amazing how well our brain compensates for these atrocious transistor radios".

I have great respect for transistor radios, as they are not pretentious. On the other hand, very expensive equipment such as the Wilson "doggies" safeguard themselves by instructing the owner to place the speakers and themselves as far from the room boundaries as possible. Why bother, just wear headphones if your environment has such an adverse effect on the speaker. But if the speaker has any flaws you're still stuffed!

We need to do reviews on people who know how to set up our systems properly, and not just review the equipment itself.

If you're not sure whether something sounds good or bad then don't buy it, no matter what the rest of your peer group may say.

Hearing (and Seeing) Things

Recently one of my clients said the too often quoted phrase " I'm hearing things I have never heard before!". Upon which one of my associates who was under the floorboards running cables to another room commented "It was all there before, but it's just like looking at an ugly woman ... the longer I look the more uncomfortable I feel". You may need an example of "discomfort" in this context ...

From time to time I drop into my local mega stores to buy the traditional wedding toasters and such. Upon poking around and looking at the racks of ill matched and poorly displayed equipment I am approached by some unsuspecting new sales person.

Salesman	Hi! May I help you ? (Yes, please call the manager, supervisor and/or owner)
Me	This rear projection TV is poorly adjusted, will you bring the remote for it please so I can see how well it can perform?
S	Yes, just one moment (off he goes and gets it) There you go.
Me	Now can you dim the lights in this section by 70% please?
S	I don't know, just hold on.

A few minutes or so later ...

S Sorry sir, the manager says that for insurance reasons the level of light may not be adjusted.

Me Can you demonstrate this set in an appropriate environment so the contrast and brightness can be reduced from maximum? (I sense his growing frustration).

S Hold on a moment!

As he goes off, I proceed to adjust all of the TVs, reducing from maximum setting all the ones that have the greatest margin of profit. The ones they don't push or don't have in stock have the contrast set very low and look washed out. Some time later ...

Manager Please ask for assistance if you require help as we pay to have them set up.

Then I proceeded to explain that I have noticed less customers in the TV section since they have been "set up" and that if all the sets had a good picture people would be less likely to walk off.

The manager pulled me up on my next visit and asked me why there were more people in the section since I had a play with the tellies. "There are two reasons" I said, "Firstly, people feel more comfortable because they don't suffer fatigue from the poorly adjusted sets, and secondly, I adjusted the boominess out of a badly installed subwoofer located nearby."

What a way to get some free blank videotapes. I am saddened by the lack of basic knowledge of how to set up audio & video systems. Even the so-called specialist dealers (especially when they get in a bind) call on subcontractors to help them out.

> The reader may have noticed that the nearby subwoofer and the TV settings are seemingly unrelated - but we use all of our senses all of the time, and will equate bad sound with a bad picture (although the reverse is not always true). Tests have been done with an identical film clip, but different soundtrack quality. The viewers were asked to rate the picture quality *only*, and to disregard the audio. Overwhelmingly, the test group said that the clip with the highest sound quality had the best picture - but the *picture never changed!* I would cite the reference for this, if only I could recall where I saw it. (Ed.)

What They Really Need to Know

No sales person will find out anything useful unless s/he asks the right questions. This is regrettably rare.

The first question should be - *What is your room like, and how long will you live there?*

If you intend to stay in the one place, floor standers may be in order, if you move around small satellites & a subwoofer will be more flexible for placement reasons. The furnishings and the structure of the room determine the tonal character of the speaker.

What type of music do you listen to, and do you actually sit and face the speakers?

If you face the speakers while listening you may be able to determine a sense of width and depth between performers and instruments). This is called the sound stage.

"Imaging" is how well defined the performers and instruments are within the sound stage. Without going into great detail get a copy of "The ultimate demonstration disc" Chesky Records UD95. You will hear things you didn't understand before.

Do you have loud parties?

If ever you have to turn the knob past the 12 o'clock position on an amplifier STOP! Get another amp with more power and maybe more efficient and powerful speakers. (Except if you are using Rod Elliott's Better Volume Pot (Project 01). For this, 3 o'clock is its limit).

> There are some possible reasons that the above may not apply. If the source is low output, or the material was recorded at low level, it is sometimes necessary to increase the volume further than indicated. If you hear sounds of distress from the speakers, *then* you have a real problem with either the amp, speakers or both. Sometimes, if the sound is too loud, you may not even hear the distress. Place your fingers in your ears to reduce the level - you will be surprised at the result. Clipping distortion from 20kW PA systems is immediately audible when you reduce the level to something your poor ears can actually cope with (Ed.)

And lastly - *What is the budget?*

This is where the juggling begins!

If these questions are not asked you will be pointed in the "we have just the one for you" direction.

"This one is on special, a bargain" (we have just done a cash deal with the supplier who may shortly have a fire sale).

Contact The Author, Val Pazin

Return to Articles Index

Back to Home Page

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Val Pazin and Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Val Pazin) and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allow that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Val Pazin and Rod Elliott.

Page created and copyright (c) 05 May 2001

The Audio Pages

Elliott Sound Products **What *is* Hi-Fi (Part II)**

Copyright © 2004 - Rod Elliott (ESP)
Page Created 12 Jan 2004
Published 27 March 2004

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Accuracy
  - 2.1 When, And Only When You Need It
  - 2.2 Linearity Vs. Accuracy
  - 2.3 Distortion
- 3.0 Conclusion

1.0 Introduction

In Part 1 of this topic some of the general principles were covered. It explained that many of the concepts applied to 'High-End' systems are at odds with one's listening pleasure, and included some reasons for this.

We tend to think that a hi-fi system should be accurate, but what exactly *is* accuracy, how do we assess it, and what happens if the system is not 'accurate' for one reason or another?

The purpose of this article is to attempt to cover the reasons that accuracy is not always

desirable or achievable, and to allow the reader to make an informed judgement of a system and its abilities. This is a somewhat daunting task, and one that may raise the ire of many, since it appears that I am suggesting that an accurate system is not useful.

It's not that an accurate system is not desirable - it is desirable in the extreme, IMHO. The problem is not so much with the system we listen to, but the source material itself. The room also plays a very important part in the equation, as the average listening room has a great many anomalies at varying frequency ranges, and can make an excellent system sound very ordinary indeed if it is unsuited to the purpose of listening to music.

I do not intend to cover room treatment in any detail - there are many articles already that do just that, and it is a non-trivial exercise at best. As an example though, I am sure we have all seen photographs on the Net and in magazines of superlative looking systems in a stark room, having no carpet and a nice tiled floor, minimal furnishings, and lots and lots of glass windows. While this may well suit the purposes of the photographer, a system in such a room will almost always be a complete sonic disaster.

Likewise, having heard a piece of music on the radio (for example), you go and buy the CD, only to get it home and discover that it is over equalised, compressed to within an inch of its life, and has excessive and boomy bass. I have several such recordings, and my system is *too* accurate, revealing every flaw with punishing accuracy. Under these conditions, an accurate system is not ideal - in fact it is a disaster! Music you might otherwise be able to enjoy is ruined, and a typical "high-end" audio system is completely inappropriate for a great many recordings. For example, it should not be necessary to switch off the subwoofer to play some CDs.

2.0 Accuracy

Accuracy in a general context could mean that the system will play the source material with no changes to the frequency response, harmonic structure, relative phase of the signals, or anything else. While this may seem to be the ideal situation, it is never achieved in a practical application. Non-linear distortion will be added, and this changes the harmonic structure of the signal. Very small amounts of non-linear distortion are added by most amplifiers and preamps, much greater amounts by all speakers. While the levels will rarely be objectionable with well designed equipment, that the signal undergoes some distortion is an absolute - there is no such thing as a zero distortion amplifier or transducer.

Certainly, the distortion levels may be able to be reduced to below the noise floor of the

listening room, but it is still there - whether you hear it or not is the only thing that is important. The concept of 'masking' - where one signal is deemed inaudible because of a higher level signal nearby does not always work as well as many designers may hope. A perfect example is MP3, where all 'masked' signals are removed as part of the compression algorithm, and 'near CD quality' MP3 recordings lack any sense of imaging or 'space' because we use the 'inaudible' signals as a cue for image placement. Since they have been removed, we lose that ability.

Frequency response will always be modified to some extent. Again, whether these changes are audible is the only question that is important - you (as the listener) may actually *want* to modify the frequency response, either to compensate for recording quality, room effects or even hearing loss. Should you suffer from a loss of high frequencies, then the ability to compensate (at least to some degree) will enhance your enjoyment, although it is probable that others will find the sound overly "bright".

Phase shifts occur in all aspects of electronics to some degree, and are particularly easily measured with crossover networks. Fortunately, our sensitivity to static phase shifts is rather low, and we can tolerate huge phase shifts with no noticeable change in sound quality. Relative phase is another matter, and any piece of electronics or transducer array that changes the phase of signals dynamically (almost always deliberately) will cause very audible effects. Rotating speakers (e.g. electronic organ Leslie cabinets), or phase shifting networks (e.g. phaser pedals for guitar effects) have a profound effect, but these are rarely found in hi-fi systems. Loudspeaker drivers inadvertently wired out of phase cause static cancellations of some frequencies, and this is almost always audible.

Phase shift is one of the least understood of the distortion mechanisms, but it is important to understand that the phase of any signal is changed radically by distance, even if it started out as perfect. Sound travels at 345m/s in dry air at sea level at around 22°C, but this changes with altitude, temperature and humidity. This works out to be .354mm/us, or 35.4mm/100us. Given that a typical listening distance may be 2 metres, that represents a time delay of 5650us, or 5.65ms. The chart below shows the relative phase shift of a perfectly phase coherent signal, at distances of 1 and 2 metres.

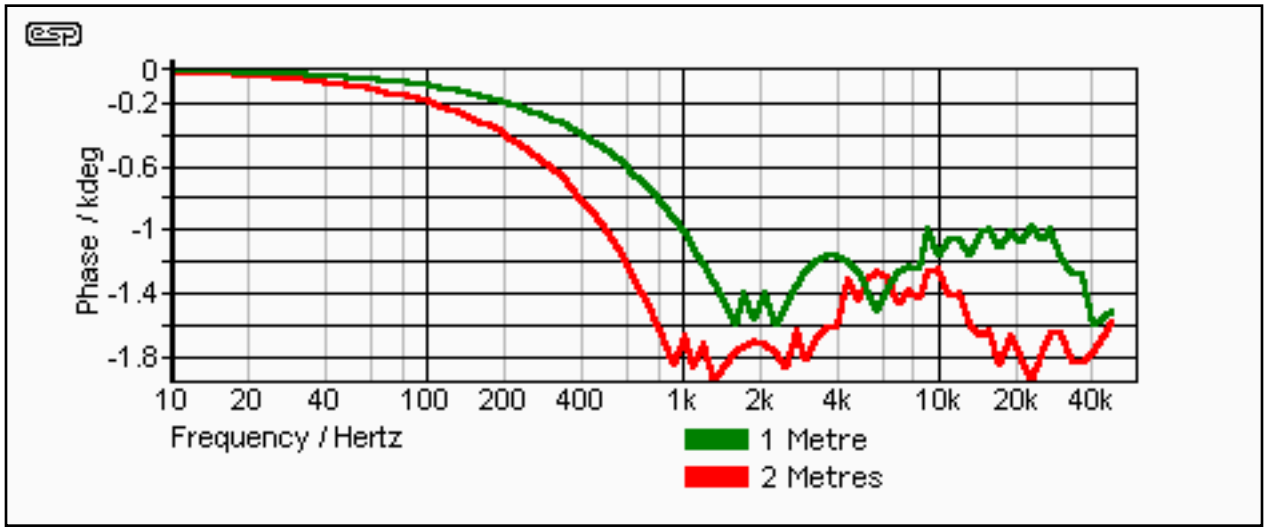


Figure 1 - Phase Shift at 1 & 2 Metres Listening Distance

At a maximum of 2,000°, this is hardly 'phase coherent' (note that the chart is in k degrees), but it happens with every speaker (including the human kind) in every room, under all circumstances. This is not distortion, but simple reality, to which our hearing is fully accustomed. The phase shift is the result of the time delay suffered by the signal as it travels through the air, and is completely normal. Hearing is not affected by this (radical) change for the simple reason that it occurs all the time, with every sound source. It also renders the claims that some speakers can reproduce a perfect squarewave rather suspect, as shown below ...

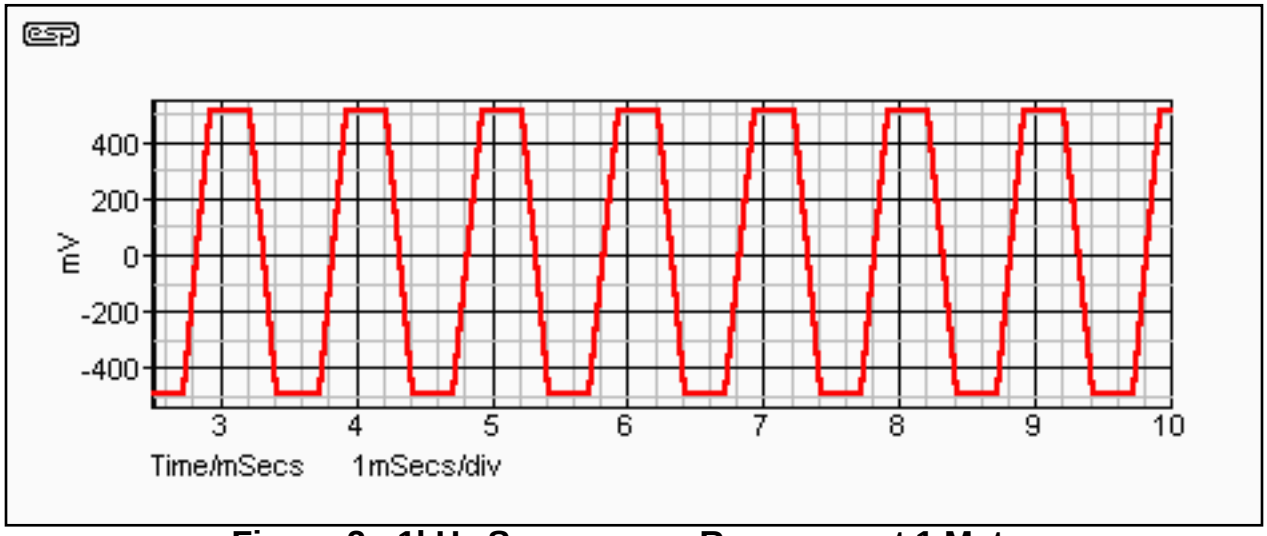


Figure 2 - 1kHz Squarewave Response at 1 Metre

Not what I would call a perfect squarewave, and naturally the shape varies with distance and frequency. The room will also have a profound effect on the waveform, and from tests I have done, I know that perfectly ordinary (non time-aligned) loudspeakers can also reproduce a squarewave - provided one is sufficiently patient to find the 'correct' microphone placement. As a matter of interest, it is almost impossible for *any* loudspeaker to introduce a phase shift of this magnitude, regardless of crossover

topology or baffle layout. As I have pointed out on several occasions, phase shift in and of itself is usually considered to be inaudible, but the effects of combining signal sources with different phase shift characteristics can be very audible indeed.

The greatest distortion is one of time itself - the re-creation of a piece of music without the musicians being present is obviously a "distortion". 100 years ago, if you wanted to listen to music, then the musicians would be right there with you, and there was no distortion - you heard exactly what you heard at your seating position. Others would hear things differently, because they would hear the same music from a different location, thus changing the sound. It is no longer necessary to have the musicians present, but for the purposes of any discussion, this distortion is usually ignored. It shouldn't be, because this is the one thing that people strive for - to make their system sound like the original performance. The fact is that this is impossible, because the original performance was recorded under conditions that cannot be duplicated in a home system, in a normal room, and with all the changes that have been made during the recording process itself. Indeed, even when present at a concert, you will hear the sound differently from different locations - which is right, which is wrong? Obviously, no position where all instruments are heard clearly is wrong, but they are different regardless.

2.1 When, And Only When You Need It

If your or my system is accurate yet cannot reproduce the music in a manner that is enjoyable, what is the point? One could scour the record shops in a quest for CDs that are well recorded and properly mixed (i.e. without excessive compression or use of effects), but if the CDs you can find have music that you don't particularly like, again, what is the point?

If you happen to enjoy "pop", contemporary or alternative music, then you absolutely need some method for correcting the often appalling mix quality, since a great many of such recordings are deliberately mixed to sound "good" on exceptionally average systems (such as factory installed car sound and department store "hi-fi"). If you like classical or string quartets, choral or organ music, you are not necessarily safe there either, since it is very common for "post processing" to be performed by the mastering engineer.

The fact of the matter is that the *only* way you will hear exactly what was intended by the recording engineers is to hear it in the studio where the recording was mixed, using the same speakers at the same level - that this is impractical in the extreme should be fairly obvious. No system and home environment will match the acoustics of the studio or the

monitor speakers used. Even if you got lucky with a "fairly good" match, it would only be applicable for a single studio at the same level that was used for mixing.

Equalisation is considered a rude word by most in the high-end (or audiophile) community, yet an equaliser, or even tone controls, can rescue an otherwise unlistenable recording. Exactly when and why tone controls got such a bad name is unclear, but my next preamp will most certainly use them. It is probable that they will be used only on a very few CDs, but at least my investment in the music is retained if I can actually listen to it.

What is really at issue here is that "accuracy" has become the catch-cry of some of the high end of the audio fraternity, and while it is absolutely essential that a system is capable of a flat response and full dynamics (accuracy), this is a very open-ended situation because the quality of recordings is so variable. It would be nice if orchestral recordings (for example) were labelled "No EQ" or "No Post-processing" so that we would know that the response had not been tampered with beforehand, but this is not the case (at least for any I have seen).

There are Audiophile recordings available, but if you can't find any with the artists of your choice or the material of your desire, then they are of no use to you. In some cases, while the recording quality may be excellent, the musicianship may be left wanting. This is a no-win situation, and will remain so until tone controls take their rightful place in systems once again.

It goes without saying that tone controls are not the real answer (even though I just seem to have said they were), but if they can be used to rescue a recording (or several), then they are worth inclusion. Ultimately, hi-fi is about music and enjoyment, and if someone *really* enjoys listening to everything with the bass turned way up, then the facility should be available. Musical enjoyment is a very personal thing, and for a great many people the audiophile world has shut them out because such systems do not have the features they desire.

No-one has the right to impose their standards on everyone else, yet by insisting that systems have a flat frequency response that cannot be adjusted (not even balance controls are used in many systems) you are allowing recording engineers to impose their preferred sound onto you - if you don't like it, there is nothing you can do about it, apart from return the CD and demand your money back (not a bad thing to do, by the way, but you may deprive yourself of a lot of music you'd otherwise enjoy).

I have commented elsewhere on these pages that RIAA phono equalisation does not

have to be super accurate, since no-one has any way of knowing what additional EQ was applied to the cutting lathe at the time of mastering. That the playback EQ should be accurate to within a dB or so is a sensible approach, and deficiencies can easily be made good by the use of tone controls - again assuming that they are available.

What of systems where the user can adjust each frequency band in a multi-way active system? By the normal high-end standards, we should be appalled by such an idea, but it makes sense. Given that music is such a personal thing, why should anyone have to listen to the system as it was designed and built?

There is no reason that people should not be able to tailor the sound to their preference if it increases their enjoyment of the music of their choice. It will rarely be accurate, but what difference should that make? If you don't like it the way the user has set everything up, that is your problem, not theirs. As long as the listeners (owners) are happy with what they can achieve, to get the sound they like, then that is all that matters. You can go home and listen to your system set up the way *you* like - naturally, I consider the ability to get a flat frequency response (as well as good transient response, low distortion, etc.) as essential in all of this, but whether it will be used like that is another matter altogether.

2.2 Linearity Vs. Accuracy

Accuracy in the sense that I use it here is primarily about frequency response, with phase and transient response following closely behind. Linearity is usually taken to mean freedom from non-linear effects, and specifically harmonic and intermodulation distortions. As noted above, these are always present to some degree, and usually vary with amplitude (loudness) and frequency. Most at risk are extremely low and high frequencies - low frequencies because they demand large amounts of movement in the transducer, and high frequencies because of limitations within many tweeters.

Quoting THD (Total Harmonic Distortion + noise) is a pointless exercise if the waveform of the distortion component is not disclosed (which it generally is not). Two amplifiers having *apparently* identical amounts of distortion can sound very different indeed, and this has led to THD measurements being declared 'useless'. Provided you know what kind of distortion is present, it is generally easy to make a prediction as to whether it will sound "good" or "bad". Solid state amplifiers with 1% crossover (or notch) distortion sound unquestionably bad, while a valve amp with 1% 2nd harmonic distortion will sound good (especially by comparison).

Linearity is important, since even relatively small amounts of THD can create much larger amounts of IMD (Intermodulation Distortion), where each frequency being reproduced beats or mixes with other frequencies to create new frequencies that were not present in the recording. This happens with electronics, but is usually much more pronounced with transducers, and for this reason (amongst others), it is common to use separate transducers for different bands of frequencies. The simplest is a woofer and a tweeter, with a crossover network that divides the signal and sends the low frequencies to the woofer, and high frequencies to the tweeter.

Transducer non-linearity usually exceeds that of electronics by a great deal, but tends to be 'softer' and less audible. Not so however when the linear excursion limits of a speaker are exceeded - the distortion produced can then be very audible and highly objectionable. For systems that are used at low levels this is not so much of a problem, but to obtain good linearity at even reasonably high levels requires that drivers be limited to a maximum of about 1 decade (a little over 3 octaves). To manage the whole audible frequency range, the signal may be split as follows ...

Low Bass	Mid Bass	Upper Mid	Treble
20 - 160Hz	160-1280Hz	1280-10kHz	10kHz-20kHz

Unfortunately, for a number of reasons, this is not a practical separation of the signal. Because of the difficulties with bass (in particular) a more common split may be as follows ...

Low Bass	Mid Bass	Upper Mid	Treble
20 - 80Hz	80-500Hz	500-3kHz	3kHz-20kHz

This too has problems - there is a crossover frequency at 500Hz that may cause problems for the natural reproduction of voices (a very critical requirement for any system). This is not insurmountable though, and the second variation is fairly common (with some minor changes depending on philosophy, driver specifics, etc.).

2.3 Distortion

Distortion actually covers every aspect of the system. Technically, if the volume is changed, you introduce distortion (amplitude distortion), since you are not listening at the same level that was used when the music was recorded or mastered. In turn, this

changes the response of your hearing, with the bass and treble seeming to be less pronounced as volume (SPL) is reduced (and vice versa). The change in amplitude is not considered as distortion, but it should be, since the relative loudness of the high and low frequencies is changed relative to the original.

The term "distortion" is usually taken to mean non-linear (harmonic or intermodulation) distortion, but the true meaning is *any* change that makes the reproduction different from the original. Since THD and IMD have already been covered, I shall only look at other changes to the system that change the way the sound is heard, versus the way it was recorded.

From this perspective, it is quite obvious that all hi-fi systems introduce distortion (of the original performance). The ability to manipulate tone controls, graphic or parametric equalisers, or crossover network frequencies and/or levels may be used to *reduce* the distortion introduced by the simple act of changing the listening level.

Years ago, it was common for systems to have a "loudness" control, often incorporated with the volume pot. That these almost never worked correctly is a given, since no attempt was ever made to provide any method of calibrating the "effect" to the SPL in the listening room. Needless to say loudness controls went away, but with them, in most high-end systems, so too did tone controls and even the balance control. While the loss of the loudness switch went almost un-noticed, the other losses should be missed, since they are useful.

It is far more convenient to be able to change a balance control or tone controls, than to have to move furniture about or listen at a specific volume all the time. Consequently, it could be said that most systems are "distorted" most of the time, since what you hear is not the same as what was recorded.

3.0 Conclusion

The dream of a system that is 'dead flat' (in frequency response) is not necessarily a good idea, since it is likely that it will be far less accurate than expected for a great many recordings. While technical excellence is always a good thing, it is unwise to place that above all else, and even less wise to assume that ever greater levels of excellence will lead to ever greater enjoyment.

What of the average person (non-audiophile) who wants a really good system? There is precious little available catering to that market at present, and the choices are limited to

the department store offerings (often with speaker boxes that look like something from outer space - and sound like it too!), or they can go and argue with the sales-thing at a high-end audio outlet, who will try to convince them that they don't want (or need) tone controls, balance controls, X-Bass or any of that "rubbish". Well, guess what? They *do* want them - all of them (well, most anyway). Will they use them with a good system? Probably not - initially they will all be used (usually radically), but after a while they will adjust and learn to listen to all of the music, and not just the bass. If it turns out that they do listen with the bass right up, then again, that is their right - it is their system, after all.

The purpose of a sound system in your home is for one thing only - to listen to music (and movie soundtracks, of course). Whether your system is optimal from a purist's perspective is immaterial. If *you* enjoy the sound, then you have a system that is perfect from your perspective. If others enjoy it too, then so much the better. The idea is for you to put together a system for yourself and (perhaps) your family - not other people. Should others in your household prefer a different sound, then there are two choices ... either you have two or more systems for different sounds, or use a system that can be tailored to suit each listener, or modified at will to suit classical music one session, rock the next, or tweaked further to suit the DVD currently playing.

The bottom line is that any sound system's ultimate purpose is for the enjoyment of music - not to follow the latest trend, or as a test vehicle for the latest magic stone, newest speaker leads or pure silver interconnects, but to enjoy music. Period.

Once you are in the position of listening for flaws in the system, you are no longer listening to the music - certainly you *hear* the sounds, but if you are so interested in the reproduction you can no longer concentrate on the music itself. You may use test instruments to analyse the response, phase alignment, spectral decay and distortion of your system, but if, after all of that, it still sounds good - who cares?

What of SET (Single Ended Triode) amplifiers and other systems that introduce often quite high levels of distortion? If this is what the listener likes to hear, then that is his prerogative, but it means that *all* music will be subjected to the same distortion. This may sound very nice with some recordings but revolting with others - again, limiting one's choice rather drastically. However, if this is what the owner wants, then fine. Such owners should resist the temptation to insist that others follow their path to 'nirvana' - most people don't want and don't need distortion.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 12 Jan 2004

Elliott Sound Products **Amplifier Basics - How Amps Work**

Copyright (c) 1999 - Rod Elliott (ESP)
Page Last Updated 06 Apr 2005

Articles Index

Main Index

Introduction

The term 'amplifier' has become generic, and is often thought (at least by some) to mean a power amplifier for driving loudspeakers. This is not the case (well, it *is*, but it is not the *only* case), and this article will attempt to explain some of the basics of amplification - what it means and how it is achieved. This article is not intended for the designer (although designers are more than welcome to read it if they wish), and is not meant to cover all possibilities. It is a primer, and gives fairly basic explanations (although some will no doubt dispute this) of each of the major points.

I will explain the basic amplifying elements, namely valves (vacuum tubes), bipolar transistors and FETs, all of which work towards the same end, but do it differently. This article is based on the principles of audio amplification - radio frequency (RF) amplifiers are designed differently because of the special requirements when working with high frequencies.

Not to be left out, the opamp is also featured, because although it is not a single "component" in the strict sense, it is now accepted as a building block in its own right.

This article is not intended for the complete novice (although they, too, are more than welcome), but for the intermediate electronics or audio enthusiast, who will have the most to gain from the explanations given.

Contents

- Basic Terminology
  - Impedance
  - Units
- Amplification Basics
  - Input Impedance
  - Output Impedance
  - Feedback
  - Signal Inversion
  - Design Phase
- Types Of Amplifier Devices
- Common Limiting Ratings
- Essential Electronics Formulae

- **Part 1 - Valves (Vacuum Tubes)**
  - Valve Characteristics
  - Valve Current Amplifier
  - Valve Power Amplifiers
  - Summary

- **Part 2 - Bipolar Transistors**
  - Transistor Characteristics
  - Transistor Current Amplifier
  - Transistor Power Amplifiers
  - Summary

- **Part 3 - Field Effect Transistors and MOSFETs**
  - FET Characteristics
    - Junction FETs
    - MOSFETs
  - FET Current Amplifier
  - FET Power Amplifiers
  - Summary
    - Junction FETs
    - MOSFETs

- **Part 4 - Operational Amplifiers (Opamps)**
  - Power Opamps
- **Part 5 - Some Basic Linear Circuit Building Blocks**
  - Current Sources and Sinks
  - Current Mirror
  - Long Tailed Pair
  - Grounded Grid (Gate or Base) Circuits
  - Cascode
- **Part 6 - Conclusions**
- References
- Copyright & Update Info

Basic Terminology

Before we continue, I must explain some of the terms that are used. Without knowledge of these, you will be unable to follow the discussion that follows.

Electrical Units

Name	Measurement of	A.K.A.	Symbol
Volt	electrical "pressure"	voltage	V
Ampere	the flow of electrons	current	A, I
Watt	power		W, P
Ohm	resistance to current flow		Ω,R
Ohm	impedance, reactance		Ω,X
Farad	capacitance		F, C
Henry	inductance		H, L

Hertz	frequency		Hz

Note: A.K.A. means "Also Known As". Although the Greek letter omega is the symbol for Ohms, I shall use the word Ohm or the letter "R" to denote Ohms. Any resistance of greater than 1,000 Ohms will be shown as (for example) 1k5, meaning 1,500 Ohms, or 1M for 1,000,000 Ohms. The second symbol shown in the table is that commonly used in a formula.

When it comes to Volts and Amperes (Amps), we have alternating current and direct current (AC and DC respectively). The power from a wall outlet is AC, as is the output from a CD or tape machine. The mains from the wall outlet is at a high voltage and is capable of high current, and is used to power the amplifying circuits. The signal from your audio source is at a low voltage and can supply only a small current, and must be amplified so that it can drive a loudspeaker.

Impedance

A derived unit of resistance, capacitance and inductance in combination is called impedance, although it is not a requirement that all three be included. Impedance is also measured in Ohms, but is a very complex figure, and often fails completely to give you any useful information. The impedance of a speaker is a case in point. Although the brochure may state that a speaker has an impedance of 8 Ohms, in reality it will vary depending on frequency, the type of enclosure, and even nearby walls or furnishings.

Units

In all areas of electronics, there are smaller and larger amounts of many things that would be very inconvenient to have to write in full. For example, a capacitor might have a value of 0.000001F or a resistor a value of 150,000 Ohms. Because of this, there are conventional units that are applied to make our lives easier (well, once we are used to using them, anyway). It is much easier to say 1uF or 150k (the same as above, but using standard units). These units are described below.

Conventional Metric Units

Symbol	Name	Multiplication
p	pico	1×10^{-12}
n	nano	1×10^{-9}

μ	micro	1×10^{-6}
m	milli	1×10^{-3}
k	kilo	1×10^{3}
M	Mega	1×10^{6}
G	Giga	1×10^{9}
T	Tera	1×10^{12}

Although commonly written as the letter "u", the symbol for micro is actually the Greek letter mu as shown. Because this would require the continuing use of a small graphic, I shall use the letter "u".

In audio, Giga and Tera are not commonly found (not at all so far - except for specifying the input impedance of some opamps!). There are also others (such as femto - 1×10^{-15}) that are extremely rare and were not included. Of the standard electrical units, only the Farad is so large that the defacto standard is the microFarad (uF). Most of the others are reasonably sensible in their basic form.

Amplification Basics

The term "amplify" basically means to make stronger. The strength of a signal (in terms of voltage) is referred to as amplitude, but there is no equivalent for current (curritude?, nah, sounds silly). This in itself is confusing, because although "amplitude" refers to voltage, it contains the word "amp", as in ampere. Maybe we should introduce "voltitude" - No? Just live with it.

To understand how any amplifier works, you need to understand the two major types of amplification, and a third "derived" type:

- Voltage Amplifier - an amp that boosts the voltage of an input signal
- Current Amplifier - an amp that boosts the current of a signal
- Power Amplifier - the combination of the above two amplifiers

In the case of a voltage amplifier, a small input voltage will be increased, so that for example a 10mV (0.01V) input signal might be amplified so that the output is 1 Volt. This

represents a "gain" of 100 - the output voltage is 100 times as great as the input voltage. This is called the voltage gain of the amplifier.

In the case of a current amplifier, an input current of 10mA (0.01A) might be amplified to give an output of 1A. Again, this is a gain of 100, and is the current gain of the amplifier.

If we now combine the two amplifiers, then calculate the input power and the output power, we will measure the power gain:

$P = V \times I$ (where I = current, note that the symbol changes in a formula)

The input and output power can now be calculated:

$P_{in} = 0.01 \times 0.01$ (0.01V and 0.01A, or 10mV and 10mA)

$P_{in} = 100uW$

$P_{out} = 1 \times 1$ (1V and 1A)

$P_{out} = 1W$

The power gain is therefore 10,000, which is the voltage gain multiplied by the current gain. Somewhat surprisingly perhaps, we are not interested in power gain with audio amplifiers. There are good reasons for this, as shall be explained in the remainder of this page. Having said this, in reality all amplifiers are power amplifiers, since a voltage cannot exist without power unless the impedance is infinite. This is never achieved, so some power is always present, but it is convenient to classify amplifiers as above, and no harm is done by the small error of terminology.

Input Impedance

Amplifiers will be quoted as having a specific input impedance. This only tells us the sort of load it will place on preceding equipment, such as a preamplifier. It is neither practical nor useful to match the impedance of a preamp to a power amp, or a power amp to a speaker. This will be discussed in more detail later in this article.

The load is that resistance or impedance placed on the output of an amplifier. In the case of a power amplifier, the load is most commonly a loudspeaker. Any load will require that the source (the preceding amplifier) is capable of providing it with sufficient voltage and current to be able to perform its task. In the case of a speaker, the power

amplifier must be capable of providing a voltage and current sufficient to cause the speaker cone(s) to move. This movement is converted to sound by the speaker.

Even though an amplifier might be able to make the voltage great enough to drive a speaker cone, it will be unable to do so if it cannot provide enough current. This has nothing to do with its output impedance. An amplifier can have a very low output impedance, but only be capable of a small current (an operational amplifier, or opamp is a case in point). This is a very important point, and needs to be fully understood before you will be able to fully appreciate the complexity of the amplification process.

Output Impedance

The output impedance of an amplifier is a measure of the impedance or resistance "looking" back into the amplifier. It has nothing to do with the actual loading that may be placed at the output.

For example, an amplifier has an output impedance of 10 Ohms. This is verified by placing a load of 10 Ohms across the output, and the voltage can be seen to decrease by 1/2. However, unless this amplifier is capable of substantial output current, we might have to make this measurement at a very low output voltage indeed, or the amplifier will be unable to drive the load.

Another amplifier might have an output impedance of 100 Ohms, but be capable of driving 10A into the load. Impedance and current are completely separate, and must not be seen to be in any way equivalent. Both of these possibilities will be demonstrated later in this article.

Feedback

Feedback is a term that creates more and bloodier battles between audio enthusiasts than almost any other. Without it, we would not have the levels of performance we enjoy today, and many amplifier types would be unlistenable without it.

Feedback in its broadest sense means that a certain amount of the output signal is "fed back" into the input. An amplifier - or an element of an amplifying device - is presented with the input signal, and compares it to a "small scale replica" of the output signal. If there is any difference, the amp corrects this, and ideally ensures that the output is an exact replica of the input, but with a greater amplitude. Feedback may be as a voltage or current, and has a similar effect in either case.

In many designs, one part of the complete amplifier circuit (usually the input stage) acts as an error amplifier, and supplies exactly the right amount of signal to the rest of the

amp to ensure that there is no difference between the input and output signals, other than amplitude. This is (of course) an ideal state, and is never achieved in practice. There will always be some difference, however slight.

Signal Inversion
When used as voltage amplifiers, all the standard active devices invert the signal. This means that if a positive signal goes in, it emerges as a larger - but now negative - signal. This does not actually matter for the most part, but it is convenient (and conventional) to try to make amplifiers non-inverting. To achieve this, two stages must be used (or a transformer) to make the phase of the amplified signal the same as the input signal.

The exact mechanism as to how and why this happens will be explained as we go along.

Design Phase
The design phase of an amplifier is not remarkably different, regardless of the type of components used in the design itself. There is a sequence that will be used most of the time to finalise the design, and this will (or should) follow a pattern.

- **Power Output vs. Impedance**
  The power output is determined by the load impedance and the available voltage and current of the amplifier. An amplifier that is capable of a maximum of 2A output current will be unable to provide more, just because you want it to. Such an amp will be limited to 32W peak (or 16W RMS) into 8 ohms, regardless of the supply voltage. Likewise, and amp with a supply voltage of +/-16V will be unable to provide more than 16W RMS into 8 ohms, regardless of the available current. Having more current available will allow the amp to provide (for example) 32W into 4 ohms (4A peak current) or 64W into 2 ohms (8A peak current), but will give no more power into 8 ohms than the supply voltage will allow.

- **Driver Current**
  Especially in the case of bipolar transistors, the driver stage must be able to supply enough current to the output transistors - with MOSFETs, the driver must be able to charge and discharge the gate-source capacitance quickly enough to allow you to get the needed power at the highest frequencies of interest. With valves, the driver needs to be able to supply enough current to supply the bias resistors only, since the valve grid draws little or no current (except for the special case of Class-AB2).

  For the sake of simplicity, if bipolar output transistors have a gain of 20 at the maximum current into the load, the drivers must be able to supply enough base

current to allow this. If the maximum current is 4A, then the drivers must be able to supply 200mA of base current to the output devices.

- **Prior Stages**
  The stages that come before the drivers must be able to supply sufficient current for the load imposed. The Class-A driver of a bipolar or MOSFET amp must be able to supply enough current to satisfy the base current needs of bipolar drivers, or the gate capacitance of MOSFETs.

  Again, using the bipolar example from above, the maximum base current for the output transistors was 200mA. If the drivers have a minimum specified gain of 50, then their base current will be ...

  200 / 50 = 4mA.

  Since the Class-A driver must operate in Class-A (what a surprise), it will need to operate with a current of 1.5 to 5 times the expected maximum driver current, to ensure that it never turns off. The same applies with a MOSFET amp that will expect (for example) a maximum gate capacitance charge (or discharge) current of 4mA at the highest amplitudes and frequencies.

  This is not normally an issue with valve amps, as the early stages of the amp are not loaded with any significant impedance. No further determinations are needed (other than the normal loading effects of valve stages in general).

- **Input Stages**
  The input stages of all transistor amps must be able to supply the base current of the Class-A driver. This time, a margin of between 2 and 5 times the expected maximum base current is needed. If the Class-A driver needs to supply a quiescent current of (say) 8mA, the maximum current will be 12mA (quiescent + driver base current. Assuming a gain of 50 (again), this means that the input stage has to be able to supply 12 / 50 = 240uA, so it must operate at a minimum current of 240uA * 2 = 480uA to preserve linearity.

- **Input Current**
  The input current of the first stage determines the input impedance of the amplifier. Using the above figures, with a collector current of 480uA, the base current will be 4.8uA for input devices with a gain of 100. If maximum power is developed with an input voltage of 1V, then the impedance is 208k (R = V/I).

  Since the stage must be biased, we apply the same rules as before - a margin of between 2 and 5, so the maximum value of the bias resistors should be 208 / 2 = 104k. A lower value is preferred, and I suggest that a factor of 5 is more

appropriate, giving 208 / 5 = 42k (47k can be used without a problem).

These are only guidelines (of course), and there are many cases where currents are greater (or smaller) than suggested. The end result is in the sound of the amp, and the textbook approach is not always going to give the expected result.

Types Of Amplifier Devices

For the purposes of this article, there are three different types of amplifying devices, and each will be discussed in turn. Each has its strengths and weaknesses, but all have one common failing - they are not perfect.

A perfect amplifier or other device (known generally as "ideal") will perform its task within certain set limits, without adding or subtracting anything from the original signal. No ideal amplifying device exists, and as a result, no ideal amplifier exists, since all must be built with real-life (non-ideal) devices.

The amplifying devices currently available are:

- Vacuum Tube (Valve)
- Bipolar Junction Transistor (BJT)
- Field Effect Transistor (FET)

There are also some derivatives of the above, such as Insulated Gate Bipolar Transistors (IGBT), and Metal Oxide Semiconductor Field Effect Transistors (MOSFET). Of these, the MOSFET is a popular choice among many designers due to some desirable characteristics, and these will be covered in their own section.

All of these devices are reliant on other non-amplifying ("support") components, commonly known as passive components. The passive devices are resistors, capacitors and inductors, and without these, we would be unable to build amplifiers at all.

All the devices we use for amplification have a variable current output, and it is only the way that they are used that allows us to create a voltage amplifier. Valves and FETs are voltage controlled devices, meaning that the output current is determined by a voltage, and no current is drawn from the signal source (in theory). Bipolar transistors are current controlled, so the output current is determined by the input current. This means that no voltage is required from the signal source, only current. Again, this is in theory, and it is not realisable in practice.

Only by using the support components can we convert the current output of any of these amplifying devices into a voltage. The most commonly used for this purpose is a resistor.

Common Limiting Ratings

All active devices have certain parameters in common (although they will have different naming conventions depending on the device). Essentially these are ...

- **Maximum Voltage** - The maximum voltage that may be applied between the main terminals of the device. This varies from perhaps as low as 25V (sometimes even less) for small signal transistors and FETS, and up to 1,200V or more for some valves and high voltage transistors. MOSFET voltages are typically up to about 600 to 800V for switching devices for use in power supplies.
- **Maximum Current** - The maximum current that the device may pass safely. Ranges from a few mA up to many amps. This will never be while the device also has the maximum voltage across it, as this would result in power dissipation far in excess of ...
- **Maximum Power Dissipation** - The maximum power that the device may dissipate (in mW or W), under any condition of voltage and current. (Called plate dissipation for valves).
- **Heater Voltage/Current** - (Valves). The operating voltage and / or current for the filament (directly heated cathodes) or heater (for indirectly heated cathodes). This should always be within 10% of the quoted value, or cathode life will be severely shortened.
- **Maximum Junction Temperature** - (Semiconductors) The maximum temperature that the semiconductor die will tolerate without failing. At this temperature, most semiconductors will be unable to perform any work, as this would raise the temperature above the maximum permissible.
- **Temperature Derating** - (Semiconductors). Above a specified temperature, the allowable power rating of semiconductor devices must be reduced to remain below the maximum allowable junction temperature. The power is normally derated above 25 degrees C.
- **Thermal Resistance** - (Semiconductors). The thermal resistance between

junction and case (high power) or junction and air (low power). Measured in Degrees C/W, This allows a suitable heatsink to be determined.

This is by no means all of the ratings, there are many more, and vary from device to device. Some MOSFETs for example will have Peak Current ratings, which will be many times the continuous rating, but only for very limited time. Bipolar transistors have a Safe Operating Area (SOA) graph, which indicates that in some circumstances you must not operate the device anywhere near it's maximum power dissipation, or it will fail due to a phenomenon called second breakdown (described later).

With most semiconductors, in many cases it will not be possible to operate them at anywhere near the maximum power dissipation, because thermal resistance is such that the heat simply cannot be removed from the junction and into the heatsink fast enough. In these cases, it might be necessary to use multiple devices to achieve the performance that can (theoretically) be obtained from a single component. This is very common in audio amplifiers.

Essential Electronics Formulae

There are some things that you just can't get away from, and maths is one of them. (Sorry.) I will only include the essentials here, but will describe any others that are needed as we go. I am not about to give a lesson in algebra, but the best reason for ever doing the subject is to learn how to transpose electronics formulae ! Transposition is up to you (unless I am forced to do it for a calculation here or there).

Ohm's Law

The first of these is Ohm's Law, which states that a voltage of 1V across a resistance of 1 Ohm will cause a current of 1 Amp to flow. The formula is

R = V / I (where R = resistance in Ohms, V = Voltage in Volts, and I = current in Amps)

Like all such formulae, this can be transposed (oops, I said I wasn't going to do this, didn't I),

V = R * I (* means multiplied by), and

I = V / R

Reactance

Then there is the impedance (reactance) of a capacitor, which varies inversely with frequency (as frequency is increased, the reactance falls and vice versa).

$$X_c = 1 / (2 * \pi * f * C)$$

where Xc is capacitive reactance in Ohms, π (pi) is 3.14159, f is frequency in Hz, and C is capacitance in Farads.

Inductive reactance, being the reactance of an inductor. This is proportional to frequency.

$$X_l = 2 * \pi * f * L$$

where X_l is inductive reactance in Ohms, and L is inductance in Henrys (others as above).

Frequency

There are many different calculations for this, depending on the combination of components. The -3dB frequency for resistance and capacitance (the most common in amplifier design) is determined by

$f_o = 1 / (2 * \pi * R * C)$ where f_o is the -3dB frequency

When resistance and inductance are combined, the formula is

$$f_o = R / (2 * \pi * L)$$

Power

Power is a measure of work, which can be either physical work (moving a speaker cone) or (much more commonly in audio), thermal work - heat. Power in any form where voltage, current and resistance are present can be calculated by a number of means:

$$P = V * I$$
$$P = V^2 / R$$
$$P = I^2 * R$$

where P is power in watts, V is voltage in Volts, and I is current in Amps.

Decibels (dB)

It has been known for a very long time that human ears cannot resolve very small differences in sound pressure. Originally, it was determined that the smallest variation that is audible is 1dB - 1 decibel, or 1/10 of 1 Bel. It seems fairly commonly accepted that the actual limit is about 0.5dB, but it is not uncommon to hear that some people can (or genuinely believe they can) resolve much smaller variations. I shall not be distracted by this!

$$dB = 20 * \log (V1 / V2)$$
$$dB = 20 * \log (I1 / I2)$$
$$dB = 10 * \log (P1 / P2)$$

As can be seen, dB calculations for voltage and current use 20 times the log (base 10) of the larger unit divided by the smaller unit. With power, a multiplication of 10 is used. Either way, a drop of 3dB represents half the power and vice versa.

There are many others, but these will be sufficient for now. I do not intend this to be a complete electronics course, so I will give you that which is needed to understand the remainder of the article - for the rest, there are lots of excellent books on electronics, and these will have every formula you ever wanted.

Next

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999-2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Elliott Sound Products Power Amplifier Efficiency Explained

Rod Elliott - Copyright (c) 1999

Page Last Updated 29 Jan 2000

ESP Home Page

Articles Index

Contents

- Introduction
- Amplifier Power Losses
  - Class-B
  - Class-A
- Supply Losses
- Further Reading
- References

Last Update 29 Jan 2000

Introduction

Amplifier efficiency is a topic that seems to be poorly understood, judging from the comments and questions I have had from readers, so I thought that I would put something together to explain away some of the mystery. As I progressed, it became obvious that I was going to have a lot more work to do than originally intended, since many of the basic concepts will also require explanation.

Efficiency in itself is not hard - power out versus power in. No electronic device can ever be 100% efficient, and all lost power in an amplifier is converted to heat. It makes some sense to try to make an amp as efficient as possible, but if this causes an unacceptable degradation of

the sound quality, then it matters not how many efficiency stars it has, no-one will want to listen to it.

At the other end of the scale, there are Class-A amplifiers, some of which are so inefficient that it almost defies belief. This waste of power does not necessarily translate into better sound, although reading some material would lead you to think that the lower the efficiency, the better it must sound. This is (of course) not true.

This article describes a single channel amplifier, using Class-A and Class-B. No allowance has been made for reactive loads (as occur in loudspeakers), which can increase power dissipation dramatically.

Amplifier Power Losses

So, where do the power losses come into play? We will need to look at a basic amplifier design first, so that the mechanism can be seen clearly. I shall start with a Class-B stage, as these are the most common (although very few are truly Class-B, most are actually Class-AB - but more on this later). For all calculations, a sinewave is the signal source, and for those who might be ready to claim that music is much more complex, music - and indeed any waveform - is derived entirely from sinewaves of differing frequencies and amplitudes.

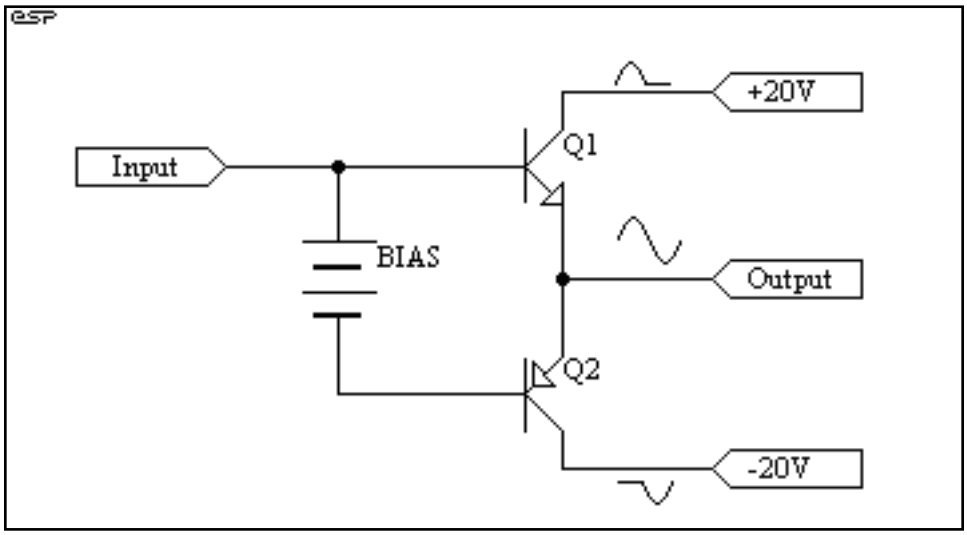


Figure 1 - An Ideal Amplifier

Class-B

We will start with the assumption that the transistors in Figure 1 are "ideal", in that there is no base current needed (they have infinite gain), and no voltage at all is lost when fully conducting. The bias voltage is adjusted so that the transistors are exactly on the verge of conduction, but no quiescent current flows. Given that the supply voltage is +/-20V, this allows

a peak swing into an ideal 8 Ohm resistive load of 20 Volts, which is an RMS voltage of 14.14 V. The peak current is 20/8, or 2.5 Amps, and this equates to 1.768 Amps RMS. The power to the load is 25 Watts.

$$P = V * I \quad \text{.... or}$$
$$P = V^2 / R \quad \text{.... or}$$
$$P = I^2 * R$$

Where P = power in Watts, V = voltage, I = current, and R = resistance. Voltage and current are RMS. Note that the term "RMS Power" is erroneous - power is the result of RMS voltage and RMS current applied to a load, and is measured in Watts. Any other "rating" (such as PMPO) is completely worthless for comparison.

The DC current waveform is shown on the diagram, and it can be seen that for each half cycle, one supply or the other supplies the current to the load. This can safely be ignored, as RMS (with a sinewave signal) implies that the waveform is symmetrical, and therefore the added individual DC currents make a "whole" RMS current.

The RMS input current (see note, below) must therefore equal the RMS output current, and we can now calculate the input power using the first equation above. The DC applied is 20V, so the input power is 20 * 1.768A, or 35.36W. The 10.36W difference is dissipated as heat in the output devices (5.18W each), and is removed using a heatsink. We do not need to take the entire supply voltage (40V) into account, since only half of it is used at any point in time.

Efficiency may now be calculated as

$$Eff = P_{out} / P_{in} * 100 = 25 / 35.36 * 100 = 70.7\%$$

This is the theoretical maximum efficiency of a Class-B push-pull amplifier, assuming no losses. Since in real life (as opposed to advertising brochures) there are always losses, this figure can only be approached - no amplifier will equal or exceed the maximum theoretical efficiency. Of course there are many amplifiers that use multiple power supplies, modulated DC, and many other electronic tricks, but at the end, when operating at full power they are all limited to this figure. Class-D (switchmode or "pulse-width-modulated") amplifiers can exceed this figure, but they are a completely different animal, and will not be discussed in this article.

Note: It is worth pointing out that using RMS current for the input power calculation is convenient, but not strictly correct. It can be shown that the average value is more accurate, and for a sinewave (the waveform of choice for all power tests) this is equivalent to 0.636 (rather than 0.707) of the peak input current. The difference is actually not at all great, especially since the amplifier should spend virtually none of its time at maximum output power

(since the likelihood of clipping is very great indeed). It is in the interests of accuracy that I let you know that ...

Average current (using the same example as above) is 2.5A (peak) * 0.636 = 1.59A, so maximum efficiency is

$$\mathbf{Eff = P_{out} / P_{in} * 100 = 25 / 31.8 * 100 = 78.6\%}$$

The "textbook" figure says 78.53%, so the above is close enough for our purposes. This still assumes that there are no losses in the devices (never the case in practice), and again is only applicable at full power. In reality, efficiency will be less, and about 70% at full power is normal for typical Class-AB amplifiers when all factors are taken into consideration.

> My thanks to Robert Milby for pointing out that average current is correct for the purposes of calculation. He also sent a spreadsheet that can be used to work this out very accurately - please contact ESP for a copy if desired.

All the above figures are quoted at maximum power, and a Class-B amp will still maintain a passable efficiency at lower powers, since the ratio of input power to output power is maintained - at least to some degree (see below). The input to output power ratio is actually quite complex. As power is reduced, efficiency is also reduced, and output device dissipation increases until a critical point is reached, after which it again falls with falling output power. The table below shows the relationship, and this must be considered when designing an amplifier, so that the safe operating power levels of the devices are not exceeded. This is based on ideal devices, so there will be about a 5% to 10% overall reduction of efficiency in real world amplifiers. In order to take the overall losses into account, I have used the RMS input current for these calculations - although not strictly accurate, this more closely approximates the actual (as opposed to theoretical minimum) losses incurred in practical designs.

Class-B Power Dissipation Versus Output

Vout (Peak)	**V**out (RMS)	**I**out	**P**out	**P**in	**Dissipation (W)**
20	14.14	1.77	25.00	35.36	10.36
18	12.73	1.59	20.25	31.82	11.57
16	11.31	1.41	16.00	28.28	12.28
14	9.90	1.24	12.25	24.75	12.50
12	8.49	1.06	9.00	21.21	12.21

10	7.07	0.88	6.25	17.68	11.43
8	5.66	0.71	4.00	14.14	10.14
6	4.24	0.53	2.25	10.61	8.36
4	2.83	0.35	1.00	7.07	6.07
2	1.41	0.18	0.25	3.54	3.29
1	0.71	0.09	0.06	1.77	1.71

Table 1

At 1/2 power (12.5W in this case), the dissipation is at a maximum, and the dissipated power exactly equals the output power, giving an efficiency of 50%. It is also apparent that at lower powers efficiency falls even more, but output stage dissipation also falls, so heating would not seem to be such a problem. However, since we are working with music in the real world, we need to consider the dynamic range (or, to be more precise, the peak to average ratio) of a typical signal. This is between 10dB and 20dB, so average power will be somewhere between 2.5 and 8 Watts. This implies from the table that the amplifier will be working fairly consistently at an average dissipation of 8 to 12 Watts at maximum power output before clipping.

This is not often taken into consideration, and many amplifiers have insufficient heatsink capacity to operate at anywhere near full power for prolonged periods. Thankfully, most music is listened to at fairly moderate levels most of the time, so the problem is rarely as bad as it looks. Sinewave testing is a most arduous test of an amplifier, quite contrary to the beliefs of some, who claim that it is too simple, and cannot give any reliable indication of the amp's performance on complex music. It is rare music indeed that will stress an amp like this, as there is a constant variation of power, allowing the heatsink time to dissipate the heat generated during loud passages.

Class-A

If the bias voltage of the amp in Figure 1 is increased to ensure that the transistor's quiescent current is about equal to the peak speaker current, then we have a Class-A amplifier. The efficiency of this amp can also be easily calculated by the same method. Since a Class-A amp conducts at all times, the DC current must be at least equal to the peak output current. A 20V peak signal will still create a 2.5A peak current in the load, so the DC quiescent current must be equal to (or greater than) this figure. The output to the load will still be 25W, but the DC input power is now 100W (2.5A * 40V). Since all of the supply voltage is in use all of the time, we must perform calculations based on this. Efficiency now (from equation 4) is

Eff = 25 / 100 * 100 = 25%

This is again the best that can be expected, and there are many Class-A topologies that are considerably worse than this. Without going into extreme detail, efficiency of Class-A stages can range from about 12.5% up to about 22% (allowing an additional amount of current to ensure that the transistors never reach a point where no current flows). The "modulated current source" (as used the the original Linsley Hood amp), can reach efficiencies up to about 30 to 35%, and although I have claimed that this is not "true" Class-A this is not strictly true, since output devices conduct for the entire waveform cycle, and such amps are capable of extremely good results. With a Class-A amp, the efficiency falls with reduced power, until at zero output power, efficiency is 0%. At an output power of 1W, efficiency is 1%, and so on (with the voltages and currents as described).

Power Supply Losses

But wait, there's more! The power supply is expected to convert it's input sinewave (at 50 or 60Hz) into a nice smooth DC that the amplifier can use. This also has losses, since the laws of physics are omnipresent. Since we used +/-20V as the power for the amp, we shall work with the same voltage output from the power supply, and will assume "typical" components.

A 14.14V RMS sinewave will provide 20V when rectified using "ideal" diodes. In real life, there will be a voltage drop of about 1V (or more, depending on current) across each diode, and also losses in the transformer, which can be expected to have an internal resistance of about 1.0 Ohm (typical of a 120VA transformer). The situation here is complicated by the fact that diode conduction is for only as very short period, so the current waveform is difficult to measure accurately. We can still use basic theory to work out what happens, and Figure 2 shows a typical supply circuit and waveforms. We will assume that the capacitance is so great that output ripple (although shown) can be considered negligible (bad news for the diodes though, since the peak curent will be very high).

As a matter of interest, the only meter that will give a passably accurate reading of the current is a moving coil (average responding) type. True RMS meters will get it (sometimes horribly) wrong. The best is to use an oscilloscope, and calculate the true average from the waveform shape and amplitude.

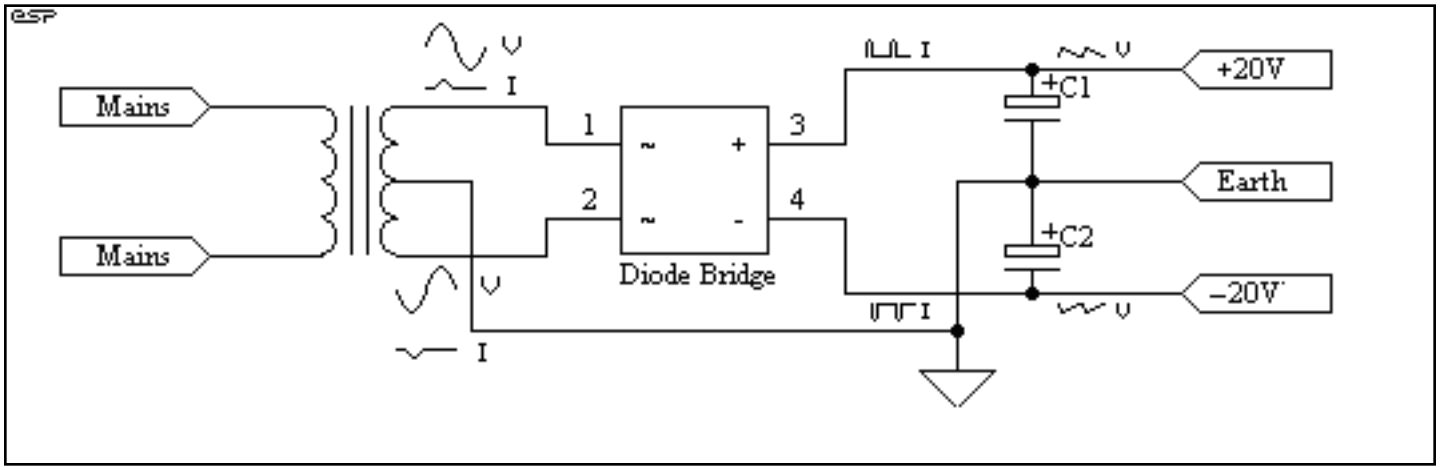


Figure 2 - Typical Power Supply

With the input of 14.14V AC, we will have a DC output of 20V (less the diode voltage drops), since the capacitors charge to the peak value of the RMS voltage. If the peak output current (on each supply rail) is 2.5A as demonstrated above, then the input power must be at least equal to the output power plus any losses. Even this configuration is open to some degree of interpretation. Is the circuit of Figure 2 a full-wave centre-tap, or a bridge. I consider it to be a dual full wave centre tap, but I could easily be wrong!.

In reality, the AC transformer current can only be accurately determined using complex analysis, but the table below is useful for passably accurate results. There are several other variations (such as choke input filters), but these are not commonly used due to the excessive cost and weight penalty, and are not shown.

Rectifier Type	Filter Type	RMS Transformer Secondary Current
Full Wave Centre Tap	Capacitor Input	1.2 x DC
Full Wave Bridge	Capacitor Input	1.8 x DC

Table 2 (1)

Assuming a Class-B amp first, 20V peak output (14.14V RMS) at an AC current of 1.768A. I have indicated that the DC current must be equal to the AC load current, so using the above table we can determine the transformer secondary current using a full wave centre tap configuration as shown above

Transformer secondary current = **DC current * 1.2 = 1.768A * 1.2 =**

2.12A
Transformer secondary voltage = **28.28 + 2 x 1V diode voltage + 2.12V** (transformer loss)
= **28.28 + 2 + 2.12 = 33V** (approx)

Since we have a transformer secondary voltage of 33, we can calculate the power rating of the transformer as

P = V * I = 33 * 2.12 = 70VA

This seems like an extraordinary amount of power needed for a simple 25W amp, and in reality it is not needed. A smaller transformer can be used, and although it will not be capable of providing enough current to charge the capacitors on each cycle, (causing the supply to droop), the maximum power will only be needed for relatively short periods. We also have to accept that there will be resistive losses in the transformer, and the resistive and junction voltage losses in the rectifier bridge.

The power lost by the transformer winding resistance and diodes is (typically)

(2V + 2.12V) * 2.12A = 8.73W

which again is dissipated as heat. These losses will decrease as amplifier power is reduced, but at full power from a little 25W amplifier we have lost about 19W in heat from the amp and its power supply. Needless to say, this is much worse with a Class-A amplifier, as shown below (using much the same calculations as before, but at the higher current).

Transformers also have "iron loss". This is power that is needed to maintain the magnetic flux in the transformer core, and is always present. As the load is increased, the magnetising current becomes less significant, and at full power, the resistive losses in the windings are usually far greater than the iron loss. It is the magnetising current that causes a transformer to get warm even with no connection to the secondary winding. Typically, small transformers have greater iron and copper losses than large transformers per volt-amp of output (i.e. they are less efficient).

Class-A amplifiers will operate at the full DC load all the time. This means that at least 2.5A DC is needed on a continuous basis, requiring a 3A rating for the transformer secondary. Since some extra voltage might be needed to allow for (say) a capacitance multiplier filter, we will give an extra 3V RMS each side. This means that the AC voltage needs to be 38V centre tapped, at a continuous rating of 3A. Using the same formula as above, this gives

VA = V * A = 38 * 3 = 114 VA

This would suggest that a 120VA transformer is needed. The losses are now much higher, too, and are continuous where before they were transient. With the 38V centre tapped transformer, we can expect about 25V DC before the filter - assuming one is used. We still have the diode and transformer losses which will actually stay about the same, since the bigger transformer will have less resistive loss, but the diodes will have more. To get the 20V supplies, we must drop about 5 Volts in the filter, so

`P = 5V * 2.5A = 12.5W` (each supply, +ve and -ve)

This has just added 25W filter dissipation and 10W rectifier and transformer loss (a total of 35W) to the already low efficiency of a Class-A amplifier. Total amplifier input power is 135W, to get 25W of audio. This will be greatly increased if you were to use a regulator, since the input voltage must be high enough to allow for power line droop (typically up to 10%). I shall leave this as a reader exercise to work this out :-)

Further Reading

From the above, you can now see where the power goes to. There are losses everywhere, and it is essential that the amplifier builder has a reasonable knowledge of losses and dissipation requirements for an amplifier before starting construction.

For further reading, see Amp Design and Heatsink Design, both of which will assist in the final determination of the supply and heatsinking requirements for your next project.

References

1. Voltage Regulator Handbook - National Semiconductor 1982
2. Amplifier Design (ESP web page)
3. Heatsink Design (ESP web page)

ESP Home Page

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Update Log: 29 Jan - added missing waveforms to drawings, fixed a few errors and ambiguous statements./ Page created and first published 19 Nov 1999

The Audio Pages

Elliott Sound Products **Troubleshooting & Repair Guide**

Copyright © 2003 - Rod Elliott (ESP)
Page Created 24 Apr 2003

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Common Problems
  - 2.1 Bad Solder Joints
  - 2.2 Incorrect Components
  - 2.3 Oscillation
  - 2.4 Schematic Errors
- 3.0 Troubleshooting Tools
  - 3.1 Multimeter(s)
  - 3.2 Signal Source
  - 3.3 Oscilloscope
  - 3.4 Dummy Load
  - 3.5 Power Supply
- 4.0 Most Common Failures
  - 4.1 Shorted Supplies
  - 4.2 Output Stuck to Supply
  - 4.3 Distortion
  - 4.4 Spontaneous Failure
- 5.0 Voltage Measurements
  - 5.1 General Principles

- 5.2 Example
- 5.3 Summary
- 6.0 Component Testing
  - 6.1 Transistor Quick Test
  - 6.2 Other Components
- 7.0 Opamp Tests
- 8.0 Earthing (Grounding)
- 9.0 Removing Dead Components
- 10.0 Testing
  - 10.1 Power Up
  - 10.2 Voltage Checks
  - 10.3 First Power Test
  - 10.4 Final Test
- 11.0 Preamplifier Tests
- 12.0 Terminology
- 13.0 Conclusions
- Copyright

1.0 Introduction

Having finished your masterpiece, it is with some distress that you find that it doesn't work. Such failures range from instantaneous meltdown upon power-up (remember the safety resistors I always suggest? Now you know why!), through to strange noises, intermittent behaviour, etc., etc.

It is not possible to write an article that covers every possibility, but hopefully the material presented here will get you underway with as little pain as possible.

One of the points I have made in a number of places on my site is that if the project does not work, you have almost certainly made a mistake. While I will usually do what I can to help you to get the project working, there is only so much I can do, and it is your responsibility (*not* mine) to find out what you did wrong.

Since basic troubleshooting techniques are not widely known - or so it seems, I get a great many requests for help, and have to try to diagnose what went wrong from the description I am given. It almost goes without saying that many of the descriptions leave me wondering what I *am* being asked. This is not to blame the person asking, but shows

that even terminology can be very misleading. There is a *huge* difference between "hum" and "buzz", but if you don't realise that, then I have to either figure out the most likely (correct) term from the description, or ask.

This information should be read in conjunction with the Amplifier Design and How Amps Work articles, and you *will* need a copy of the amplifier's circuit diagram (schematic) as reference. I have been very general in this description, since there are a great many amplifier designs both on my site and elsewhere, and if the descriptions were specific to only one design, you may have great difficulty when working on something different.

To be excellent at fault finding, you need to understand how the circuit works - this allows you to make informed decisions, to know what to look for (and where), and to recognise instantly if a voltage reading is right or wrong. I never said that this was simple!

I have made one major assumption in this article - the forward voltage drop of a diode (or a transistor junction) is nominally 650mV (0.65V), but it can vary by a considerable margin. In most descriptions that follow, I will assume 650mV, but expect to see anywhere between 0.55V to 0.75V, depending on type, current, etc.

2.0 Common Problems

There are quite a few common issues that you will be faced with from time to time. If one channel of a stereo project works and the other does not, this makes fault finding a lot easier, since you have a reference that you can use. This applies to voltage readings, resistance measurements, etc., and also eliminates some of the more common errors - for example leaving the zero volt return line off when the amp is wired into circuit. (If you do that, neither channel will work.)

Having said that, with any of the PCB projects on The Audio Pages, if it does not work, then you have made a mistake. There are some occasions where new components are faulty or incorrectly marked, but other than fake power transistors these are very rare. You do need to be aware that new components can be faulty, but in general, suspect your own work first.

The following basic guides indicate some of the most common project failures ...

2.1 Bad Solder Joints

When completed, a solder joint should be clean, shiny, and show a perfect adhesion to both the component lead and the PCB. If there is any sign of the solder being "frosty",

sitting on the PCB as a "blob", or not flowing up the component lead in a nice smooth arc, then the solder joint is incorrectly made. It may appear to work, but the contact is/may be pressure based, rather than alloy based as it should be (solder forms an "alloy", or molecular metal bond between the solder, component lead and PCB). For an excellent tutorial on basic techniques (and what *not* to do), see www.epemag.wimborne.co.uk/solderfaq.htm. There are a great many such sites on the web, and a web search for "soldering techniques" will find you a broad cross section for reference.

The most important thing about making an excellent (as opposed to disgusting or barely ok) solder joint is cleanliness! The component leads, PCB and soldering tip must be completely free from any contamination - burnt flux, melted plastic, oxides, old solder, etc. must be removed. Make sure that the component cannot move as the solder cools, and ensure that your soldering iron (or station) can supply the right amount of heat. To much heat will burn the flux (and even the solder itself!) and may damage the component. Too little heat makes a "cold" joint, where the solder just sits in blobs but does not make a metallic bond.

2.2 Incorrect Components

All components must be inserted in the correct place, as shown on the PCB overlay and/or other instructions. While this may seem obvious, it is the most common form of "component failure" - the component is not faulty per se, but if it is in the wrong place it will affect the operation of the circuit. This is made worse by the fact that many components use "strange" markings, and it is not always easy to figure out what the value is supposed to be.

With resistors, unless you know the colour code very well, it is a good idea to measure all colour coded resistors before insertion. This is especially true with 1% 4-band codes, as they can be very confusing - even for professionals! There is some information about basic components, markings, etc., in the articles section of this site (see Articles). This is not comprehensive, and cannot be - there are just too many different devices available to cover them all.

Always, *always*, make sure that you download the manufacturers data sheet for transistors, ICs etc. It is not uncommon that suppliers will substitute brand name parts with "equivalents" - these may (or may not) be as good as the original, but they may also have different pinouts. The only way to know for sure is to get the data sheet from the company who actually made the device you have - this applies mainly to semiconductors, but also may be of concern for relays, some electrolytic capacitors (especially power supply filter caps), and other components as well. For semiconductors, most will be fine, but expensive power output transistors are regularly

counterfeit! See Counterfeit Transistors for more information on this topic.

Occasionally, you will get a brand new, brand name component that is faulty. Irritating? Of course it is, but also inevitable. This is where you really do need to hone your fault finding skills, since it is clearly not the result of a mistake on your part. These faults can be difficult to find, and require a disciplined approach to troubleshooting to repair.

2.3 Oscillation

The zobel resistor in most amps is at the output, and is in series with a cap - typically 100nF, but this varies. If the resistor goes up in smoke and/or the amp gets hot *fast*, either the amp is oscillating, or you are trying to amplify too high a frequency.

Oscillation is caused by using the wrong value compensation cap (typically between 47 and 220pF), or (and more likely) having input wiring too close to speaker wiring. Input cables to power amps should always be shielded, and kept as far away as possible from DC power leads, mains transformers and wiring, speaker leads and connectors, etc. In some cases, it may be necessary to provide shielding between the input circuits and power amp.

During testing, the heatsink may not be earthed to the power supply common. In some cases, this can cause oscillation because the heatsink acts as an antenna - as does the input lead if it is not shielded. Always earth the heatsink - even for just a quick test.

Speaking of heatsinks, *never* operate any power amplifier without its heatsink. The devices can overheat very quickly, and are easily damaged by the excess temperatures. A small clamp may be used to attach a temporary heatsink if you are in that much of a hurry, but make sure that you monitor temperatures carefully.

2.3 Schematic/ PCB Errors

I am pleased to be able to say that there are few (if any) schematic errors on The Audio Pages. This is not always the case however, and there are many errors to be found in schematics on the Web (some circuits published won't work at all, or will stress all components way beyond their ratings), and even established and normally reliable magazines can (and do) make mistakes. Sometimes these mistakes will prevent a circuit from working at all, so be warned.

While there may be the occasional PCB error in some of the project PCBs, the error is clearly explained in the construction notes, and will usually only be minor - major mistakes require the artwork to be re-done (which is expensive), but no ESP circuit board requires any modification - track errors are fixed, regardless of the cost.

It goes without saying that if you do find a mistake in any of the ESP projects (PCB, schematic or construction details), then please let me know - this is the main reason that there are so few - people *do* let me know, and I appreciate the feedback. If you find a mistake in someone else's project however, I *don't* want to know - tell the author - not me.

3.0 Troubleshooting Tools

3.1 Multimeter - To be able to do even the most basic fault finding, you will need at the very least a multimeter, and preferably two. Most people prefer digital meters, but if you know how to use an analogue meter you may find things that a digital will miss.

You need to be able to measure ...

- Volts, both AC and DC, from a few millivolts to 100V (or more)
- Amps, DC only is sufficient, but preferably up to at least 2A
- Ohms, from less than 1 ohm up to 10 megohms
- Other functions (transistor tests, capacitance, frequency) are useful, but not essential

3.2 Signal Source - You also need a signal source. While a walkperson (for example) is useful, it is not a good source of proper test signals, and is therefore limited. There are several PC based audio oscillators available on the Web, and these are fine (if a little inconvenient). Ideally, an audio oscillator should be used, see the Projects Pages for details of test gear you can make quite cheaply.

3.3 Oscilloscope - For some tests, an oscilloscope is almost essential. While few hobbyists can justify the purchase of such an expensive piece of test gear, for many professionals the CRO (Cathode Ray Oscilloscope) or 'scope, is the first thing that is attached to anything that does not work. Again, there are many PC programs that allow you to use your computer as a basic oscilloscope. By nature, most soundcards are limited to 20kHz upper frequency, so such PC based tools will not find all problems.

> **Warning:** An oscilloscope cannot be used in the same way as a multimeter (unless a self contained hand-held unit is used), since one probe terminal is connected to the chassis, and thence to mains safety earth. Never, ever disconnect the safety earth from an oscilloscope - this is an invitation to disaster, death and/or destruction of something or someone, at some time. This is an

extremely dangerous practice.

3.4 Load - A "dummy load", usually a high powered resistor or bank of resistors, and ideally switchable to 4 or 8 ohms. This enables you to perform full power tests without the noise, and if a fault develops, the load just gets hot, but your speakers do not get fried. If desired, you can have a 47 ohm 10W resistor from each terminal of your load to an external speaker, so you can monitor the output signal.

A load resistor bank can also be immersed in oil (light engine oil is fine) or water if you have a lot of power to dissipate. Water is the best for removing heat, but may cause corrosion. Glycol (engine coolant) would probably be good too - I've not tried it, but I don't envisage any problems.

3.5 Power Supply - A bench power supply is immensely useful, but possibly even more useful is a variable voltage transformer ("Variac™"). This allows you to make any power supply variable, and the amplifier voltage can be slowly increased while monitoring the amplifier's output voltage (and supply current with your second multimeter). Another useful test tool for those who cannot justify the expense (again, Variacs are not cheap) is a "lamp lead" - a standard light bulb (100W is usually about right), carefully wired in series with a mains lead (and properly insulated!). An amp with a short circuit fault will cause the lamp to glow at full brightness, but a normal load will cause the lamp to flash brightly for a moment, the settle down to a steady dull glow.

One of the most essential power supply tools is a pair of 10 Watt resistors, between 10 and 22 ohms (or as suggested in the project article). These must be used in series with the supply leads before applying power, and limit the current to a (hopefully) safe value, especially when used in conjunction with a Variac or lamp lead.

Now that you have the fault finding tools, we can continue on to making some actual measurements.

4.0 Most Common Failures

This is the part where it all comes together. The first thing to do when you know an amp is faulty, is to determine the exact nature of the fault. Does it short out the supply (safety resistors get hot), or does the output swing to one rail or the other and refuse to leave? Perhaps it seems to be alright, but is badly distorted. Make sure that you identify the fault completely - there is no point chasing a fault that was incorrectly diagnosed!

4.1 Shorted Supplies

First, let's look at a "shorted" supply. This is most commonly caused by shorted output or driver transistor(s), but may also be the result of any of the following ...

- Incorrectly installed transistors - PNP instead of NPN (or vice versa), either as drivers or output devices
- Shorts between transistor case and heatsink, due to a punctured mica washer
- Open circuit bias servo circuit. The bias servo is the transistor and pot that generates the bias voltage needed to keep the transistors conducting at just the right level to avoid crossover ("notch") distortion. An incorrectly installed transistor, faulty (open circuit, wrong value, or improperly adjusted) pot, dry solder joint or broken track can all cause the output transistors to turn on fully when power is applied. In some designs, the bias "servo" is simply two or more diodes, and may also have a series resistor.
- Solder bridges between tracks or component pads.

The first thing to determine is if the short is "hard" or "soft". A hard short will show up as a very low resistance between the supply rails (less than 1 ohm), when measured with a multimeter with no power applied. Hard shorts always indicate either blown transistors, solder bridges or punctured mica washers. If you are lucky, it will be either of the second two, but don't get your hopes up. Hard shorts are unusual in an amp that has just been built and is being tested for the first time (using the safety resistors!).

A soft short is identified by the fact that a resistance measurement between the supply rails to each other, the output and earth (ground) does *not* show a very low resistance (less than (say) 650 ohms or so). Resistance readings of around 600-700 ohms are possible in one direction (this is actually a voltage, and is developed across diode junctions either on real diodes, or within the junctions of transistors). Resistance may be the same or much higher in the other direction - swap the meter leads for all such tests so you measure with both polarities. You almost certainly have a component (power transistor or driver) installed incorrectly if you get a soft short, but a faulty bias servo will create the same effect.

If you can vary the voltage, determine the voltage where the soft short comes into play. It is very rare for soft shorts to be present at extremely low voltages (less than +/-1 or 2 Volts), but if it is, then something is installed incorrectly.

See Component Tests (below). These techniques will isolate 99% of all soft short problems.

4.2 Output Stuck to Supply

When the output voltage "sticks" to one supply or the other, there are (as always) several possibilities. In order of likelihood, these are as follows ...

- Incorrectly installed components
- Solder bridges between tracks or component pads.
- Dry (or cold) solder joint(s)
- No earth return between amplifier and power supply
- Broken tracks
- Faulty transistor(s)

If one output or driver transistor is shorted, this does not cause rail sticking, it causes a soft short. Rail sticking can be the result of an open circuit transistor, possibly in conjunction with its opposite being shorted. These faults can be found with a multimeter (as described above). It is important to eliminate blown devices early, or you will spend a lot of time trying to find the problem in the wrong place. A common error is to leave off (or forget) the power supply earth return - this gives a similar effect to supply rail sticking, but it usually happens slowly (several hundred milliseconds to several minutes)

A stuck rail can be caused by any of the following faults closer to the input ...

- Open circuit feedback resistor (or track)
- Open circuit (or simply non conducting) class-a driver transistor
- Open circuit (or simply non conducting) current source/sink transistor
- Open circuit bootstrap resistor chain
- Solder bridges between tracks or component pads.
- Incorrectly installed transistors, diodes, LEDs, etc. (As always, and anywhere)
- Non-functional long tailed pair input circuit/ error amplifier.

Again, the hard part is finding the fault, and this is where the next section will be useful. The most common problem by far is still incorrect components, but when a visual check fails to find the problem, then you need to measure voltages.

4.3 Distortion

Distortion comes in many flavours, but may be roughly categorised as "gross" or "subtle". Both are in reality "gross", but from a testing perspective it is essential to separate the two somehow. I would consider gross distortion as being a state where only half of the signal is reproduced. From a listening perspective, this is way beyond mere gross - it is totally unlistenable! "Subtle" distortion is also unlistenable, but some people don't notice it (true).

Should only half the signal be reproduced (or a small amount of one polarity and the full amount of the other), then you almost certainly have an open circuit somewhere in the driver or output stage. It may be an open transistor (rare) but is more likely to be ...

- A bad solder joint, leaving part of the output stage inoperable
- Incorrectly installed components (as always)
- A fault in the Class-A driver circuit, such that there is insufficient current to drive one output or the other

This is where an oscilloscope is almost essential - faults of this nature are very hard to diagnose if you can't see the waveform. The above fault list will help you to solve most gross distortion problems relatively easily.

Subtle distortion is more insidious, as there are several possibilities. Again, very difficult to determine without an oscilloscope, but voltage measurements *will* isolate some of the more likely issues. The things to look for are ...

- Incorrectly set quiescent (idle or bias) current
- Faulty or shorted bias servo
- Solder bridges between tracks or component pads.
- Parasitic oscillation

The first three are easy enough to test, requiring only a multimeter. A few measurements will isolate the problem fairly quickly, and all should be well.

Parasitic oscillation is a lot harder, and usually needs an oscilloscope. I can say with confidence that the designs on the ESP website will be free from parasitic oscillation provided that all normal precautions against continuous oscillation are taken - earthed heatsink, shielded input leads (separated from output or DC wiring), and input connectors a sensible distance from output connectors (or shielded by an earthed metal cover).

Also, make sure that bypass capacitors are fitted as required, and keep DC supply lines as short as possible. Distortion tests will almost *always* require a load to show up. While a small amount of distortion may be visible with no load, most will reveal themselves either fully or partially with as little loading as 20 ohms or more at the speaker output.

4.4 Spontaneous Failure

The amplifier has been working for some time (from minutes to weeks) then fails. You

have eliminated nearly all of the potential construction faults, since the amplifier has shown that it does (or did) work. Unfortunately, this does not make your job any easier.

One of the most common problems in the case of spontaneous failure is counterfeit power transistors. See the article Counterfeit Transistors for more information on this topic. Other things to look for are ...

- Infant Mortality - the term usually used to describe components that fail a short time after a device is first used. Most common failures are semiconductors - specifically transistors or ICs, but other components may be affected as well (electrolytic capacitors - rare but it happens, diodes, zeners, etc.) Infant mortality failures are not as uncommon as you may hope for, but are still comparatively rare. This is completely normal (albeit very irritating).
- Dry solder joints - seemed Ok when first tested, but failed after use (not uncommon!)
- Excessive supply voltage, causing component stress
- Inadequate heatsinking, allowing components to overheat
- Incorrectly set bias current (too high), causing overheating
- Incorrectly mounted power transistors, with inadequate thermal contact to the heatsink
- Short circuited outputs, usually the result of inadequate (or no) insulation, or changing speaker leads around with the amp on and with a signal present - not recommended, even if an amp has short circuit protection

Since most cases of spontaneous failure result in shorted power transistors, these are usually easily found with a multimeter. The amp's fuse(s) may blow, but transistors (aka "3-legged fuses") are much faster than any conventional fuse :-(If a fuse blows, apply proper testing procedures (look for shorts, etc) rather than replace the fuse and hope that all will be well. This rarely happens, but additional damage (and to more components) is common.

For all projects on The Audio Pages, there are quite specific absolute maximum supply voltages, and minimum specified load impedances (which may vary with applied voltage). It is extremely important that this information is adhered to, or individual device ratings may be exceeded, resulting in premature failure. None of the project amps is designed to work with a 2 ohm load, and simply adding parallel output transistors (for example) just makes the driver transistor subject to failure, and having failed, it will nearly always result in output device failure as well.

It is worth noting that the principle failure mode for a BJT is short circuit. Open circuit devices will be found, but this occurs when another device shorts, and the internal bonding wires then fuse. The device then measures as open circuit, but the die (the silicon "chip" inside the transistor) has failed short circuit. External fuses are *not* intended to protect the transistors - they are there to prevent catastrophic failures (including fire) should an output device fail.

When one or more output devices fail, it is usually a good idea to replace all output and driver devices, even though they may seem to be alright. It is almost certain that they have been stressed, and may be more prone to failure at some later date. In some cases, an output stage fault can also damage the Class-A driver (and/or the current sink, if used). This rarely affects the input stage, which normally survives even the most destructive failure. Note that in some cases the fault current can be so high as to open-circuit the emitter resistors (usually not all, but one or two can fail). *Always* check these if an amp fails, preferably after you have removed the power transistors and drivers.

Unless the failure can be positively attributed to counterfeit transistors (which will fail at much lower power levels than the genuine device), try to determine exactly what went wrong *before* re-commissioning the amplifier. Check speaker leads, supply voltages and speaker impedance - something caused the amp to fail, and it is better fixed than allow it to happen again.

5.0 Voltage Measurements

Voltage measurements must be done with the greatest of care. A simple and cheap fault can easily turn into a complex expensive one with just the slip of a probe!

In keeping with the general nature of this article, I will not refer to any specific voltages, but will rather give an overview of what to look for. At this point, a good understanding of the basics of transistor operation is expected and necessary, otherwise you will not be able to understand what you are seeing on you meter or oscilloscope.

Always measure your supply voltages first !

Countless man-hours (person-hours?) have been wasted trying to locate "bizarre" faults, when all that has happened is that the supply voltage(s) are either not present or are incorrect. This is the very first voltage measurement you should make - always !

5.1 General Principles

In the most general of terms, with any bipolar transistor (FETs and MOSFETs are completely different!), there should be about 600-700mV measured between the emitter and base, and in linear circuits (such as conventional amplifiers) there will be some higher voltage of the same polarity between emitter and collector as that between base and emitter. For example, on a PNP transistor, with the red meter lead to the emitter, there will be around -650mV between emitter and base, and anything from (negative) a few volts to several tens of volts between emitter and collector.

An oscilloscope will show perhaps almost no AC voltage at all on the base, but a large AC signal on the collector - this is usually quite normal. The DC voltage readings will tell you if the transistor is correctly biased, and therefore able to do its job. A voltage of 650mV between emitter and base, but full supply voltage on the collector is not necessarily wrong - you must read the voltage with reference to the circuit diagram.

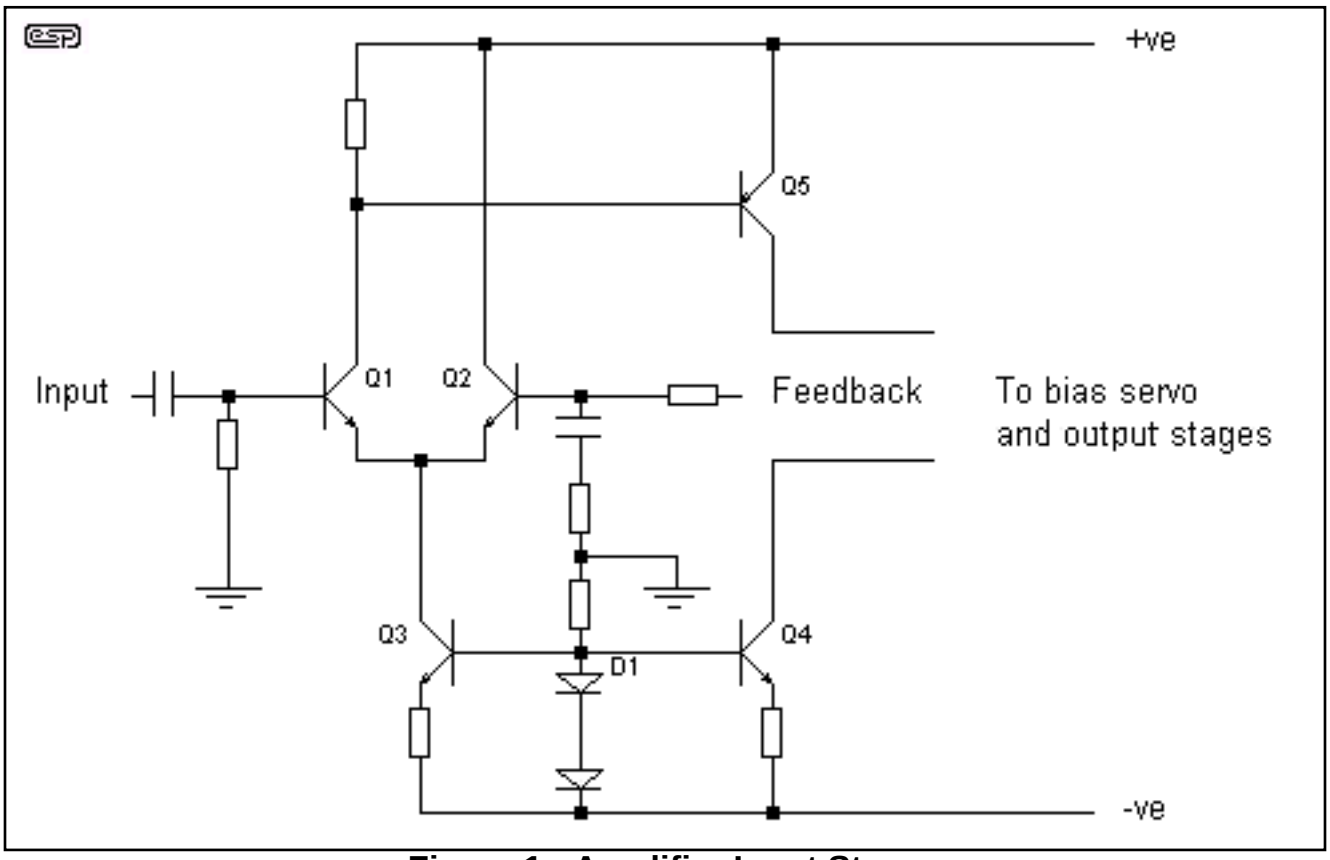


Figure 1 - Amplifier Input Stage

5.2 Example

Let's assume for a moment that you have a conventional NPN long-tailed pair for the input circuit (Q1 and 2, Fig 1). The emitters are tied together, with perhaps small resistance values in series with each emitter in some designs. The voltage at the bases will probably be a few millivolts negative, and the emitter to base voltage should be

around 650mV. The collectors will be at almost the full supply voltage in most circuits (there are exceptions though). If you were to see that the output was stuck to one of the supply rails, then that will upset the long-tailed pair, and all voltages will be wrong. This could mean that one of the long-tailed pair transistors is faulty, but maybe not!

This is where you need to play detective, to ascertain *why* the output is stuck to the supply rail (having eliminated all the previous fault types - incorrect components, bad solder joints, etc.). The next device to test is the class-a driver (Q5). Check the emitter-base voltage, and make sure that it is around 650mV. If that is correct, then the collector should be at close to zero volts, but it won't be. Instead, you may find that it is sitting at (or near) one supply rail voltage. Look at the circuit - the class-a driver is PNP (using the previous example) and the collector is at full positive supply, that means that the transistor is fully turned on ... why? Or is it?

The next step is to look at the current sources (Q3 and Q4). Between the emitter and base of each there should be 650mV or thereabouts, and the current through each is easily determined. Measure the voltage across each emitter resistor - it should be about ... 650mV (can you see why this would be so?). The current is equal to V/R so if the emitter resistor is (say) 100 ohms, then the current should be 0.65/100 = 6.5mA (close enough).

The collector of Q3 should be at around -700mV, and that of Q4 at around zero volts. If this is the case then the amp should be working. Assume that the collector of Q5 is almost the full supply voltage, and likewise that of Q4 - there are either of two possibilities - Q5 is shorted (or turned on fully), or it has no collector current. The job of Q5 is to pull the output high as it turns on, and let it swing low when it turns off, but if Q4 were supplying no current, then the output will swing high. The input stage will try to turn Q5 off, but will become unbalanced by the voltage at the feedback input. This will make the circuit inoperable until the fault is located - this is your mission, should you choose to accept it, of course ;-)

So, Q5 has full positive supply at its collector, give or take a volt or so (not important at this stage). The collector voltage at Q4 should be about the same, and the current should be about 6.5mA, but wait! If everything were working as it should, the amp would be functional, so there is something amiss - but we knew that already. What is the voltage at Q4's collector? Is the voltage across Q4's emitter resistor 0.65V as it should be?

If the collector voltage is near the negative supply rail, or the emitter voltage is a lot lower than 0.65V, then Q4 is open circuit at the collector - this is not a common failure

mode for a bipolar transistor, so it is likely that there is a bad solder joint at the collector of Q4 (or perhaps a hairline crack on the PCB). If the collector voltage were at close to positive supply, then the emitter resistor could be open - probably a bad solder joint, as resistors rarely go open without a lot of smoke and fuss. Check the value carefully - was a 100k resistor inserted by mistake?

5.3 Summary

The purpose of this exercise was to demonstrate the general processes of elimination that should be used to locate the type and nature of a fault, and then it can be easily corrected. It is not possible to cover every possibility here, even with the simple circuit shown, but by carefully measuring the voltages you will be able to track down the most likely cause, without having to rebuild the whole circuit!

The answer to the little riddle above ... There must be about 650mV across the emitter resistor of the current sink because there are two diodes in series. D1 balances out (or "cancels") the emitter-base voltages of both Q3 and Q4 - also 650mV. Whatever voltage exists across D2 (and we know it must be 650mV), must also appear across the emitter resistors. It really is that simple, but it may take a bit more experience before you see it clearly.

A useful thing to remember about transistors - if it gets hot, it is working (or trying to). Looking at Fig 1 again, if Q4 gets hot and Q5 is dead cold, then Q5 is the faulty device - not Q4 as you may think at first.

6.0 Component Testing

Transistors can be tested for basic functionality with a multimeter. If you use an analogue meter, be aware that when on the ohms range, the red probe is negative. Digital meters retain the "correct" polarity. A BJT (Bipolar Junction Transistor) can be thought of in terms of two diodes, as shown in Fig. 2. As with any diode, they should conduct in one direction, but not in the other. All BJTs may be tested this way, revealing open circuit, leaky or shorted junctions. The test tells you nothing about gain, voltage breakdown, or anything else, only that the device is *likely* to be functional.

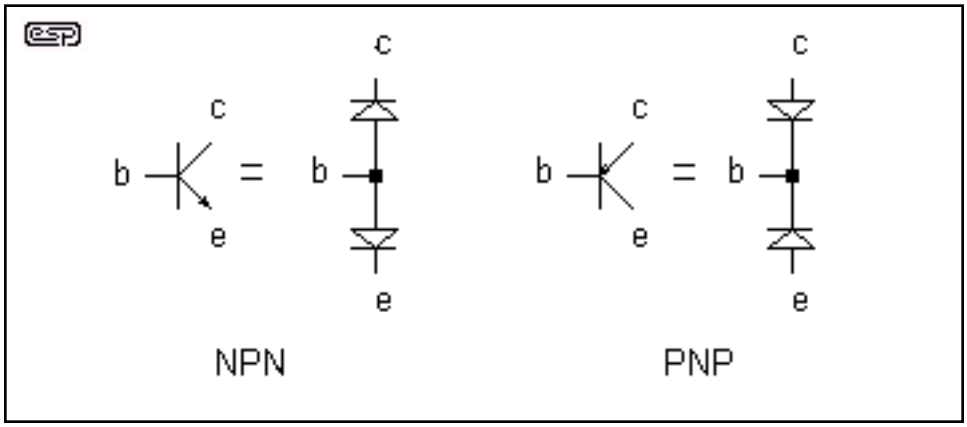


Figure 2 - Basic Transistor Test Model

6.1 Transistor Quick Check

Check in both directions with your multimeter between the base and emitter/collector of each power and driver transistor. An NPN transistor will show a "resistance" of 600-700mV (shown as ohms, but is actually voltage with 99% of digital meters) with the positive (red) lead of a digital meter connected to the base, and the black lead on emitter and collector. Reverse red and black and measure again - in some cases, one connection may still show 600-700mV because of a connected power or driver transistor - this is normal.

By using this method, the proper conduction of each diode can be checked - as with any diode, the forward voltage drop is around 650mV (which as explained above shows on most digital multimeters as 0.65k ohms), and the reverse bias condition should show infinity(keep your fingers away though). In-circuit tests can also be done like this, but the results may be misleading because of other devices in the circuit.

In case you were wondering (and you are by no means the first to do so), you cannot use two ordinary diodes wired as shown as a transistor. Transistor operation relies on the junction between the "diodes" (hence bipolar *junction* transistor).

6.2 Other Components

Resistors should read their correct value, but again, in-circuit tests can be misleading. All diodes should show proper conduction and blocking as the probes are switched from one end to another. This is not a useful test for LEDs or zener diodes, but at least you will know if it is open or short circuit.

Capacitors really need a capacitance meter (as well as an ESR [Equivalent Series Resistance] meter) to test properly, but you can still get a fair idea with a multimeter. Shorts are uncommon in film caps, but can occur, although in most projects this is highly unlikely. Electrolytics should show a low resistance at first, which will rise as the cap

charges. Reverse the leads and make sure that the cap discharges (expect to see silly resistance values at first), and charges up again. Low voltage reverse polarity will not harm electros.

Most other components (transformers, connectors, wiring) needs only to be checked for continuity, and that all wiring is connected to the proper place.

7.0 Opamp Circuits

There is not much that can go wrong with an opamp circuit. Most linear circuits (as used in preamps) have one thing in common - the two inputs should be at almost exactly the same voltage, and so should the output. The most common problem is oscillation - especially with very fast opamps. The ESP boards are designed so that bypass capacitors are as close as possible to the opamps, and there is also additional filtering using small electrolytics.

It is still possible to make an opamp circuit oscillate though, so sensible precautions should be taken - keep inputs and outputs shielded and apart, and always use a 100 ohm resistor in series with the output of any opamp that connects to a cable - regardless of length.

Other problems can occur, but normally they will be the result of bad solder joints (as always), or incorrectly installed components. All ESP boards will function first time, every time if assembled according to the instructions, but if yours doesn't, then there is a mistake in the component placement, or the opamp is faulty. Yes, opamps can be faulty from new - it doesn't happen very often, but it does happen.

As with power amps, leaving off (accidentally or otherwise) the power supply earth (zero volt) line is quite a common "fault". A +/- supply means that the earth (or ground) lead is required - it is *not* optional!

8.0 Earthing (Grounding)

This is the area where it all goes to pieces. Hum or buzz is the usual symptom, but unfortunately, there are no fixed rules that can be applied in all cases to cure the problem.

The distinction between a "hum" and a "buzz" is extremely important! If you describe a

noise as a hum, then the expectation of anyone knowledgable in the field will think "low frequency, no (or few) harmonics". This describes the noise made by an earth loop - a situation where two or more pieces of circuitry are joined by the mains safety earth lead and the shield of an interconnect (for example), forming a loop. This can inject a very low voltage (but sometimes surprisingly high current) into the loop, and the signal is picked up by the inputs. You hear hum - a single low frequency tone.

"Buzz" has a sharp edge to it - there is usually a low frequency component, but it has a hard sound that may even be audible in tweeters at times. Buzz is caused by any number of things - input leads close to mains wiring, power transformer or bridge rectifier (and associated wiring), bad or no earth connection, loops (they can cause buzz as well as hum), the list is almost endless.

Sporadic oscillation in an amp can also create a buzz or hum in some cases - follow the guidelines above to ensure that the amp is stable under all conditions - low level oscillation can usually only be detected with an oscilloscope, but you may be able to detect it using an RF "detector" probe - see the projects page for a suitable example.

With any of these problems, it is almost impossible to give a standard "fix". The solution is different in nearly every case, and sometimes the best result is obtained with an arrangement that should not work at all. My normal approach is to keep lots of separation of input cables from anything else, and for locating the optimum earth location, I use the following methods ...

- Make a "soft" earth connection with a 10 ohm resistor, usually from the amp's input to a convenient place on the chassis
- Get an alligator clip lead, connected to the amp's input earth point, and probe likely places for low noise - the star earth point for the filter caps, right next to the input connectors, etc., until the quietest place is found. Just hope that it doesn't end up being to some part of your anatomy or the cat :-)
- Make the "sweet spot" a solid connection, and do tests to determine if there is any possible way to improve matters.

This method *usually* works well, and if you really do find the optimum location, you will need another amplifier to be able to hear any noise. It should be possible to earth the end of the input lead to any of your other equipment without adding noise, but there is a point where it is fruitless to try to make it better.

If you have to put your ear right next to a speaker to hear anything, then from the listening position it is effectively completely silent. Further improvements will not yield

any audible benefit.

A common mistake (and an excellent source of undesirable noises) is to take the DC from the rectifier. DC must always be taken from the filter caps, and never from the rectifier. Those short leads will develop serious noises when the amp draws current, and may create a background "haze" that is audible as a background noise from the amp - but only when it is playing! This is very insidious, since the amp seems dead quiet with no signal. An oscilloscope and/or distortion meter is essential for locating problems of this type.

9.0 Removing Dead Components

Once you have determined that a component is dead (or is probably dead), you need to remove it from the PCB. *Never* attempt to just heat the leads and prise them from the board, and resist the temptation to use a solder sucker (or solder wick) to remove the solder prior to prising the component out. This will nearly always result in damaged PCB pads and tracks. It is far better to use a very slim cutter, and cut off the legs first. Cut as close to the PCB surface as you can (for both component-side and copper side mounted parts - be careful that the cutter does not damage the pad or track!), and then use a sucker or wick to extract remaining solder and the remnant of component lead.

While desoldering, a clean soldering tip is just as important as when soldering. A clean (and properly "tinned") tip requires less heat and time than a dirty one, and therefore lessens the chance of damage. In some cases, it may be necessary to apply a small amount of new solder to an existing joint to facilitate desoldering - it is important to keep temperatures as low as possible and make the desoldering activity as short as possible to prevent PCB damage.

None of this will guarantee that you will not damage the board, but such damage is a lot less likely. Should a pad be lifted while desoldering, do not rely on that pad when the replacement component is resoldered. Instead, fold the component lead down flat against the PCB, directly along the connected track (5mm or so is recommended). Carefully solder the lead and track to make a solid connection - the pad is now almost irrelevant, but the component is properly connected.

In the case where more than one track runs from a pad, make sure that the remaining section of track is intact - use a multimeter!

Hairline cracks

Cracked copper traces (hairline cracks) caused by PCB damage are quite common, and are very hard (sometimes almost impossible) to see. It is far easier to make a quick measurement whilst doing the repair than to try to find the fault later.

The most common reason for PCB failures (lifted pads and tracks, hairline cracks, etc.) is excessive heat and/or force. The copper is held onto the PCB laminate by an adhesive, and there are very few adhesives that can withstand soldering temperatures for any length of time. A damaged PCB can always be repaired using tinned copper wire soldered along the faulty section of track and to the component lead(s), but future repairs will be more difficult, and the result may be very untidy.

10.0 Testing

Once repairs are done, or at least thought to be done, the next step is to test the amp (or preamp, etc) to make sure that everything is now working properly. There is a natural tendency to want to hear it working immediately, but it is important that you resist the temptation, lest you be showered with fire, brimstone, and capacitor guts.

Make sure that you have everything needed to hand - multimeters, signal source, dummy load (for power amps), etc. In all cases, the first power-up should be done with a low voltage, current limited supply (if available), or use safety resistors in series with the supply leads. The idea is that since you have most likely replaced possibly expensive components, it is preferable that they don't blow up because there is a secondary fault that you did not find the first time.

Such secondary faults are very common, and their destructive capabilities should never be underestimated. By using limiting resistors and a Variac (or a 100W light globe wired in series with the mains cable) the energy available is greatly reduced, and the chances of (further) component damage are minimised.

For the remainder of this section, I have assumed a power amplifier, since they cause far more grief than anything else. Before testing, make sure that all power transistor hold-down screws are very firm - do not overtighten, but screws must exert sufficient pressure to ensure that a small amount of heatsink compound is forced out around all transistors.

10.1 Power Up

Initially, do not connect your dummy load. Use alligator clip leads from your multimeter, and connect to the supply rails (positive and negative). Any quiescent current setting pot

should be set for minimum current (see the original project article). If you use a Variac, advance the voltage slowly and observe the voltage. Check the safety resistors for heat - they should remain cool! Any heat (from anything) as the voltage is advanced is an indication that something is amiss. Should you find heat do not advance the voltage further. Measure the voltages before and after the safety resistors to determine the nature of the fault.

If you use the series light globe method, then initially, turn the power on for a brief period - the lamp should flash brightly, then settle to a dull glow. When turned off again, the voltage should decay relatively slowly (typically a few seconds) - if the lamp remains at full brightness and the voltage collapses quickly, then there is a fault. The safety resistors will probably be warm to hot.

If everything is ok, leave the amp on for a few minutes, and check *everything* for heat. Some components are expected to get warm, but *anything* that causes you to exclaim "RudeWord!" is an indicator that something is amiss. Refer back to the fault finding sections of this page, and locate the fault. Any remaining fault calls an immediate halt to the test process.

10.2 Voltage Checks

Having successfully reached this part, you should now verify that all voltages are normal. The voltage across each safety resistor should be only a few millivolts to up to a volt or so, depending on the amp design. The output voltage should be close to zero - as a rule of thumb, any offset in excess of 100mV or so is excessive, but be aware that some amplifiers may not be able to better this while the safety resistors are still installed (uncommon but possible).

If the output voltage is at zero volts, and the supply rails are close to normal voltage, there is a good chance that the amp will work. I don't suggest that you start shouting "Hurrah!" just yet though (sorry ;-)

10.3 First Power Check

Making sure that your dummy load is set to 8 ohms (or more), connect it to the output (do not disconnect the safety resistors!). Connect your signal source (output set to zero!), and then slowly advance the level while monitoring the voltage at the amplifier side. Advance the level until there is about 5V across each safety resistor. The voltages should be the same across each resistor. Voltage across one but not the other indicates an open circuit power stage. You may also use a speaker, with a series resistor of around 22 to 47 ohms. If the sound is heavily distorted, then there is something wrong. Advance the level a little more to determine if you are hearing crossover or "rectification"

distortion. If distortion lessens as the signal level increases, then it is crossover distortion, and you have nothing to worry about (the quiescent current is set to minimum, remember).

10.4 Final tests

Having verified basic functionality, at least to the best of your ability, just do a quick re-cap of the tests ...

- Supply voltage at amp terminals (with safety resistors) is within 1V of the voltage across the filter caps
- Nothing gets hot - including safety resistors (slightly warm only)
- Output voltage is within 100mV or so of zero volts (preferably less than 50mV)
- A basic power test shows that both sides of the amp draw current, and a reasonable signal is heard on a speaker.

You are now ready to test the amp at full power (but still use the Variac or light bulb lead for safety). Disconnect the safety resistors, and replace any fuses you removed with the correct value. You may leave the speaker connected with the series resistor - just in case. Turn on the power (or wind up the variac) - all voltages should come up to normal, but they may be a little low if you use the "lamp lead". Check carefully for anything that may be hot (or getting hot).

If all is well, adjust the quiescent current to about 1/2 the recommended value, inject a signal, and verify that the amp sounds clean. Do not try to get full power if you still have the lamp lead. Leave everything for at least 10 minutes, checking that bias (quiescent) current remains stable, and that no components are too hot. Note that some devices do run quite warm in many amps, but you should be able to hold all small/ medium transistors without being burnt.

When you are satisfied that everything is working properly, switch off, remove the Variac or lamp lead, and switch on. Double check the supply and output voltages, and apply a signal - if you still have the series resistor in series with your speaker, you can advance the input level until the sound becomes distorted (clipping).

Check all temperatures (again! I know this is tedious, but it is worth the effort). Set quiescent current to the recommended value, using the method suggested in the project article (this can vary considerably). Leave the amp to stabilise, monitoring temperatures - transistors, resistors, heatsink. After the temperature has settled (typically about 15 minutes), verify the quiescent current, and adjust as needed.

Note: If the heatsink temperature continues to increase (and so does quiescent current), you have a problem! Switch off immediately, and reduce the current setting. Do not apply power until the heatsink cools (a fan will make this a lot faster). The amplifier may have a thermal stability problem - verify that the bias transistor (if used) is mounted according to instructions. Thermal runaway (as this problem is known) is usually the result of insufficient thermal feedback. Consult the project designer for information on how to solve the problem.

Never use an amplifier while this problem persists - it will overheat and fail.

11.0 Preamplifier Tests

This is a very short section, since the range of tests on preamps is minimal - especially for opamp designs. As always, an oscilloscope is very useful, but not everyone has the luxury of ownership. As a result, other methods need to be found to enable you to track the signal through the preamp, until the point where it disappears is found. This narrows down the "search area", and makes it a lot easier to find the problem.

First, check that the opamp outputs are at (or near) zero volts, and that you have +/-15V (or whatever your supply voltage is supposed to be) at the supply pins. This is fairly obvious, but when you are desperate, it is often overlooked. Also, make sure that the 0V (GND) connection from the power supply is connected - this is a common error.

Then, apply a signal from a PC sound card, tuner or CD player, and use an old active PC speaker as an audio probe (it *must* be an active speaker, with its own internal amplifier). That way, you can check the signal level at the input, output of 1st opamp, etc. You need to find where the signal stops, and you can do that easily by just following the signal path, and monitoring the signal. Check for level, distortion, or anything else that may be amiss.

Remember that the PC-speaker input lead's shield must be connected to the ground (common) of the preamp, and the centre wire is used to probe the circuit under test. These speakers are low power, but I recommend that the level be adjusted so that you have a clean signal in the speaker. Never connect the PC-speaker input to the supply voltage(s), as you may damage the internal amplifier. A 1k resistor in series with the signal lead is a good idea.

In nearly all cases, there is a mistake somewhere, as all PCBs have been fully verified and definitely work if wired properly. Faulty opamps are possible, but very rare. Because

of the ground-plane in some designs, there are a lot of very closely spaced tracks, and solder bridges are easy to make.

This method can be used on almost any opamp based circuit, be it preamp, equaliser, or whatever - however it is limited to audio frequency circuits for obvious reasons.

12.0 Terminology

There are all sorts of strange noises that electronics can make through loudspeakers, and only a few are intentional. Clicks, pops, farts, hum, buzz, hiss, thumps - all have meaning, but the meanings are sometimes confused (such as "hum" and "buzz"). I can't (and won't) attempt to cover them all, but will go over some of the more important (and common) ones.

Hiss: The sound of an FM tuner that is not on a station, or the sound of air between your teeth. The "sss" sound in the letter ess. Those are all hiss - anything that does not sound like one of those is not hiss. Some hiss is inevitable - all components make noise - even cables, and the use of low noise components (especially opamps and resistors) will reduce, but never eliminate, hiss.

Hum: As described above - a smooth, 50/60Hz low frequency, with no harshness or high frequency energy at all.

Buzz: Any frequency, but usually 100 or 120Hz depending on your mains frequency. Buzz has a harsh tonality, and is typical of rectifier noise, TV frame synchronisation noise (run a signal lead behind the telly to hear that one), and general mains noise. Probably the hardest to fix.

Crackle: Sometimes crackles can be caused by a faulty component (transistor, opamp, etc.), and dry solder joints are also a good source of crackles, pops, farts, and other noises of similar ilk. Occasionally (especially in valve equipment), crackles may be caused by the valve itself, or a faulty capacitor. Usually easy enough to find, except crackles and pops (etc.) almost never occur when test equipment is nearby - can be very sneaky. Try lightly tapping the board and components with an insulated probe of some kind (a plastic screwdriver handle works well), as many such noises are vibration sensitive.

Distortion: Extremely variable. A couple of the more common types are ...

- Crossover (or "notch") distortion sounds similar to tearing paper, and is only present when there is a signal. It gets worse as the level is *reduced*. Severe crossover distortion will allow an amplifier to stop reproducing anything other than signal peaks at low volumes.
- There are many other forms of distortion as well - "rectification" distortion can occur if only half the signal is reproduced, and is typical of open circuit power transistors, faulty drivers or current sources. It is sometimes difficult to hear some forms of distortion (such as that caused by an amplifier oscillating).

An oscilloscope is essential for detecting oscillation induced distortion - there is no other way to see exactly what is happening to the signal.

Thump: A low frequency noise, often as power amps (in particular) are turned on or off. This is usually a design issue, and most thumps are not of concern - unless you use an amp that thumps to drive tweeters in a biamp or triamp setup! You will almost always need a speaker delay relay to get rid of thumps, since the only real alternative is to redesign the amp. Some opamp circuits also thump (or Crack!) as power is applied or removed. High speed opamps in the P09 electronic crossover is a prime example. Again, relays and a delay circuit are the best solution.

Chirps: Some power amps (notably some otherwise excellent power opamps) emit a "chirp" or bird-like noise as power is removed. Like thumps and cracks, this is a design issue that is very difficult to prevent. If it really annoys you, then a loudspeaker mute relay is the only real solution.

13.0 Conclusions

As stated right at the beginning, there are just too many possibilities to try to cover them all. Nonetheless, I hope that this article helps you to debug your project, and allows you to concentrate on listening to music. I doubt that there are many things more disheartening than a shelved project that you just could not get to work properly.

As I find out more of the problems that people face, I will add to this article, and will try to ensure that it is up to date with the latest and greatest faults. It is probable that the vast majority have been covered here, but there is always a new one, or one I didn't think of at the time of writing.

Should you go through the steps outlined for your particular problem, and still not find the solution, then let me know - I will help you solve it if I possibly can.

Note that this *only* applies to ESP projects - if you found a circuit or project elsewhere, then you must discuss it with the designer or publisher - **not** with me. I not only do not have the time to discuss other people's designs, but have a distinct lack of inclination to do so.

Of all the issues I have discussed here, the most common problem of all is failure to follow the instructions, and/or failure to place all components where they are supposed to be. I would estimate that well over 90% of all faults are caused by incorrect component placement - "I have checked and double checked" is a phrase I have seen in e-mails countless times, but the problem still turns out to be wrongly installed parts. This *must* tell you something ;-)

This is not to be taken as an insult to your intelligence by any means - everyone does it. When one of my prototypes fails to work as expected, there are two possibilities ...

- I made a mistake in the board layout
- I installed something incorrectly

So far, incorrect components have accounted for maybe 60% or more of original failures - in some cases because I goofed in the design phase, but mostly I just did what everyone else does - put something in the wrong place.

If it helps at all, just remember ... "*The person who makes no mistakes, makes nothing at all !*"

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 24 Apr 2003

The Audio Pages

Elliott Sound Products **Amplifier Bridging**

Copyright © 2001 - Rod Elliott (ESP)
Page Created 20 Oct 2002

Articles Index

Main Index

Contents

- Contents
- Introduction
- Description
  - How Does it Work?
- Conclusion
- Copyright

Introduction

The so-called "Trimode" system used by most car amps seems mysterious, as there is no speaker switching, and commonly only a simple stereo/ Mono switch. Even without the switch being operated, a speaker connected between the +ve (Red) terminal of one amp (typically the left channel) and the -ve (black) terminal of the other will still work with a typical stereo signal.

There is nothing magical about this, and the basic concept is used in almost every car amp known. What follows is a simplification, but describes the process and how it works.

Decription

A schematic of a trimode amp is shown below, in Figure 1. As you can see, there are two

completely normal power amplifiers, and a simple inverter circuit is used in front of the Right channel amp. The "trick" is that the output terminals for this inverted channel are reversed, so the +ve Red terminal is actually connected to ground (chassis), and the -ve Black terminal connects to the amp's output. For any normal stereo signal, this maintains the correct polarity of the system, and the speakers will not be out of phase.

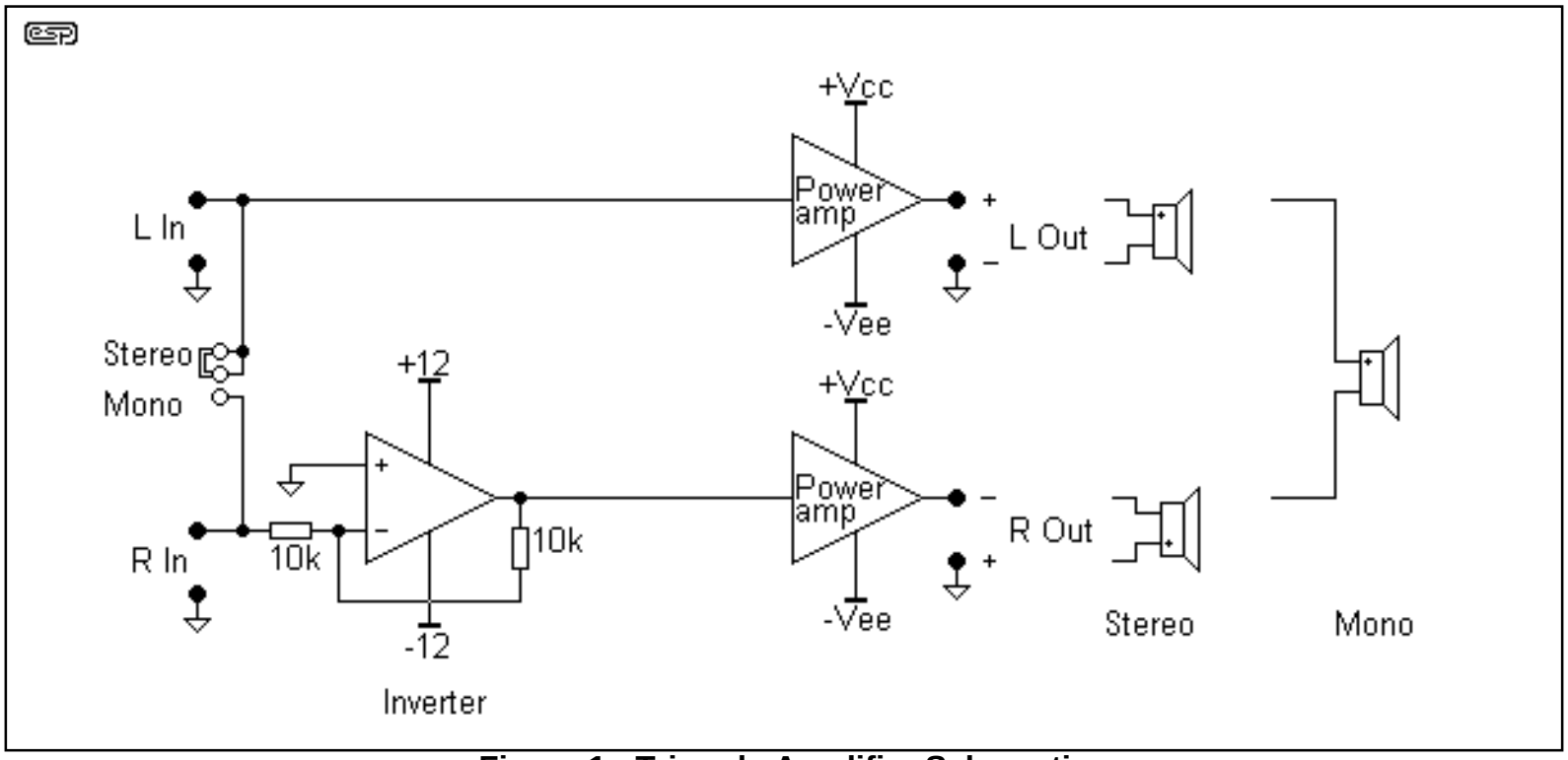


Figure 1 - Trimode Amplifier Schematic

The opamp is usually nothing special, and virtually any standard opamp (such as a TL071) can be used. If another opamp is used in the front end, then the device may be a dual opamp - I have deliberately not added pin numbers, since you will need to work out for yourself what you need here.

Since this is not intended as a design as such, if you want to use this method, you will have to work out the details for yourself. Do **not** send me e-mails asking for a complete design! There is enough detail here for anyone who is capable of wiring a simple circuit, and who understands the basics of amplifiers in general terms. The +/-12V for the opamp is commonly taken from the main switching supply, either using a separate small winding, or by using regulators off the main supply.

Again, you will have to work this out for yourself - I will not do it for you.

How Does it Work?

A bridge tied load (BTL) amplifier applies a normal signal to one terminal of the speaker, and an inverted signal to the other. If a single amp is capable of producing 20V RMS across the speaker, this equates to $P = V^2 / R$, so in this case, 20^2 / 4 = 100W.

When connected in BTL, the same speaker "sees" 20V at one terminal, and an inverted 20V signal on the other - a total of 40V RMS (I shall leave the proof of this to the reader :-) Using the same formula, 40^2 / 4 = 400W - **four** times the power. But ... each amp now sees only half the load impedance (think of an imaginary centre tap in the voice coil, connected to ground). The amplifier must be stable into 2 ohms, or this method will not work. Of course, you can use an 8 ohm speaker and still get 200W if the amplifier cannot drive 2 ohms safely.

Conclusion

The method described here is in almost universal use for car amplifiers, is simple and works well. Strangely enough, it is rarely used in (home) hi-fi amplifiers, possibly because it would be considered "unacceptable" by audiophiles (for whatever reason).

As stated above, please don't e-mail me for further information on the construction of high power car amps. There is a project article for a high powered switchmode converter, and that can be used with almost all amplifiers shown on these pages - however, none of the amps I have presented will drive 2 ohms safely, and they are not recommended for car audio use. Most car amps are very rugged, but are typically of "sub-hifi" performance, often having relatively high levels of crossover distortion (especially audible at low impedance and low volumes).

A power amplifier for car audio use may be offered at some later time, but at the present there is nothing available on these pages that I consider suitable. Several people have enquired about using LM3886 power op-amps, and although these will work, they really need to be connected in parallel. There is an excellent Application Note (AN-1192) on how best to do this available on the National Semiconductor site (the document is a PDF file - approx 529k in size). There is also a simplified method of achieving the same result as shown above, but bear in mind that input impedance will be low (and there are possible stability concerns as well).

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 20 Oct 2002

The Audio Pages

Elliott Sound Products **A Beginners Guide to Audio Bulletin Boards**

Contributed by Geoff Moss
(de facto editor of the Audio Pages)

Articles Index

Main Index

A useful source of information for those starting out in audio electronics is the Audio Bulletin Board or Forum. There are a number of these on the Web where you can exchange views with and seek help from other people with similar interests and possibly greater experience. However, before venturing forth into a forum, it helps to know the customary procedures and protocol.

The following notes will give you a basic understanding of what to expect and how to behave when visiting these sites.

- When you register with a forum or submit a posting (message), use a silly name; it helps to preserve your anonymity and adds an air of gravitas to your postings.
- Ignore all spelling, punctuation and grammar; not only is it time-consuming to get these right, but somebody might actually understand what you are trying to say.
- Use as many abbreviations and acronyms as possible; if there isn't a standard abbreviation, make one up (but please don't define it when it is first used).
- When you start a new thread (topic), give it an ambiguous title; that way, more people will access your thread to see what it is about.
- When adding to an existing thread, try not to respond to the original question; it is

preferable to go off at a tangent and expound upon your latest obsession.

- Try to keep adding to postings in an existing thread, even if it means repeating that which has been said before; this will ensure that the thread has chance of becoming one of the longest on the Forum and will receive the appropriate accolade.
- Do not treat any comments on your postings as being constructive; respond vehemently, as if you have suffered a great personal affront.
- Never, *ever* admit you are wrong or apologise if you are proved to be wrong.
- Don't be hidebound by facts and accuracy; these are of little importance to the overall debate.
- The views of certain correspondents carry more weight than others; this is usually in direct proportion to the number and length of their postings (it has nothing whatsoever to do with their competence, knowledge or ability).
- Feel free to comment on the quality of an amplifier even if you have never built or heard it; your views and observations are still relevant.
- Simulations are a good basis for comparing the quality of different circuit topologies; you don't need to build or listen to anything to determine the relative merits.
- An amplifier should only be judged subjectively; objective measurements don't mean anything and are a complete waste of time.
- An amplifier should only be judged by objective measurement; you do not need to listen to it to determine its abilities.
- All amplifier designs originating from the USA are better than any 'foreign' alternatives.
- All amplifiers built in the USA are superior to any 'foreign' alternatives.
- MOSFETs are always better than BJTs (because Nelson Pass uses them); BJT designs are archaic and of historic interest only.
- Nelson Pass designed amps are always best; Nelson Pass is always right!
- Try to ensure that you always have the last word in any debate, even if you have

no more constructive comments to add.

Though I have referred to amplifiers in the above notes, the comments equally relate to all other items in the audio chain.

On a more serious note: I started out by suggesting that an Audio Bulletin Board or Forum is a 'useful' source of information for the beginner. This is not necessarily the case.

There is a lot of incorrect information being expounded in these fora (forums) by people who do not know what they are talking about (though I am sure that they genuinely believe what they are saying). All information obtained from these questionable sources should be treated with caution and thoroughly checked against an authoritative source before being used in a project. Either that, or restrict your reading and research to reliable sources such as the ESP Audio Pages.

No-one gets out of here alive ... (well, maybe not that bad, but I have to add my piece). From Rod ...

The above is intended as humour, and I hope that readers will see the funny side to Geoff's comments. When next you visit one of the various bulletin boards, read a little more closely than usual, and you will see that a great many respondents follow these instructions to the letter!

Humour aside, the topic has a serious side. Geoff added in an e-mail ...

> *"I had two serious intentions, one to warn the uninitiated against believing everything they read on the Web and the other to try to shame the correspondents into better behaviour (probably a waste of time as this is likely to prove impossible). It is a pity that all fora seem to degenerate to the same abysmal level. A good one could be both interesting and informative and of great benefit to the DIY community."*

There is a vast amount of "disinformation" offered by the chronic bulletin board addicts - in some cases this is simply because the respondent(s) don't know enough to know when not to say anything, and in other cases it is because they have their own (usually hidden) agenda. For example, one can find posts from "high end" cable dealers who will blithely say that anyone who can't tell the difference between "cords" has tin ears, but do they say anywhere in their posts that they make a living from reeling in the suckers? Noooooo.

The list of charlatans is endless and depressing, and the poor audiophiles lap it up as if it were nectar from the Gods.

Every so often I visit some of the 'boards, these days mainly to see what they are saying about my site and its contents. There was the great "flame out" of '99, and I think that came close to the record for the longest thread - all over mains leads! During the entire thread, barely a civil word was uttered, as the respondents became more and more determined to have it go their way. This (of course) never

happened, and eventually *everyone* gave up in disgust.

Heaven help anyone who disagrees with one of the self appointed BB experts, for surely will the wrath of the (anonymous) snake oil dealer descend upon s/he who dares to offer a word of dissent. If you even try to be sensible or factual, you are dead meat, mate, since everyone there knows that to detect by ear is divine, but to measure is heresy - punishable by death, defenestration and/or disembowelment (but not necessarily in that order) if they had their way.

Articles Index

Main Index

This article was written by Geoff Moss (with parts added by Rod Elliott) and is Copyright © 2001, all rights reserved. Reproduction, storage or republication by any means whatsoever whether electronic, mechanical, or any combination thereof is strictly prohibited either in whole or in part without the express written permission of the author, with the sole exception that readers may print a copy of the article for personal use.

Article created 08 Apr 2001

Elliott Sound Products **Audio Designs With Opamps**

Copyright (c) 2000 - Rod Elliott (ESP)

Articles Index

Main Index

Contents

- Introduction
- Essential Formulae
- Basic Rules Of Opamps
- Essential Opamp Information
  - Configurations
  - Applying Power
  - The Ideal Opamp
- Basic Opamp Circuits
  - Non-Inverting Amplifier
  - Inverting Amplifier
  - Inverting And Non-Inverting Buffers
- Interesting Variations Of Basic Circuits
  - High Impedance Amplifiers
  - Simulated Inductor
  - All-Pass Filter
  - Phase Shift Oscillator
  - Schmitt Trigger Oscillator
- Part 2 - Pure Audio Circuits
- Part 3 - More Filters, Comparators, Etc.
- References / Suggested Reading
- Copyright Notice

Introduction

There are few audio designs today that do not use operational amplifiers (opamps). Over the years, the poor opamp has been much maligned, with mainly spurious claims about 'audibility', distortion, and other so-called defects.

Although some of the more basic opamps do have limitations which make them less than desirable in some cases, most of the new breed are unsurpassed for linearity, with total harmonic distortion figures as low as 0.00008%. Even the basic types still have their uses in control circuits and other non-demanding applications.

The operational amplifier was first used in the 1940's, as the basis of analogue computers. Since integrated circuits were unknown at the time (this was before the invention of the transistor), the earliest versions were made using valves. The basic concept is to have an amplifier with differential inputs, whose operation is controlled only by the external feedback components. By rearrangement of the feedback circuit, different "operations" could be performed. Typically, these early opamps could add and subtract, and these are essential functions to this day (even in audio). With the advent of the IC and mass production techniques, the opamp became very popular and remains so - with considerable justification.

This article will concentrate mainly on audio applications, but there are some configurations that are just so wonderful that I cannot resist the temptation to include them. For the most part, any of the configurations shown can use the simplest (and cheapest) opamp you can get (especially for testing), unless extremely wide bandwidth or low noise is a prime consideration. For any of the test circuits this is not an issue.

I also suggest that you build up the Opamp Design and Test Board (Project 41), which is ideal for the experimenter. Most of the circuits shown can be built using this test board, and will function perfectly, although there will be limitations as to bandwidth and noise because of the LM1458 dual opamps recommended for the project.

Essential Formulae

To understand this article, you need to know Ohm's law and it's derivatives. Ohm's law is fundamental to electronics, and with little more it is possible to derive most of the other resistance based formulae. Ohm's law states that a potential of 1 Volt through a resistance of 1 Ohm will cause a current of 1 Ampere to flow. This is expressed as:

R = V / I ... where R is resistance, V is voltage and I is current, or ...
I = V / R ... or ...
V = I * R

Later on, we will also use the formulae for inductive reactance and capacitive reactance, as well as calculating frequency response and some filter design. These will be presented as needed.

You will also see references to 'an instantaneous level of "x" volts AC'. At any point in time, an AC voltage has an instantaneous voltage - this is the voltage that is present at that moment, and for

analysis can be treated as DC. This is valid only when we consider this 'DC' level as a transient thing, since many of the circuits do not operate down to DC at all (many others do, but this is beside the point :-)

Basic Rules of Opamps

Many years ago I used to teach electronics, and I devised what I called the 'Basic Rules of Opamps' for the purposes of explanation. There are two Rules, and although real life is never like theory (I could fill the page with suitable examples, but shall refrain), they describe the operation of all opamp circuits very accurately:

1. An opamp will attempt to make both inputs exactly the same voltage (via the feedback path)
2. If it cannot do so, the output will assume the polarity of the most positive input

Needless to say, this requires some explanation. So let's look at Rule #1 ("No Poo....", oops, sorry, wrong Monty Python sketch).

An opamp will attempt to make both inputs exactly the same voltage
When an opamp is operated in its linear mode (which is most of the time for audio), the feedback circuit will cause a voltage to appear on the inverting input (-in) that is almost exactly equal to that present on the non-inverting input (+in). Any change of voltage on the +in terminal is reflected by a change in the output that causes more or less current to flow in the feedback circuit to maintain equilibrium.

If this is unclear to you, see the further explanations below - but remember the 1st Rule !

If Rule #1 cannot be satisfied, the output will assume the polarity of the most positive input
There are many circuits that use opamps in non-linear mode, and this can also happen if the output cannot swing its voltage fast enough. In these cases, should the +in terminal be the most positive, then the output will (or attempt to) swing positive. If the -in terminal is more positive, the output will swing negative.

There is almost no opamp circuit that you cannot understand once these Rules are firmly established in your thinking. Even circuits that use external transistors in strange ways will obey the Rules. An opamp that does not perform as above is being used outside of its normal operating parameters, and the results will be unpredictable and almost always unsatisfactory.

It is more often explained that an opamp reacts only to the difference between the two inputs, and not to their common voltage. The ability to ignore the common voltage is called the Common Mode Rejection Ratio (CMRR), and will be covered later in this article.

Some Essential Opamp Information

Before we cover the circuits themselves, we need to look at some of the parameters you will come across, how to apply power and bypass the supply rails and so on.

There are many parameters that you will see in data sheets, and these are covered in more detail a little later. There is no point doing it now, as the importance will be lost until you know more about the opamp itself.

Configurations

Opamps come in a variety of configurations, but the most common are:

Single - one opamp in an 8 pin DIL (Dual In Line) package
Dual - Two separate opamps, sharing only the power supply pins - commonly in an 8 pin DIL package
Quad - Four separate opamps, again sharing the power supply - most commonly in a 14 pin DIL package

Amazingly, nearly all opamps use the same pinouts, and these were established many years ago by the venerable uA741 for single opamps, and the likes of the LM1458 dual opamps set the stage for the others that followed. Many of the quads use the same pinouts as well, and this has enabled people to swap opamps for 'better' ones for a very long time.

However - Don't count on it! There are some variations, and although uncommon, they do exist. I shall not be concerned with any of the different devices - only the common pinout versions will be shown.

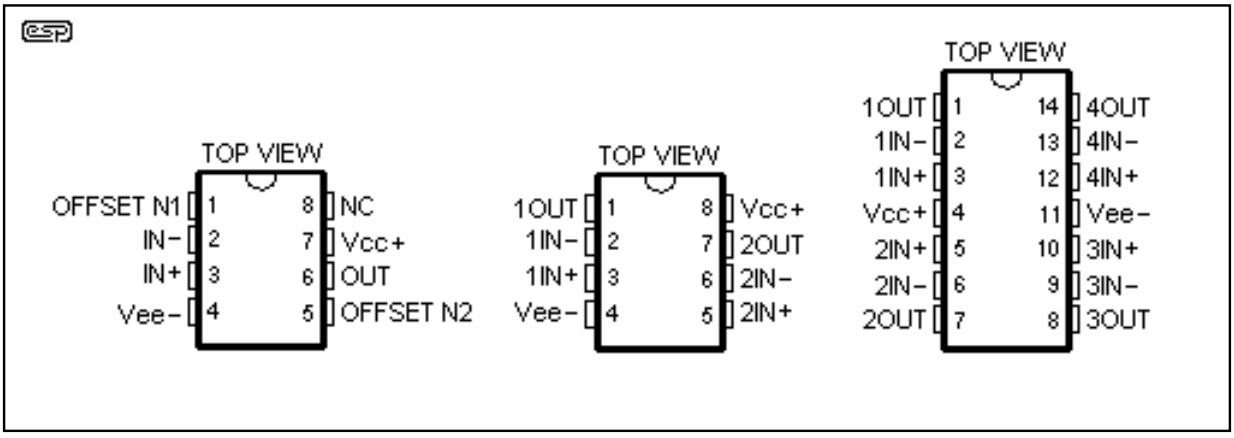


Figure 1 - Common Opamp Pinouts

Figure 1 shows the standard connections for single, dual and quad opamps, but be aware that the remaining pins on the common single devices can occasionally have uses other than those shown. Single opamps have additional connections available. These are most commonly:

- Offset Null - used to adjust the amplifier to ensure that the input transistors are perfectly balanced, so that with no input signal, there is zero volts output. This is important for DC amplifier circuits.
- Compensation - Some single opamps do not have inbuilt frequency compensation capacitors, and these are connected externally instead. This allows the designer more freedom, and the opamp's high frequency performance can be optimised for the design objective.

You may at times see these connections used in unconventional ways. This may be to obtain greater bandwidth than might normally be available, or perhaps just so the designer can show how clever s/he is. Either way, I shall not be delving into these aspects of the design process :-)

Applying Power

No (active) circuit works without power, so this has to be the first step. Most opamps will operate with a maximum of 36V between the supply terminals. As this is the absolute maximum, operation at a lower voltage is the general rule, and the most common is to use +/-12V to +/-15V to power the circuit. Some opamps are rated for higher voltages, and others for less, so consult the spec sheet from the manufacturer.

A dual supply is not required, but it does simplify the design, and is recommended for most applications. A dual supply has the advantage that all inputs and outputs are earth (ground) referenced. This can eliminate a great many capacitors from a complex design, and is the most common way to power most opamp circuits. Note that from a commercial perspective, elimination (or reduction) of capacitors is done for economic reasons rather than any great desire to simplify the signal path.

Since the pinouts are nearly always the same, Figure 1 will be applicable in most cases, but as I said earlier "Don't count on it!". When in doubt, get the specification sheet from the manufacturer. When not in doubt, get the specification sheet anyway.

Power Supply Rejection Ratio

The PSRR (Power Supply Rejection Ratio) of an opamp is a measure of the amount of power supply noise that finds its way into the output signal. Most specification sheets give the test conditions for this measurement, and this should be consulted if an unusual design is contemplated. Mostly it can be ignored.

Bypassing

Although most opamps have a very good PSRR, it is always recommended that the supply be bypassed with capacitors - especially with high speed opamps. Bypassing should always use capacitors with good high frequency performance, and ceramics are probably the best in this regard. It is common for designs to use electrolytic capacitors, themselves bypassed by low value (100nF) capacitors. This ensures that all noise sources are minimised, and helps to prevent

oscillation. When this occurs with a high speed (HS) opamp, it will commonly be in the MHz region, and is extremely hard to see on most oscilloscopes.

A sure sign is inexplicable distortion, that mysteriously disappears (or appears) when you touch the opamp, or a component in its immediate vicinity.

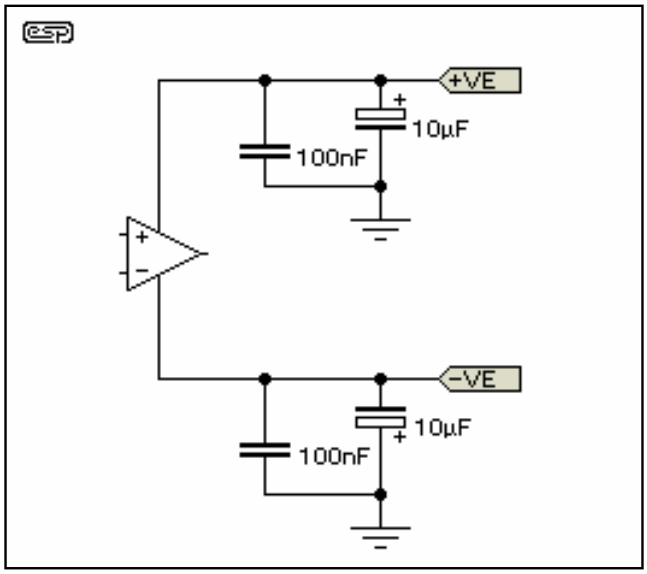


Figure 2 - Bypassing The Opamp Supplies

Even with HS opamps, electrolytic capacitors are not needed for each device, but the use of ceramic bypass caps between the supply pins of each device is highly recommended. Figure 2 shows the most common method of bypassing power supplies for opamp circuits, but there are others. In some cases, the supplies may not be bypassed to earth (ground), but just to each other. This is relatively uncommon.

Claims have been made that supply bypassing ruins the sound (rubbish), or that ceramic caps should never be used in audio, even for bypassing (more rubbish), and even that high value capacitors (> 1uF) slow down the sound (unmitigated drivel). These claims are made by frauds and charlatans, and should be completely ignored - they have no basis in fact whatsoever, and indeed, quite the reverse may be true in each case.

The Ideal Opamp

An ideal opamp has an infinitely high input impedance, and therefore needs no bias current. It is also capable of infinite gain without feedback, so there are no errors between the two inputs (i.e. Rules 1 & 2 will hold for all cases). The ideal opamp also has zero ohms output impedance, and is capable of supplying as much current as you will ever need. The ideal opamp does not exist :-(

Although it does not exist, the ideal opamp is the common model for nearly all opamp circuits, and few errors are encountered in practice as a result of designing for the ideal, and actually using a real (non-ideal) device. The tolerance of even the best resistors will ultimately limit the accuracy of

any opamp circuit at low frequencies (where gain is highest).

The primary practical limitations are:

- Input Impedance - Typically from one to several hundred Megohms. FET inputs are used for very high impedance inputs
- Gain - 100dB at up to a few hundred Hertz is common
- Common Mode Input Voltage - typically limited to the supply voltages, but may be up to 0.6V greater with some designs
- Bandwidth - opamps with a usable high frequency limit of 1MHz at unity gain are now common
- Output Current - most common opamps are limited to about 20mA of output current.

There are others, such as input offset voltage and current, but we shall not concern ourselves with these parameters just yet. Power opamps may be capable of up to 10A, but these are outside the scope of this section of the article.

The use of ideal opamps is assumed for much of the following, but all are designed to function properly with real world devices. In practice the difference between an ideal opamp and the real thing are so small as to be ignored, but with one major exception - bandwidth. This is the one area where most opamps show their limitations, but once properly understood, it is quite easy to maintain a more than adequate frequency response from even basic opamps.

The common mode input voltage can be important in some applications. Ideally, an opamp only reacts to the voltage difference between its inputs. Provided this does not change, in theory, the actual voltage between the two inputs and the common (zero volt line) may be anywhere within the specified range *with no change in the output voltage*. In other words, the inputs can assume any voltage between the negative and positive supplies, and there will be (almost) no change at the output.

With a real (as opposed to ideal) opamp, there will be some change, and this is specified as the common mode rejection ratio. An opamp with a CMRR of 100dB (not uncommon) will ensure that the change in output voltage is 100dB less than the change of input voltage (as applied to *both inputs simultaneously*. Any *difference* between the inputs is amplified normally. CMRR is affected by the open loop gain of the opamp, so is usually worse at high frequencies.

The Basic Opamp Circuits

The following collection shows the most common configurations for amplifiers. These are intended as linear amplifiers, in that they are essentially distortion free (within the capabilities of the opamp itself, of course).

As we progress, most of these original circuits will be seen over and over again, since they are the

very foundations of building an audio circuit using opamps.

In all cases, a dual power supply is assumed, and this is not shown on the circuits. This partly for clarity, since the additional circuitry makes the diagrams harder to understand, and partly because it is a convention not to show all the supply connections anyway. We all know they have to be there, so there is little point in showing the obvious over and over again. Likewise, bypass capacitors and other support components are not shown - only the basic opamp and its associated components.

You will also see reference to the "instantaneous value of the AC waveform". This is like a snapshot, and we simply freeze time while we analyse the operation of the circuit. At any point in an AC waveform, it can have only one value of voltage and current, *regardless* of the complexity of the signal source. A sinewave is no different from any other signal - provided its amplitude and frequency are within the capabilities of the opamp.

The Non-Inverting Amplifier

The most common of all configurations is the non-inverting amplifier. I will therefore use this as a starting point, because it is also the simplest to understand. Figure 3 shows a completely conventional non-inverting opamp voltage amplifier.

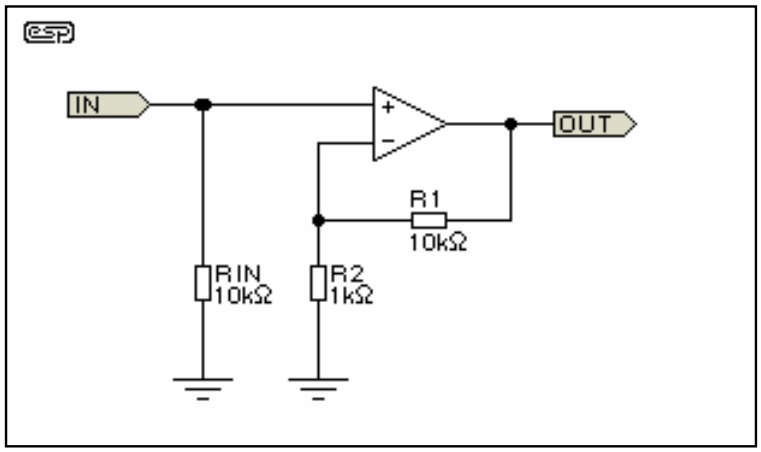


Figure 3 - Non-Inverting Opamp Amplifier

Rin is the input resistor, and is needed because an opamp needs a reference voltage at the input. In this case the reference voltage is the zero volt (earth) bus. Input impedance is equal to the value of Rin in parallel with the opamp input impedance. Generally the latter can be ignored because it is so high.

The gain (Av - Amplification, Voltage) is set by the ratio of R1 and R2, and is equal to:

$$Av = (R1 + R2) / R2$$

The gain of this stage cannot be less than unity, regardless of the resistor values used. As shown

in the diagram, the gain is 11 times, so a 100mV input will become a 1.1V output. To re-examine Rule #1, it is obvious that if 100mV (instantaneous AC or DC) appears at +in, the amplifier must have 1.1 volts at the output, since the voltage divider R1/R2 will ensure that 100mV also appears at -in. This is obtained from the simple voltage divider formula, which is strangely familiar ...

$$Vd = (R1 + R2) / R2$$

This will hold for any gain and any output within the capabilities of the power supply and the opamp's ability. A signal at 10MHz will not follow the rule, since the opamp will almost certainly be incapable of amplifying such a high frequency. An input voltage of 10V with a gain of 11 will also break the rule, since the opamp has only +/-15V supplies, and the output cannot exceed the supply voltage. Likewise, an 8 ohm load will break the rule, since the opamp cannot supply the current needed to drive such a load.

To see how the opamp behaves in these abnormal conditions, I suggest that the circuit be built, and run the tests if you have access to an oscilloscope. Examine the inputs as well as the output, since the inputs are by far the most interesting when the opamp is appearing to break the Rules.

Inverting Amplifier

Once, all amplifiers were inverting. A single valve or transistor stage (other than a cathode or emitter follower buffer stage) always inverts the signal, and this is how it must be (see Amplifier Basics - How Amps Work for more info).

With the advent of the opamp, all this changed, and the inverting amp is a very different beast from the simple discrete designs. The gain ratio is again set by a pair of resistors, but the +in terminal is earthed, either directly, or via a resistor. This configuration is also called a virtual earth stage, and is common in mixing consoles and many other signal processing circuits.

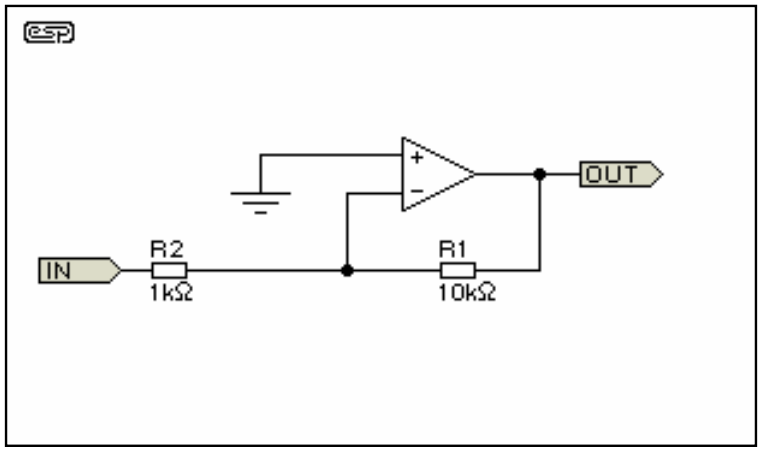


Figure 4 - Inverting Amplifier

Since +in is earthed and Rule #1 says that both inputs must be the same, -in will "appear" to be at earth potential as well (i.e. zero volts). Assume an input of 100mV DC. The output will be at -1V

DC, a gain of -10 (the minus indicates only that it is inverting, not that the circuit has "negative gain" which is actually a loss).

Input impedance is equal to the value of R2, and voltage gain is R1/R2, or 10 as shown. Note that this configuration is capable of negative gain (loss). If R2 is larger than R1 (say 20k), then the gain is equal to R1/R2 as before, so is now -0.5.

With the input at 100mV (again instantaneous AC or DC), the input current will be 100mV / 1000 (using Ohm's law), which is 100uA. The current through the feedback resistor must be exactly equal and opposite to ensure that zero volts is at the -in terminal (so we don't break Rule #1). As it happens, 1V / 10k equals 100uA, so the requirement is satisfied, since the output is negative.

As before with the non-inverting amp, the limitations of the opamp and its supply may cause Rule #1 to be broken, but the amp is now no longer operating in its linear mode, and Rule #2 will take over. Observation of the -in terminal will show a distorted waveform when the opamp can no longer operate in linear mode.

Inverting and Non-Inverting Buffers

A very common opamp application is the buffer stage, which (for the non-inverting configuration) can have an extraordinarily high input impedance, and a low output impedance. As with all opamp circuits, the output impedance may be very low (typically < 10 ohms), but the output current capability will not allow the circuit to drive such an impedance at more than the 20mA or so that is typical of most opamps. This would limit the output voltage (before clipping) to a maximum of +/- 160mV, or about 113mV RMS into 8 ohms. Distortion will be unacceptably high, and the end result is not worthy of further consideration.

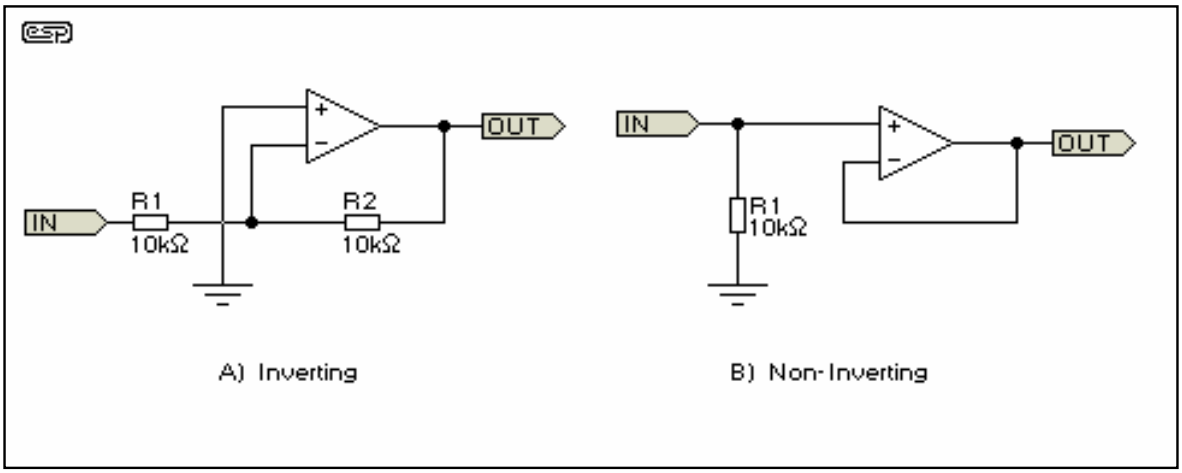


Figure 5 - A) Inverting and B) Non-Inverting Buffer

In many cases the non-inverting buffer can be replaced by an emitter follower, but performance is nowhere near as good. Input impedance is lower, output impedance is higher, and the gain is not quite unity. In addition there is more distortion and lower output drive capability.

The inverting buffer is more of a convenience than anything else, and is simply a normal inverting amplifier with unity gain. Input impedance is the same as R1, and very high values are not possible without excessive circuit noise.

Some Interesting Variations On Basic Circuits

It is now time to look at a few of the many variations on the basic circuits discussed above. It is not possible to cover all the different circuits that have been made using opamps, since there are so many that I could easily end up with the world's longest web page. I doubt that this would be appreciated by most of you :-)

I shall only cover the more common, or most interesting, as this will give a better appreciation of how versatile these building blocks really are. All of the circuits that follow will work - they are not theoretical, but real designs, and can all be made on the opamp test board.

High Impedance Amplifiers

The inverting buffer has been used in some very interesting ways. For example, a standard low cost TL071 opamp has an input bias current of about 65pA, and a claimed input resistance of 10^{12} ohms. To put this into perspective - assuming we have a way to supply the bias current without affecting input resistance - the input impedance could be as high as 10,000,000,000,000 ohms. That is 10Tohms (1 Tera-Ohm is 1000 Gig-Ohms). We will be completely unable to achieve this in practice, since the insulation resistance of a PCB is nowhere this figure, and the smallest amount of contamination will reduce the impedance dramatically.

In reality, we can easily expect to be able to get an input impedance of 100M ohms or more, but care is needed, since with high value resistors additional noise is produced. Since noise in a resistor is proportional to the voltage across the resistor and its resistance, it is easy to see how a simple circuit can become a real noise generator. Figure 6 shows the circuit and PCB layout for a very high impedance amplifier.

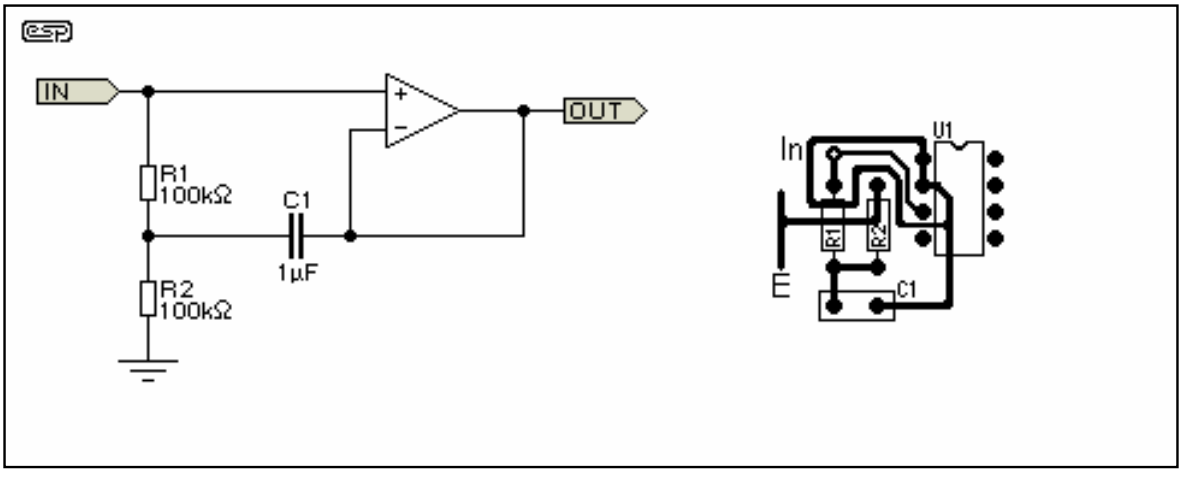


Figure 6 - High Impedance Amplifier

The bias resistor is 'bootstrapped' from the output, and this allows a lower resistance (for lower noise), while maintaining an extraordinarily high input impedance. A circuit such as this could be used for a capacitor microphone (for example), which will typically have such a small capacitance that any loading will reduce the low frequency performance to an unacceptable degree. The guard track can be seen encircling the input and the input end of R1. What on earth is a guard track? Read on ....

To prevent the resistance of the PCB from causing a problem, the input section may be 'guarded' with a section of track connected back to the output. Bootstrapping and guarding work in the same way. The guard track works by maintaining a voltage from a low impedance source around the input circuit that is the same voltage as the input. If they are the same voltage, no leakage current will flow. In reality it is not quite that simple.

Assume that the opamp has 100dB of gain at 1kHz (our test frequency). This equates to 100,000 - a little shy of infinity! Since the opamp has a finite gain, the 'unity gain' buffer will actually have a gain of 0.99999 - not 1 as we had assumed. This error reduces the ability of the opamp to bootstrap the circuit perfectly, so the 100k input resistance will only be effectively increased to 10G ohms.

But wait .... how does it increase the effective resistance at all? This is very simple. Assume an instantaneous AC voltage of 1 volt input to the amp. Normally, this would cause a current of 10uA into the 100k resistor of Figure 6. Because the bootstrapping action causes the voltage at the junction of R1 and R2 (Fig 6B) to be 0.99999V, there is actually only 1 - 0.99999 = 10uV across the resistor. The input current is now 10uV / 100k = 100pA (1 pico-amp is 10^{E-12}A). We can now calculate the equivalent resistance as R = 1V / 100pA = 10G ohms. This will fall at increasing frequencies as the opamp starts to run out of gain.

Oh yes, the term "bootstrap" comes from the unlikely picture of a man "lifting himself off the floor by his own bootstraps". As you might have guessed, the term is somewhat antiquated, but there has never been any move to change it (thank goodness). It is intended to show that the impossible can be done, but it is not really impossible, and is just a very clever example of lateral thinking.

The bootstrapped circuit cannot be used at DC, since it requires a capacitor for its operation. This is not as much of a limitation as may first be thought, since DC is quite inaudible

---

Simulated Inductor

This circuit has to be one of the all-time classics. Although it can also be done with a single transistor, the performance of the opamp version is so much better that the alternative is not really worth considering.

Inductors have always been a problem in electronics, as they are by nature relatively large, and being made from a coil of wire, tend to pick up mains hum as well as other noise in the electromagnetic spectrum. The idea of simulating an inductor using an opamp has been about for a long time. I have no idea who invented the circuit, but I am deeply impressed!

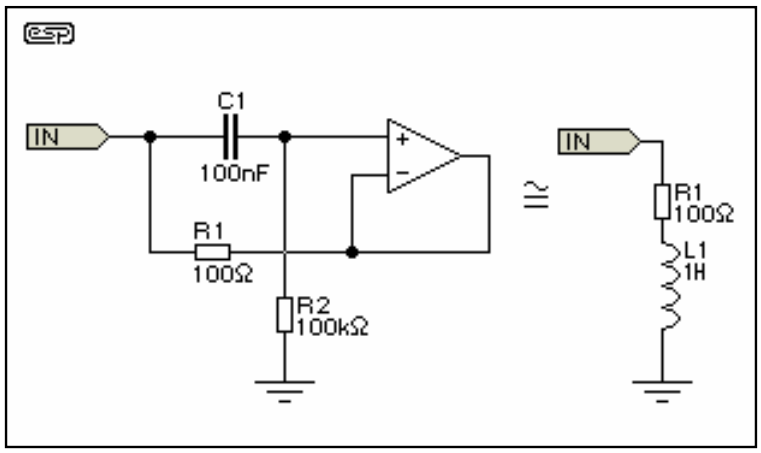


Figure 7 - Simulated Inductor

The circuit is so much smaller than a real inductor (at least for the larger values), and does not suffer from noise pickup. It does have a limited Q (quality factor), but it is rare that very high Q circuits are needed in audio, so this is not really a problem. It is also variable over a moderately wide range, something that is very difficult with the wire wound genuine article. So, how does it work?

The idea is very simple, but operation is less easy to understand. Essentially the circuit uses a capacitor, and 'reverses' its operation, thus making an 'inductor'. The essential character of an inductor is that it resists any change in its current, so if a DC voltage is applied to an inductance, the current will rise slowly, and the voltage will fall as the external resistance becomes more significant.

An inductor also passes low frequencies more readily than high frequencies - the opposite of a capacitor. An ideal (that word again) inductor has zero resistance, so will pass DC with no limitation, but will have an infinitely high impedance at infinite frequency. These limits are generally considered to be outside the audio range.

To understand how the circuit works, remember that the output of the opamp is (almost) exactly the same as the non-inverting input. Imagine a DC voltage of 1V is suddenly applied to the input, via resistor R1. The opamp will ignore the sudden load because the change is coupled directly to the input via C1. The opamp will represent a high impedance. Just as an inductor would do.

With the passage of time, C1 charges via R2, the voltage across R2 falls, so the opamp sees less and less of the input signal, and starts to draw current from the input via R1. This continues as the capacitor approaches full charge, and the opamp has close to zero input, so the output is also close to zero volts.

Eventually resistor R1 becomes the only limiting factor to current flow, and this appears as a series resistance within the inductor in the same way as the resistance of the wire in a real inductor behaves. This series resistance limits the available Q of both the simulated and real inductor, with the main difference being the magnitude - real inductors generally have much less resistance than the simulated variety.

Inductance is measured in Henrys, and for the simulated inductor is equal to ....

$$L = R1 * R2 * C1$$

A more accurate version of the formula (due to Siegfried Linkwitz) is shown below, but normally the error from the simple version will be very low with typical values - a ratio of 100:1 would normally be the lowest one would use, and this will have an error of only 1%. Component tolerance will have more effect, but for completeness, here is the accurate version ...

$$L = C1 * R1 * (R2 - R1)$$

so for the circuit shown is 1 Henry. This is a large inductance, and would be very expensive and bulky if made conventionally. The real inductance will have lower resistance and higher Q, but will need to have a large iron core to be able to withstand even a small amount of DC, and the inductance will change depending on how much DC is present. The simulated inductor is limited by the current capability of the opamp, so can handle up to +/-20mA with no change in performance.

There are some limitations to the simulated inductor.

- First (and most annoying) is that one end of the inductor is earthed. Although simulated inductors have been made that are floating (can be connected in any way you like), these are expensive and uncommon. Fortunately the standard version is quite suitable for many audio applications, so this is not too great a burden.
- The simulated inductor cannot be made with high Q, since the value of R1 cannot be made low enough to allow a Q of more than about 10. This is due to the limitations of the opamp - a minimum value of 100 ohms is usually specified for R1, although lower values are sometimes used. This represents a series resistance (equivalent to wire resistance in a real inductor.
- There is also a resistance effectively in parallel with the simulated inductor, equal to the value of R1 + R2. Although this can be measured, it is not generally a hindrance to practical circuit design.
- Although the simulated inductor acts in many ways like the real thing, it does not have the same energy storage, and cannot respond like a proper wound inductor. When the input voltage is suddenly removed from a real inductor, the collapse of the magnetic field causes a large voltage pulse of the opposite polarity - this does not happen properly with a simulated inductor, since there is no magnetic field involved. The simulated inductor will still try, but is limited to the voltage swing of the opamp, so the flyback pulse is limited to this value.

Figure 8 shows two simple LC filters. One is using a real inductor, and the lower circuit has a simulated inductor. They are both series resonant circuits, and are tuned to the same frequency (159Hz). The reference level (near the top of the graph) is 0dB, and neither circuit exhibits any appreciable loss outside the stop band.

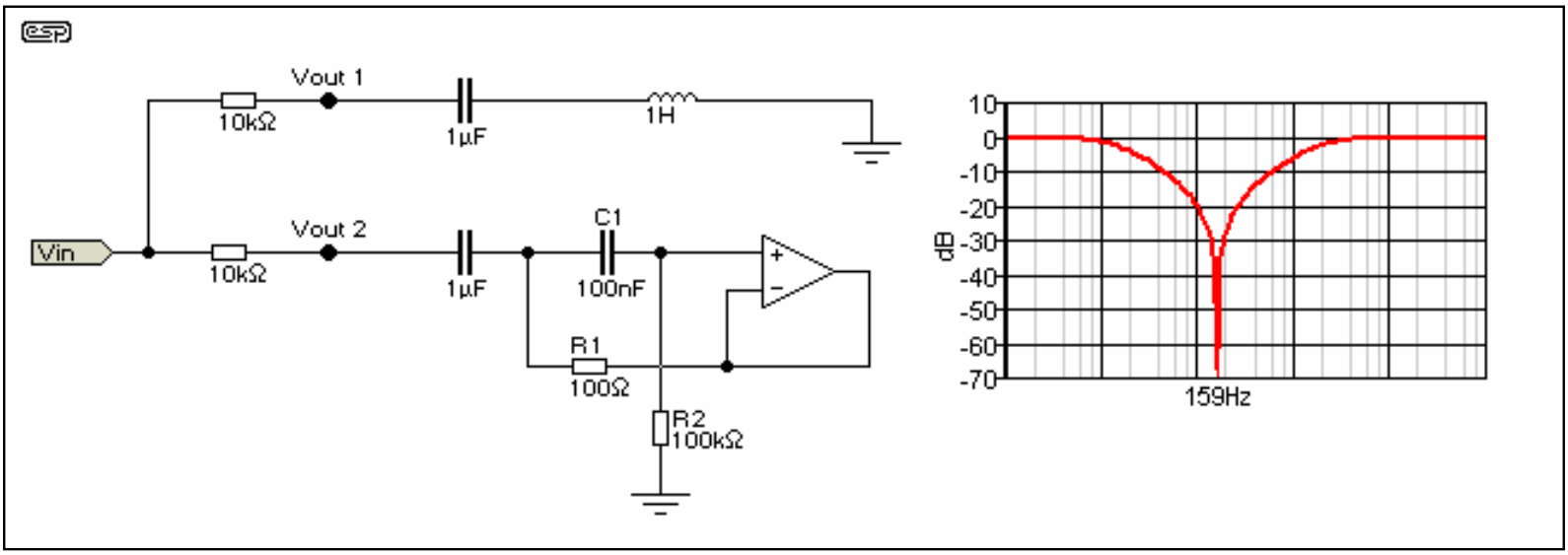


Figure 8 - LC Filters, Real And Simulated

The performance of both is almost identical as a simple filter as seen from the response plot, except the thing you don't see (mainly because I didn't include it) is that the simulated inductor has a slightly shallower notch, at about 37dB instead of 40dB. The frequency is calculated from ...

$$F = 1 / (2 * \pi * \sqrt{(LC)}) \text{ Hz}$$

A series resonant circuit has minimum impedance at resonance, and in the configuration shown will act as a notch filter, reducing the level at the resonant frequency. Because of the relatively low Q, the notch is not very sharp, but the simulated inductor is an important building block for equalisers and spectrum displays, and is quite common in audio.

Note that at the junction of Cin and the inductor, the voltage is higher than the input voltage. This is normal behaviour for a series resonant circuit.

All-Pass Filter

The all-pass filter is one of the strange ones. It passes all frequencies perfectly, with no attenuation at all within the capabilities of the opamp used. All it does is change the phase of the signal, and this circuit is used in everything from phase correction circuits for sub-woofers to guitar effect pedals. It is a versatile and useful building block, and the circuit is shown in Figure 9.

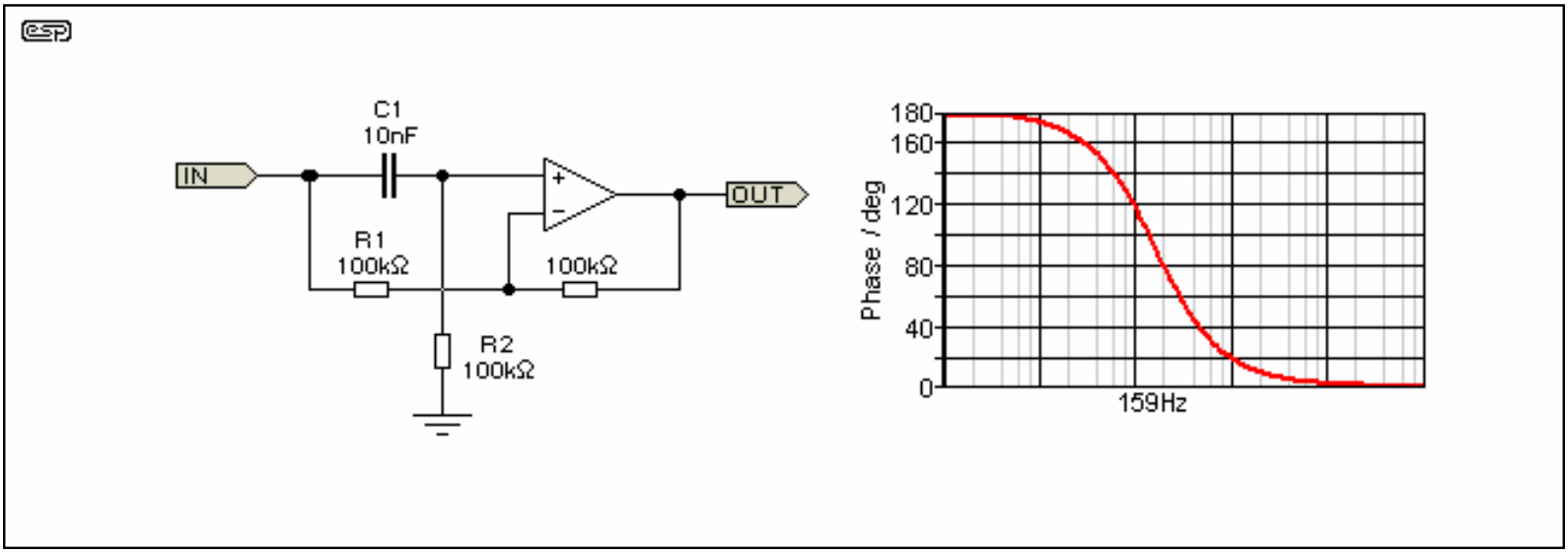


Figure 9 - All-Pass Filter

The circuit shown will have a 90 degree phase shift at 159Hz. At DC, phase shift is 180°, and at high frequencies it is 360° (note that 360° phase shift is *almost* the same as 0° - there is a subtle difference for transient signals, so the two can be considered identical only for steady state signal conditions). The shift of phase about the centre frequency is completely inaudible, but if a pot is substituted for R1, the phase can be swept back and forth. This is audible, and by cascading a number of these circuits "phaser" or vibrato (frequency modulation) effects pedals can be made. One of the latter is described in my projects pages.

The input signal is effectively applied to both opamp inputs, but there is always a small phase difference except at DC or infinite frequency. The value of C1 and R2 determine the frequency at which there is a 90° (or 270°) shift, and the frequency is determined with the formula ...

$$Fo = 1 / (2 * \pi * R2 * C1)$$ ... where Fo is the 90°phase shift frequency

A quick analysis will show how this works. Assume a DC input of 1V; at DC the cap has no effect, so the circuit operates just like an inverting buffer. The output is therefore -1V, so there is a 180° phase shift. At high frequencies, the reactance of C1 is negligible, and the full input is supplied to the opamp's +in terminal. Remembering Rule #1, the opamp output will be such that both inputs will have the same voltage, and in order to do this, the output must be equal to the input at high frequencies.

At intermediate frequencies, the combination of R1 and C1, along with R2 and R3 will ensure that the output amplitude remains constant, but the phase will change. The relative positions of C1 and R2 may be reversed, which will modify the characteristics of the circuit.

Phase Shift Oscillator

There are many things in life I do not understand, but one of the simpler ones is the phase shift oscillator implemented using an opamp. Don't get me wrong - the circuit I understand perfectly.

The bit I *don't* understand is how come (up until recently) I have never seen this circuit published - anywhere ???

In its heyday, this circuit was used almost anywhere a simple sine wave oscillator was needed, and I have seen it made with valves, transistors and even FETs. What I had *not* seen until I designed one was a phase shift oscillator using an opamp. As it transpires, although *I* had not seen this done, my trusty editor in the UK has.

He has seen it in a number of publications including John Linsley-Hood's "The Art of Linear Electronics". JLH also supplies the equation for frequency calculation ...

$$Fo = \sqrt{6} / (2 * \pi * C * R)$$

Loop gain must be 29.25 dB according to Mr Linsley-Hood, and lacking further information I must assume that the formula only applies if all resistors and capacitors are equal, and gain would be the minimum required for the circuit to oscillate.

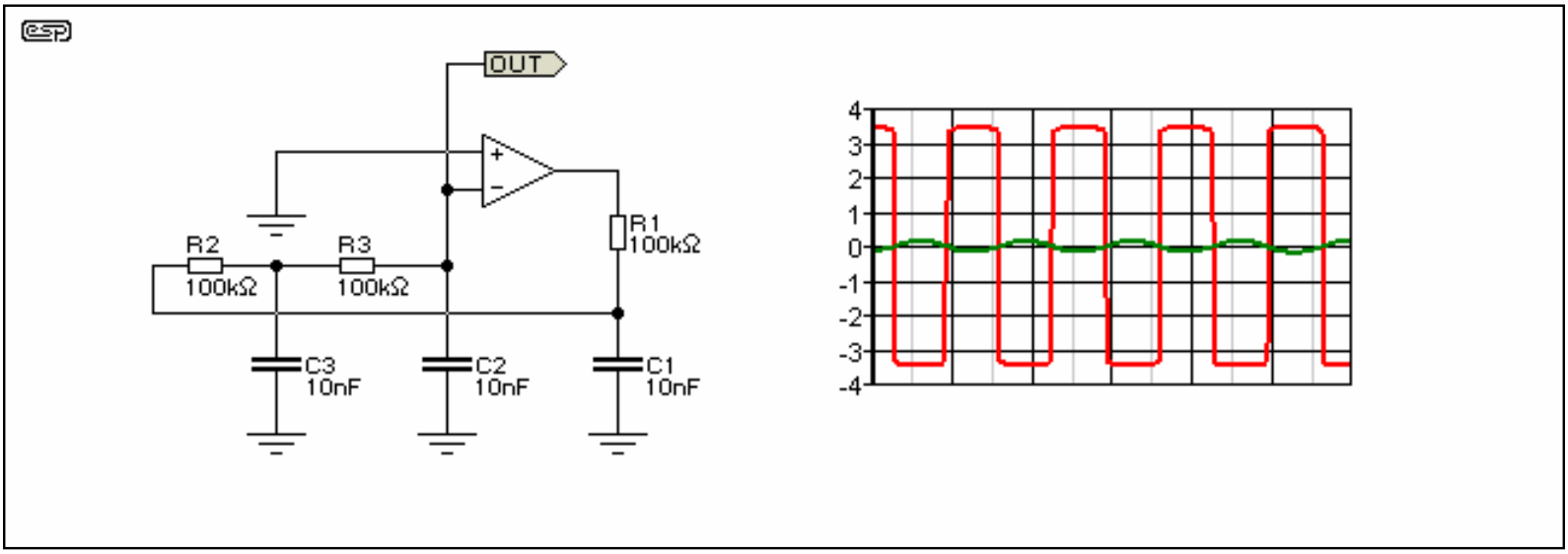


Figure 10 - Phase Shift Oscillator

The frequency stability of this circuit is quite good, but as with all phase shift oscillators the amplitude varies when the frequency is changed. Any resistor can be varied to change the frequency, and the use of a pot allows continuous variation over a 5:1 range (or more if you experiment with the component values).

This is a perfect example of an opamp being unable to obey Rule #1, and its operation is governed by Rule #2. The circuit is essentially unstable, and the opamp is always trying to play catch-up (without success, or the circuit would stop oscillating).

The frequency is a cow to determine if different values are used for R or C, and although I believe there *is* a formula, it is apparently a very tedious process (I've not seen it myself). The circuit shown above will run at about 420Hz, with a sinewave output of around 500mV (with +/-15V

supplies) - although JLH's formula seems to indicate that it should oscillate at 390Hz If you *really* want to know, you will have to build one. Changing the value of any resistor or capacitor will change both frequency and amplitude. The square wave at the output is at almost the full +/-15V supply voltage (limited only by the output circuit of the opamp).

The sinewave shown on the oscilloscope trace is obtained from the Vout terminal, and the square wave is obtained from the opamp's output (these are not to scale). The string of resistors and caps acts as a phase shift network, and oscillation takes place at that frequency where there is an exact 180 degree shift, converting negative feedback into positive feedback. The circuit is stable at DC, since it has negative feedback through the string of resistors.

Let's have a look at how it works. Remember Rule #2? Now have a look at the signal at the inverting input. As you can see, the output takes the polarity of the most positive input, so when the -in terminal is positive, the output is negative. Over a period of time based on the resistance and capacitance, the voltage on the -in terminal will fall towards zero volts, and will eventually become negative - the output promptly swings positive, and the cycle repeats. Like all filter circuits, the resistor / capacitor (R/C) network introduces a time delay, and it is this (plus the simple low-pass filter formed) that produces a sinewave with less than 1% distortion. By no means wonderful, but quite adequate for a number of simple applications.

The sinewave output is at relatively high impedance, and should be buffered with an opamp before use. Any loading will alter both amplitude and frequency.

Schmitt Trigger Oscillator

Also known as a free-running multivibrator, the Schmitt trigger oscillator is one that is much more conventional in terms of opamp designs. Like the phase shift oscillator (indeed, like all oscillators) it is an inherently unstable circuit. Also like the preceding example, this circuit cannot obey Rule #1 (since that would make it stable), so follows Rule #2 instead.

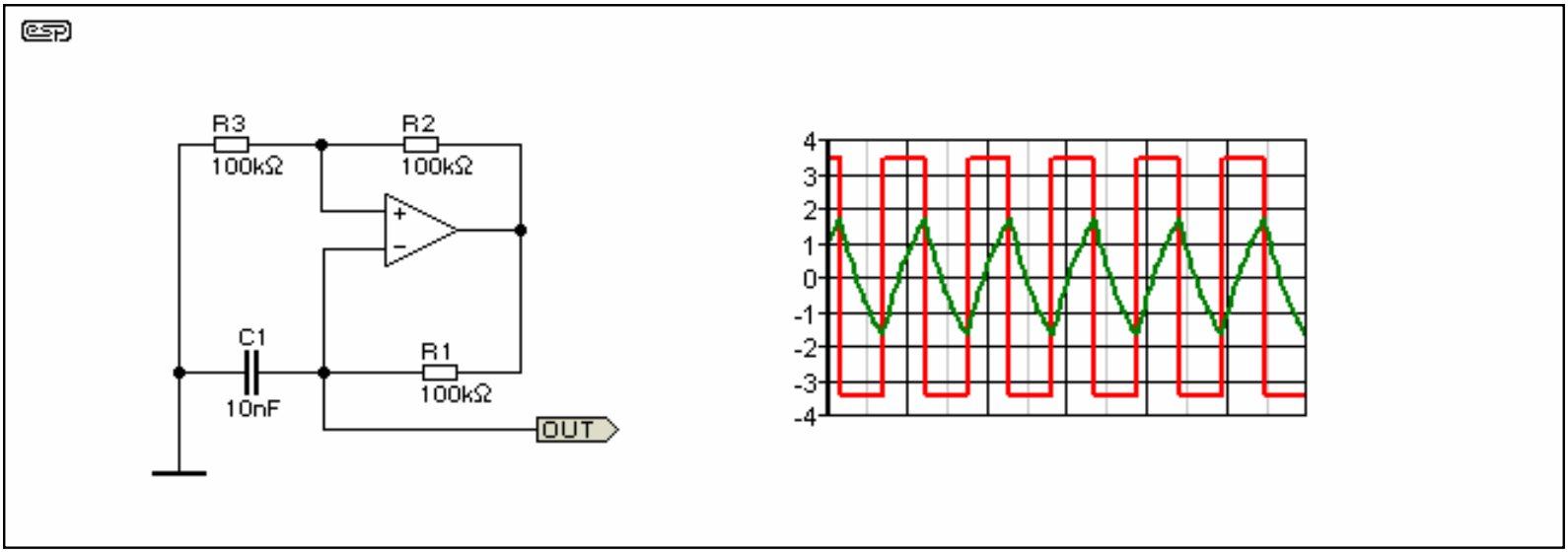


Figure 11 - Schmitt Trigger Oscillator

This circuit is very common where an oscillator is needed, but as shown produces a triangular waveform that is quite high in harmonic content. Note the use of positive feedback, via R2 and R3. This particular connection creates a Schmitt trigger, a useful and fairly inscrutable circuit for the beginner. Although this is a simple circuit, understanding how it works is not.

Assume a supply voltage of +/-15V, and ignore the losses in the opamp output stage. We shall start at the point where the opamp's output is at +15V. The +in terminal will be at 7.5V, since there is a voltage divider from the output to earth. C1 will therefore charge to a positive voltage, until such time as the voltage is very slightly greater than 7.5V. Since Rule #2 must be obeyed, the -ve input is now the more positive, so the output will swing negative. The cap now must discharge its positive voltage and start charging to a negative voltage. Once the negative voltage is less than (more negative than) -7.5V, the output will swing positive and the cycle repeats.

At least it is possible to determine the frequency of this oscillator, and it is approximately equal to ....

$$X = R3 / (R2 + R3)$$
$$Fo = 1 / (2 * R1 * C1 * \log_e((1 + X) / (1 - X))$$

For the example above (after noting that it is the natural log *not* base 10 :-) frequency is ...

$$X = 100 / 200 = 0.5$$
$$Fo = 1 / (2 * 100k * 10nF * \log_e(1.5 / 0.5))$$
$$Fo = 1 / (0.002 * \log_e(3))$$
$$Fo = 455\ Hz$$

The triangle wave output is at relatively high impedance, and should be buffered with an opamp before use. Any loading will alter frequency, but not amplitude (this is fixed by the voltage divider of R2 and R3).

Part 2 Part 3

References

I have used various references while compiling this article, with most coming from my own accumulated knowledge. Some of this accumulated knowledge is directly due to the following publications:

National Semiconductor Linear Applications (I and II), published by National Semiconductor
National Semiconductor Audio Handbook, published by National Semiconductor
IC Op-Amp Cookbook - Walter G Jung (1974), published by Howard W Sams & Co., Inc. ISBN 0-672-20969-1

Data sheets from National Semiconductor, Texas Instruments, Burr-Brown, Analog Devices, Philips and many others.

Suggested Reading

AN166 - Basic Feedback Theory, Philips Semiconductors Application Note, Dec 1988

Opamps For Everyone - by Ron Mancini, Editor in Chief, Texas Instruments, Sep 2001

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © Rod Elliott, 25 Apr 2000./ Updated 03 May-Added more info

Elliott Sound Products **Audio Designs With Opamps - 2**

Copyright © 2000 - Rod Elliott (ESP)

Articles Index

Main Index

Contents

- Part 1 - Opamp Basics
- Part 3 - More Filters, Comparators, Etc.
- Pure Audio Circuits
- Active (Baxandall) Tone Control Circuit
- Active Filters and Crossovers
- References
- Copyright Notice

Pure Audio

The next set of circuits are pure audio, and include tone controls, equalisers and filters, as well as balanced line driver and receivers. There is also a simple power amplifier suitable for use with headphones or low powered speakers.

When we speak of audio, the commonly accepted range is from 20Hz to 20kHz, and at some point someone decided that the centre frequency is 1kHz. I have never known why, since if you look at the octaves covered by the audio range we have a sequence like this ...

20 40 80 160 320 640 1280 2560 5120 10240 20480

This is a total of 10 octaves, so the centre frequency must actually be 640Hz, and indeed this is much closer to reality than 1kHz. If we were to divide the musical spectrum based on energy content, the centre frequency is about an octave lower, or 320Hz (approximately - it depends to a large degree on the style of music).

I am going to break with tradition and suggest that the 640Hz frequency is the midpoint, and 1kHz shall be ignored for other than comparative purposes.

So that the limitations of opamps in an audio circuit can be fully understood, we shall also look at the compensation curves typical of standard and high quality opamps, and examine the alternatives for very wide

bandwidth.

Baxandall Tone Controls

This is the most common of all modern tone control circuits, and was named after PJ Baxandall who came up with the idea many years ago. The original design was article was entitled "Negative Feedback Tone Control - Independent Variation of Bass and Treble Without Switches", and was published in Wireless World (now Electronics World) in 1952. This type of control is fully symmetrical, and there is no interaction between the controls, unlike the older passive controls. When centred, there is neither loss nor gain, and the opamp acts as a buffer. Frequency response is absolutely flat, provided the pots centre precisely.

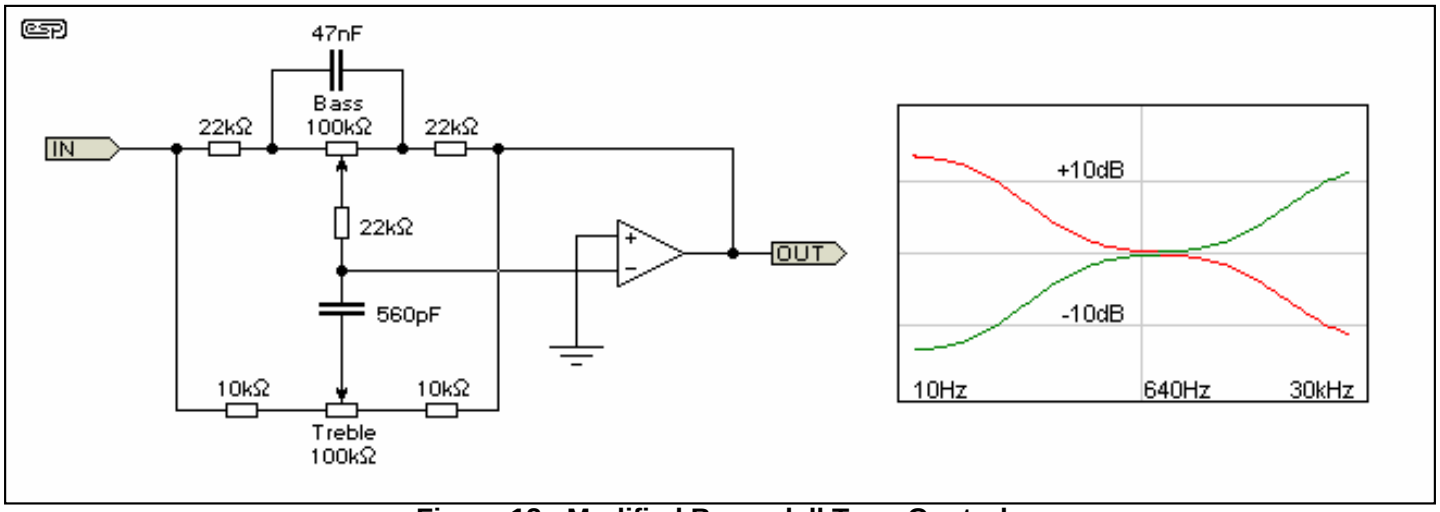


Figure 12 - Modified Baxandall Tone Controls

The circuit is a frequency dependent feedback arrangement, and provides boost and cut for high and low frequencies. It is common for designers to make the turnover frequency for both treble and bass centred around 1kHz, but this is essentially a stupid thing to do (IMHO). Ideally, bass boost/cut should start from no higher than about 160Hz, and treble boost/cut should start at no lower than around 2.5kHz - this is two octaves either side of the 640Hz nominal centre frequency for audio. In practice it will be found that this is more natural, and provides the boost where it will do the most good (or harm, for the purists!).

Figure 12 shows a more or less conventional circuit, with the turnover frequencies set to my preferred values. This has the effect of limiting the maximum boost and cut within the audible spectrum to somewhat less than the 20dB often quoted. IMO, this is an outrageous amount of control, and far exceeds what is needed for normal system balancing. If 20dB of anything is needed, then I suggest that the system is grossly inadequate, and should be replaced!

The maximum boost is about 14dB, which is still far more than necessary, the lower turnover frequency is about 150Hz, and the upper turnover frequency is about 2.5kHz. These can be changed by modifying the values of the bass and treble capacitors, and the amount of boost and cut is varied by changing the series resistors for each pot.

This circuit must be driven from a low impedance, so connecting it after the volume control (for example) is a no-no. Ideally, the output of an opamp will be the source, thus ensuring the required low impedance.

The formulae for designing these active tone control circuits are fairly simple, but can be quite inaccurate under some circumstances. At least it will give you a starting point to have the formulae, so the formulae for

the bass control lower turnover frequency (Fb0) and +/-3dB frequency (Fb1) are (respectively) ....

> Fb0 = 1 / (2 * π* C * Rv) ... where C is the cap across the pot, and Rv is the pot value
> Fb1 = 1 / (2 * π* C * Rs) ... where C is the cap, and Rs is the series resistance to the pot

These give (again, respectively) ...

> Fb0 = 1 / (2 * π* 47nF * 100k) = 33 Hz
> Fb1 = 1 / (2 * π* 47nF * 22k) = 154 Hz

The treble is a little trickier, since there is an interaction with the bass control, due to the bass feed resistance from the bass pot wiper. In theory (ha ha) the frequency is determined by

> Ft = 1 / (2 * π* C * Rb) ... where C is the treble cap (560pF above) and Rb is the bass feed resistor from the pot wiper.

While this seems reasonable, the formula comes up with a frequency of nearly 13kHz, which is obviously wrong. There are series resistances in the entire treble control network, and the interactions are interesting to say the least. In fact, the +/- 3dB frequency for the treble control changes with the pot setting. For example, at 80% of rotation, treble boost is +3dB at just over 3.5kHz, with a maximum boost of 6.3dB at 20kHz. This may seem to be a bad thing, having the frequency shifting about as the pot is varied, but in practice it works very well. To obtain the results *you* want, I suggest that experimentation is in order !

A final word about the tone control circuit. Note that it is a 'virtual earth' circuit, so the feedback at all times will maintain the -ve input at zero volts. The bass and treble content of the input waveform will force the amplifier to provide just the amount of boost or cut at any frequency to maintain the 0V condition on -in. You will find this useful as you work towards an understanding of the complete circuit.

Active Filters and Crossovers

The active filter is one of the most common opamp applications after basic amplifiers. There are essentially four different types of filter, but each can be constructed using a myriad of different techniques. I shall only use the most common arrangements, since the complete array is truly mind-boggling. The area of filters is the subject of complete books (from a number of authors) and I cannot hope to cover any but the most common.

Firstly, there are many different possibilities for filters. We shall look at each of the different responses, orders and types before anything else. The filter "order" gives us essential information about how well (or otherwise) a given filter will reject unwanted frequencies. The standard ones are:

Order	**Sections**	**Nominal Rolloff**
First	1	6 dB / Octave
Second	2	12 dB / Octave
Third	3	18 dB / Octave
Fourth	4	24 dB / Octave
"n"	Where n > 4	6 dB / Octave / section

Table 1 - Filter Orders

In audio work, the first four are most commonly used, since as the filter order increases, the transient response becomes worse and greater phase disturbances become evident. All filters affect the phase of the signal, and all filters have some effect on transient response. These are unavoidable, regardless of whether the filter uses valves, transistors or opamps, or is completely passive, using only capacitors, resistors and/or inductors.

There are 4 main responses that are obtainable from filters. Some are - or can be - derived from others, so the base types are given first. In particular, band pass and band stop filters can be simple filters with a narrow response (typically used for equalisation circuits), or band pass filters made from a combination of a low frequency high pass filter, followed by a higher frequency low pass section. These are commonly used in crossover networks.

Response	Passes ...	Blocks ...
Low Pass	Low frequencies	High frequencies
High Pass	High frequencies	Low frequencies
Band Pass	Selected frequency	All other frequencies
Band Stop	All other frequencies	Selected frequency

Table 2 - Filter Responses

Low and high pass filters are usually conventional enough, but band pass and band stop filters can be made in many different ways.

Essentially, there are also a few basic filter alignments, which are as follows (not all are included, and many are applicable to filters of more than one section, or more complex arrangements):

Filter Type	Q (Typ)	Characteristics
Bessel	0.57	Best time delay
Butterworth	0.707	Flattest amplitude
Chebychev	0.8 - 1.3	Fast initial rolloff
Cauer / Elliptical	0.7 - 1.3	Very fast initial rolloff

Table 3 - Filter Types

Of the above, the Butterworth is the most common in audio. Although some filter's responses may be closer to Chebychev, this is commonly more by accident than design. A Chebychev alignment is very common in acoustical filters (a loudspeaker - box - port combination, for example), but is not generally considered desirable in electronic filters for crossovers or other purposes.

You may have noticed that the Linkwitz-Riley filter was not mentioned in the above table. This is because it is an alignment between filters, and not an alignment type itself. The Linkwitz-Riley filter is a rearrangement of two cascaded 2nd order Butterworth filters, and relies on the characteristics of the two sections (high pass and low pass) to provide the total amplitude and phase response.

Finally, there are different circuit arrangements that are commonly used in audio to create the filters described above. As you can see by now, the combinations and permutations of all of these different

possibilities is immense.

Circuit Type	Gain	Characteristics
Sallen-Key	Unity	Simple design
Equal component value	Depends on Q	Simpler design
Multiple feedback	Depends on Q	Relatively complex
State Variable	Depends on Q	Relatively complex
Biquadratic	Depends on Q	Relatively complex
Passive	< unity	Simple or complex

Table 4 - Circuit Types

This is a condensed listing, and there is some overlap between the topologies in some cases. Indeed, in some cases it is difficult to see any appreciable difference at all. For the sake of simplicity (and to keep this section to a readable length), I shall concentrate on the Sallen-Key filter type, using Butterworth alignment.

Before we go there, we need to define some of the terms you will see. Again, I have kept this list to the minimum for the sake of simplicity, but armed with this knowledge you will be able to understand the circuits and descriptions that follow.

Terms and Definitions

- **Cascading:** Simple filters can be made with one (or none) active devices. As filter orders become greater, it is no longer feasible, so filters are joined together in series to obtain the desired response. This complicates the design (often dramatically).
- **Cutoff Frequency:** Also shown as Fo, this is the -3dB frequency of the filter, relative to the highest peak (if any exist) in the passband.
- **Decade:** A 10:1 (or 1:10) ratio of frequency. For example 100Hz to 1000Hz is one decade.
- **Decibel:** (dB) - the most common way to describe amplitude in audio. The dB scale is logarithmic, and describes the amplitude as we hear it. A 3dB drop in gain equates to half the power in an amplifier.
- **Quality Factor:** Commonly known as Q, this is the inverse of the filter's damping. For example, a Butterworth filter has a Q of 0.707, which equals a damping of 1.414. Higher Q gives more ripple for low and high pass filters, or makes a bandpass or band stop filter more selective.
- **Octave:** A 2:1 (or 1:2) ratio of frequency. For example, 440Hz to 880Hz is one octave (Musical "A" note).
- **Order:** The order of any filter determines its rolloff frequency response. First order filters roll of at 6dB per octave, and as the order increases, so too does the rolloff rate. An additional 6dB / octave is gained for each additional order, starting from first. A 3rd order filter will therefore roll off at 18dB/octave, for example.

Designing Our First Filter

The first active filter must be the first order low pass. By simply reversing the filter components, this becomes a first order high pass. In each case, the filter only consists of the resistance and capacitance, with the opamp simply isolating the filter from the following stages. Q is not variable in a first order filter, and the only options are high pass and low pass. A bandpass filter can only be created by cascading a high and low pass filter. Although the least useful of all filters, they are easy to understand, so make a good starting point.

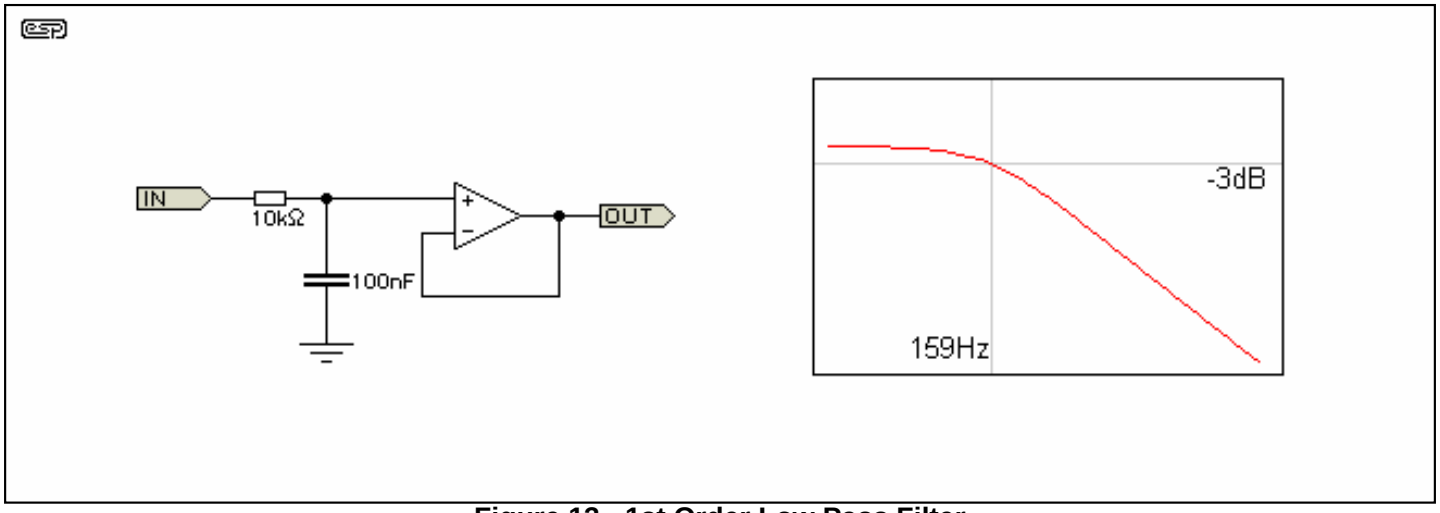


Figure 13 - 1st Order Low Pass Filter

As you can see, at low frequencies, the capacitor has little effect on the signal, which is simply passed through the resistance and buffered by the opamp. As the frequency increases, the capacitor will shunt more and more of the signal to earth, until at very high frequencies, no appreciable amount of the signal is passed. As the capacitor's reactance becomes significant with respect to the resistance, the signal will be subjected to a phase shift as well as reduction in amplitude. When the capacitive reactance is equal to the resistance, the amplitude will be 3dB lower than at low frequencies.

Rc = 1 / (2 * &pi * F * C) ... where Rc is capacitive reactance

This is the cutoff frequency of the filter, and is determined with the formula

Fo = 1 / (2 * π* R * C) commonly shown simply as ...
Fo = 1 / (2 πR C)

Note that the input of this filter *must* have a low impedance return to earth at the input, or the opamp will not have any bias current or voltage, and will not work. This rule applies to low pass filters in all cases without exception.

For the filter in Figure 13, the frequency is determined by ...

Fo = 1 / (2 π* 10k * 100nF) = 159 Hz
Rc = 1 / (2 π* 159 * 100nF) = 10k ohms

By reversing the positions of R and C, we obtain a high pass filter. The formula remains the same, and the two filters will have a complementary response centred at the 159Hz frequency. There is no need for an earth return at the input for a high pass section, as this is provided by the resistor.

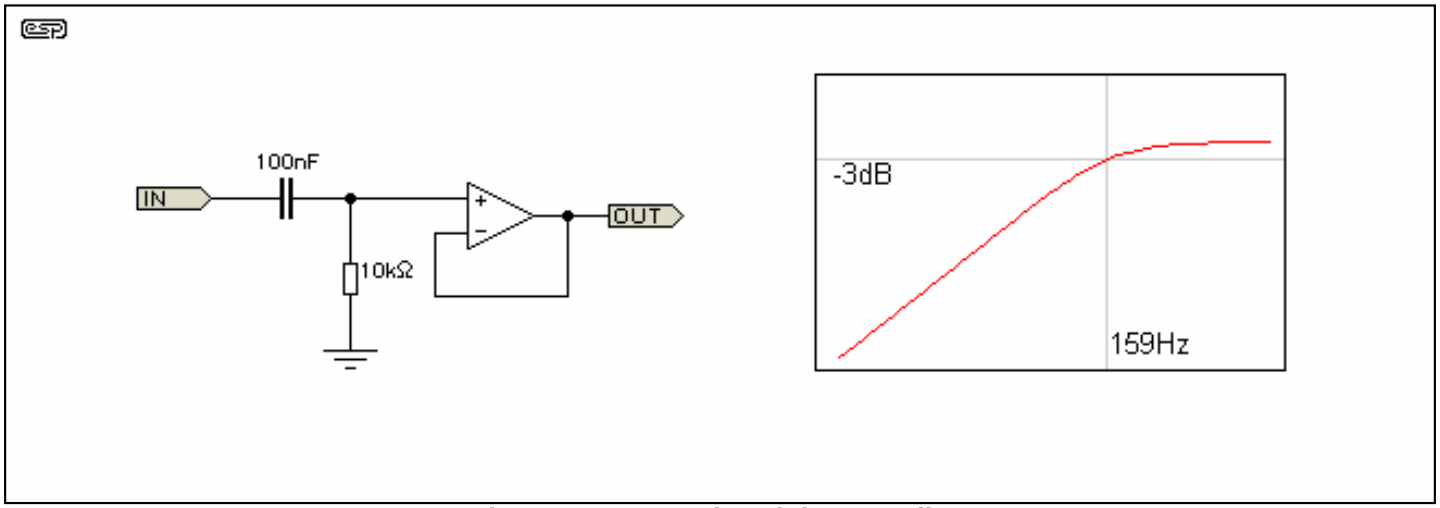


Figure 14 - 1st Order High Pass Filter

In their simplest forms, these two filter sections are also known as integrators (low pass) and differentiators (high pass). This is of some consequence, as these terms are commonly used in electronics.

A bandpass filter is created by cascading a high pass and a low pass, as shown in Figure 15. An opamp may be used in between the two sections to prevent any interaction. As shown, it is also possible to scale the first filter so that interaction is minimised. This works well enough in practice to be a useful technique. Scaling merely means that the ratio of values remains the same, but the resistance is reduced and the capacitance increased to make a lower impedance filter with the same characteristics. The generally accepted scaling factor is one decade (an "order of magnitude", or a 1:10 ratio). For Figure 15, I simply selected the lowest sensible resistor value of 1k, and the capacitor was chosen more or less at random to give an acceptable graph of the response.

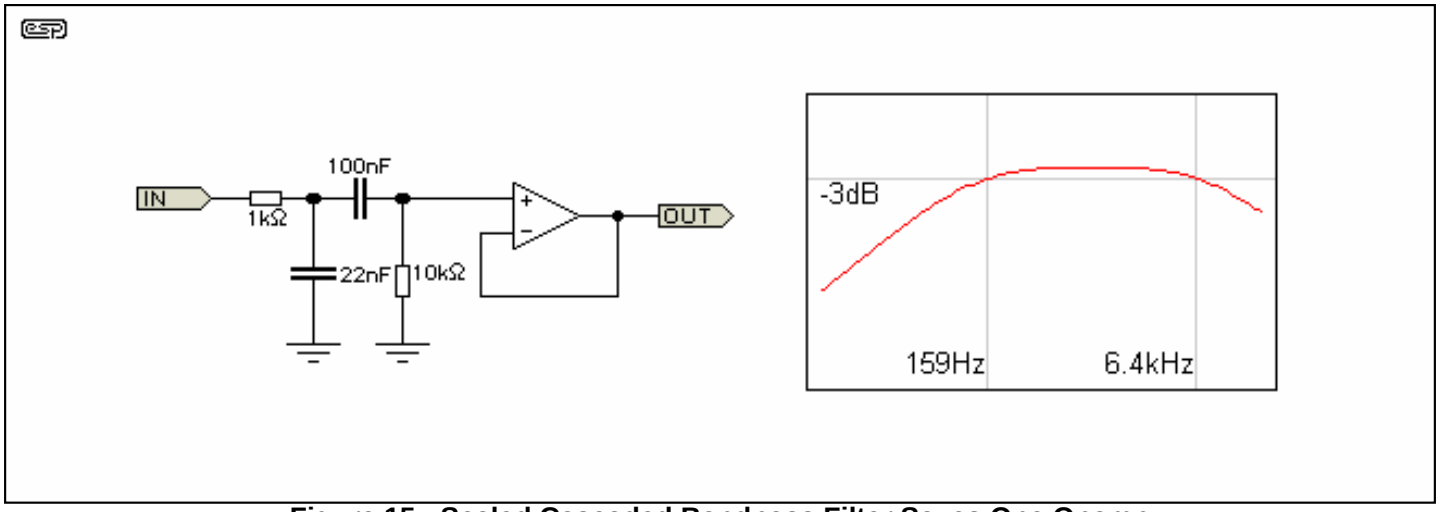


Figure 15 - Scaled Cascaded Bandpass Filter Saves One Opamp

In reality, it may be found that the value of capacitance becomes so great that the capacitor will be more expensive than a section of an opamp. Alternatively, the resistance may become so low that no opamp can drive the load. In all cases, it is essential to ensure that impedances are within the acceptable range for the opamps used. Generally, impedances of less than 1k at any frequency are not recommended. For noise considerations, very high resistance values are also not recommended, and I suggest that 100k is a reasonable compromise. There will be occasions where this is not practicable and higher values may be the only sensible solution, but these should be few and far between.

If the high pass section of Figure 15 is verified by calculation, it is discovered that the cutoff frequency (Fo) should be 7234Hz, and not 6400Hz as shown. The difference is caused because the second filter loads the first, shifting the frequency. This is a very good reason to isolate the sections using an opamp.

Component Selection

All following filter sections will use resistance within the range of 1k to 100k. This gives two decades of freedom, and this is more than enough to allow capacitors of sensible sizes to be used. Very low capacitance values are to be avoided, since the capacitance of wiring (PCB tracks etc.) will modify the filter characteristics to a possibly unacceptable degree. I generally try to keep capacitance above 1nF wherever possible, and this may save you some grief as we progress. Likewise, any capacitance above 1uF becomes large and relatively expensive, but within this range we also have two decades of freedom, and there is virtually no audio filter that cannot be designed within these constraints.

Generally, it is sensible to select the capacitor value first, as these have less available values within a decade than resistors. Capacitors have 12 values per decade (the E12 series), while resistors have up to 96 values per decade (E96 series). The latter are not usually easy to get, but the E24 series is now very common, and has (surprise!) 24 values per decade. The E12 and E24 series are shown in Table 5.

E12	1.0		1.2		1.5		1.8		2.2		2.7		3.3		3.9		4.7		5.6		6.8		8.2	
E24	1.0	1.1	1.2	1.3	1.5	1.6	1.8	2.0	2.2	2.4	2.7	3.0	3.3	3.6	3.9	4.3	4.7	5.1	5.6	6.2	6.8	7.5	8.2	9.1

Table 5 - E12 and E24 Component Values

These values are multiplied or divided as needed for any decade range from 0.1 Ohm up to 10M Ohms, and from 10pF up to 10uF. Generally it will be found that at the extremes of the ranges (such as from 10pF to 100pF), most stockists do not have the full range of values. This is another good reason to stay within reasonable limits for all component values wherever possible - not just for filter designs.

Second Order Filters

Note that all response graphs shown in this section cover the frequency range from 10Hz to 10kHz, and all but the Chebychev response are from 0dB to -20dB (the Chebychev is from +10dB to -20dB). The reference input voltage is 1V RMS, or 0dBU.

By using multiple feedback paths, a second order filter can now be designed. These are the first genuinely useful filters for crossovers and the like, and are the most commonly used in both electronic and passive crossover networks. There are some unpleasant side effects to the second order high and low pass filters, but this has never stopped anyone from using them. As we shall see, these effects can be cancelled to some extent, but unless exotic configurations are used you can never get phase coherency.

A phase coherent filter gives the design some special characteristics that are extremely useful in audio. This simply means that at any frequency within the pass band or stop band of either filter, the output signal from each is in phase, preventing any peaks or dips in the combined response. These will be covered in more detail a little later.

Of all the topologies available, I will concentrate on the Sallen-Key (also known as unity gain) Butterworth filter. The equal component value filter is useful in some areas, but is a nuisance because of the gain that

each section adds. This is sometimes useful, but mostly is not necessary, and the filter section needs more components. They are easier to design, but the difference is slight, and often the resistor needed to set the gain precisely will work out to be impossible to obtain. For the values I am using, a 14.414k resistor would be needed, which as you can see from Table 5 is not a standard value.

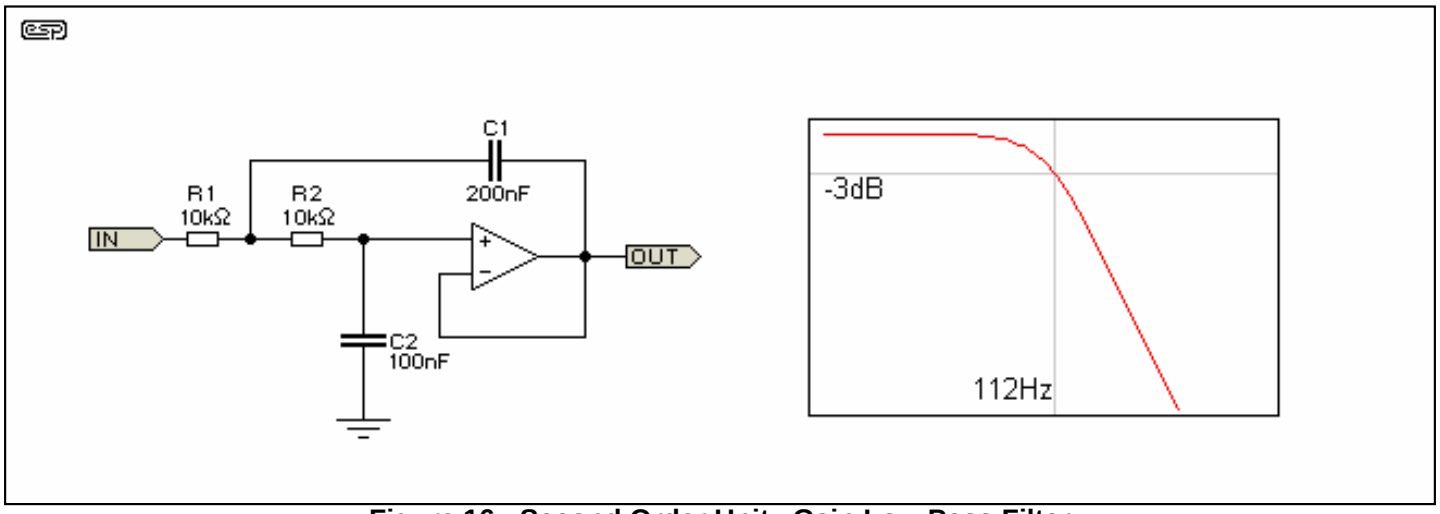


Figure 16 - Second Order Unity Gain Low Pass Filter

As you can see, the opamp is used only as a buffer, contributing no gain. For Butterworth alignment, the component values are as shown. The filter can be made into anything from Bessel to Chebychev by changing the component values about, but these other alignments are not generally as useful in audio work.

If R has been selected first, the capacitor values for C1 and C2 are chosen from the equations

$$C1 = 4 / d^2 * C2 = 2 * C2 \quad \text{... where } d = 1 / Q$$
$$C2 = 0.707 / (2 * \pi * F * R)$$

The formula for calculating the value of C1 *only* applies to a Butterworth filter. Of course if you were to follow my advice from above and select the capacitance first, you need a different formula ...

$$R1 = 0.707 / (2 * \pi * F * C2)$$

Again, the equation for resistance or capacitance for each example *only* works for Butterworth filters! For a Q of 0.707 or damping of 1.414 (Butterworth) it works out that C1 is exactly double the value of C2. R1 and R2 must be equal in value, or the filter's response will be something other than that desired.

Any change of Q by varying the resistance ratio or capacitor ratio also changes the frequency. The formulae become quite complex, and I am not going into further detail.

For example, if the C1 is four times as great as C2, this creates a Chebychev filter with a Q of 1, as shown in Figure 17 for the sake of example. Equal capacitor values would be used for a 'sub-Bessel' alignment with a Q of 0.5 - the possibilities are endless, as I am sure you can now appreciate.

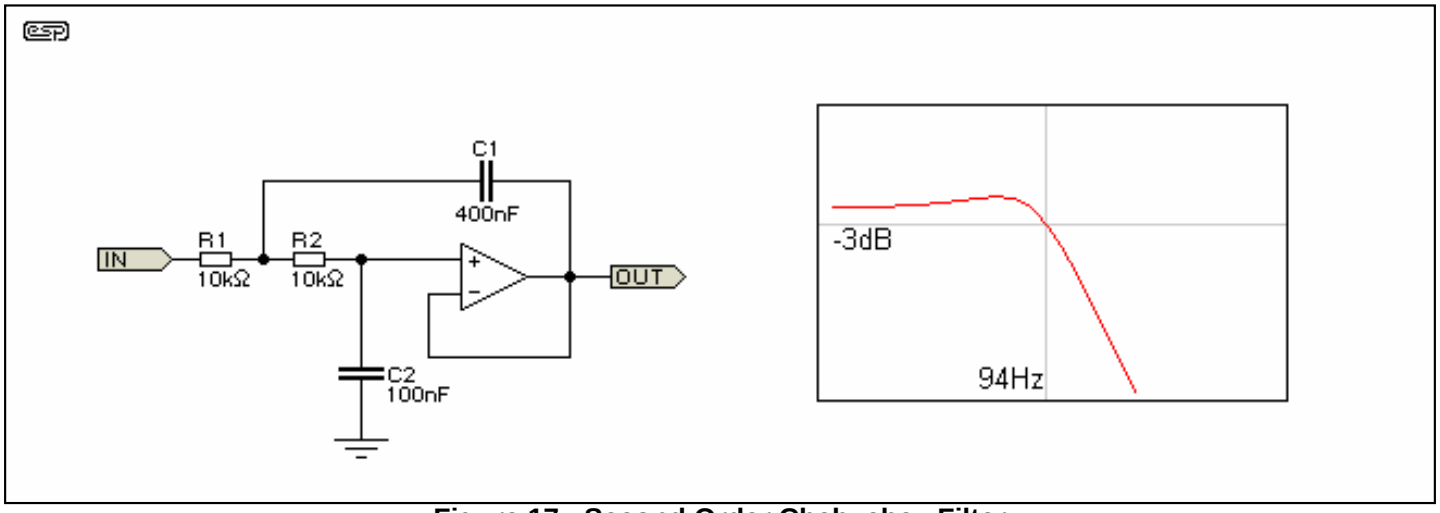


Figure 17 - Second Order Chebychev Filter

The filter of Figure 17 clearly shows the peaking obtained from a Chebychev filter. The peak amplitude is +1.25dB from the nominal 0dB value, and the cutoff frequency is determined to be 3dB below this maximum. I remain unconvinced that this is the right way to measure the cutoff frequency - I think I prefer to use the "genuine" -3dB point. Either way, the cutoff frequency is not easy to calculate, and although it would be no great strain for me to give you all the equations, I doubt that you would want to know!

The high pass equivalent of the Butterworth low pass filter is shown in Figure 18, and as you can see is a re-arrangement of the other design. Frequency is the same as for the low pass filter, but note that now C1 and C2 are the same value, and R2 is double the value of R1. Frequency is calculated on C1 and R1, and it is easy to become confused when designing the circuits just which values determine the frequency.

In both cases it is nearly always easier to use paralleled capacitors and series resistors as I have shown, since these will be more accurate than simply doubling (or halving) the values - especially so since the standard values do not always have exact double or half values. Any variation of these values will shift the response away from Butterworth and towards either Bessel or Chebychev responses, and will change the frequency.

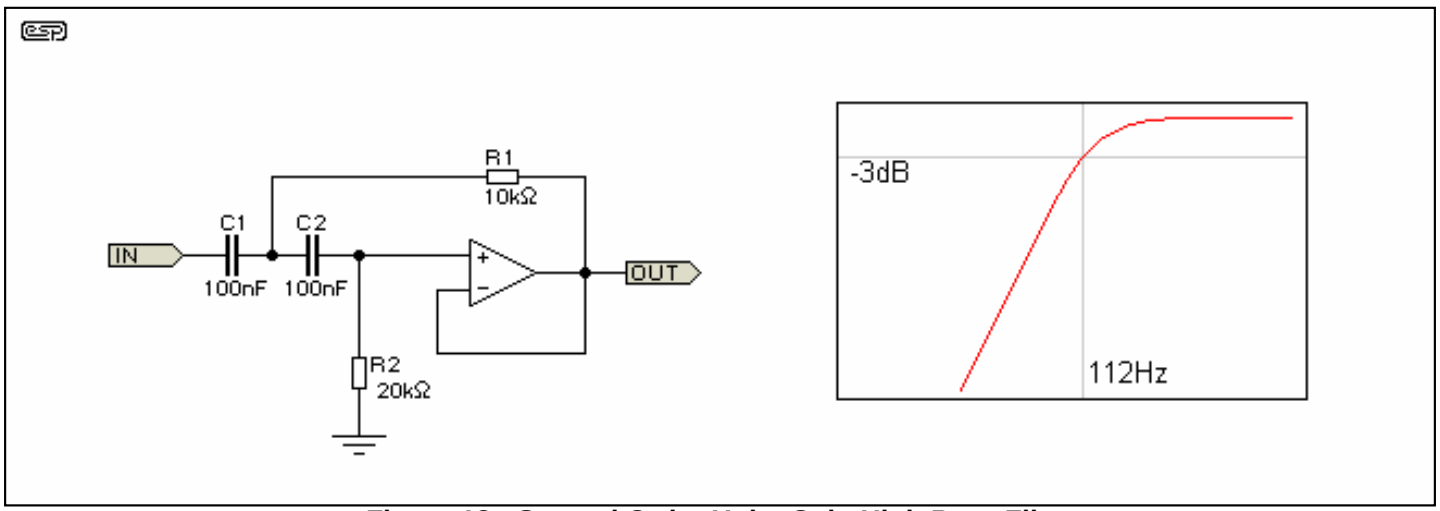


Figure 18 - Second Order Unity Gain High Pass Filter

Second order filters can also be used for bandpass and band stop, but are generally of limited use in audio circuits. They are sometimes used as equalisation circuits, and indeed the simple inductor / capacitor (LC) filters shown in Figure 8 are 2nd order types. A second order filter requires that there are two reactive

elements, with an filter LC there is the capacitance and the inductance, and with an active filter there are two capacitors.

Crossover Networks

When combined, high pass and low pass filters are commonly used to create electronic crossover networks. There are many different ways to do this, but the most common is still the second order Butterworth filter. Although these are essentially at least as good as the passive speaker level counterpart (but not all that much better), the results are often vastly superior. The reasons for this are discussed at great length in my article Bi-Amplification - Not Quite Magic (but Close) and I shall not repeat them here.

In a nutshell, using an electronic crossover eliminates many of the problems that beset loudspeaker designers when they have to design and build the crossover network. The passive crossover is influenced by any aberration in the driver's impedance, especially at or near the crossover frequency. Since the electronic crossover supplies the signal to a separate amplifier for each frequency band, there is no interaction and each amp only needs to be concerned with a much smaller bandwidth.

All crossover networks are a combination of high pass and low pass filters, although this is not always achieved in the same way. Some crossovers use an opamp as a subtracting amplifier, so rather than using a separate filter, the bass (for example) is subtracted from the main signal to provide the midrange and high frequency. Alternatively, the mid+high output from the filter is subtracted from the main signal to separate the bass, and this configuration is shown below. These crossovers are phase coherent (both outputs are always in phase at any frequency), but are asymmetrical. A typical design is shown in Figure 19, and the response clearly shows that the high pass rolloff is 12dB/octave, but the low pass is only 6dB/octave.

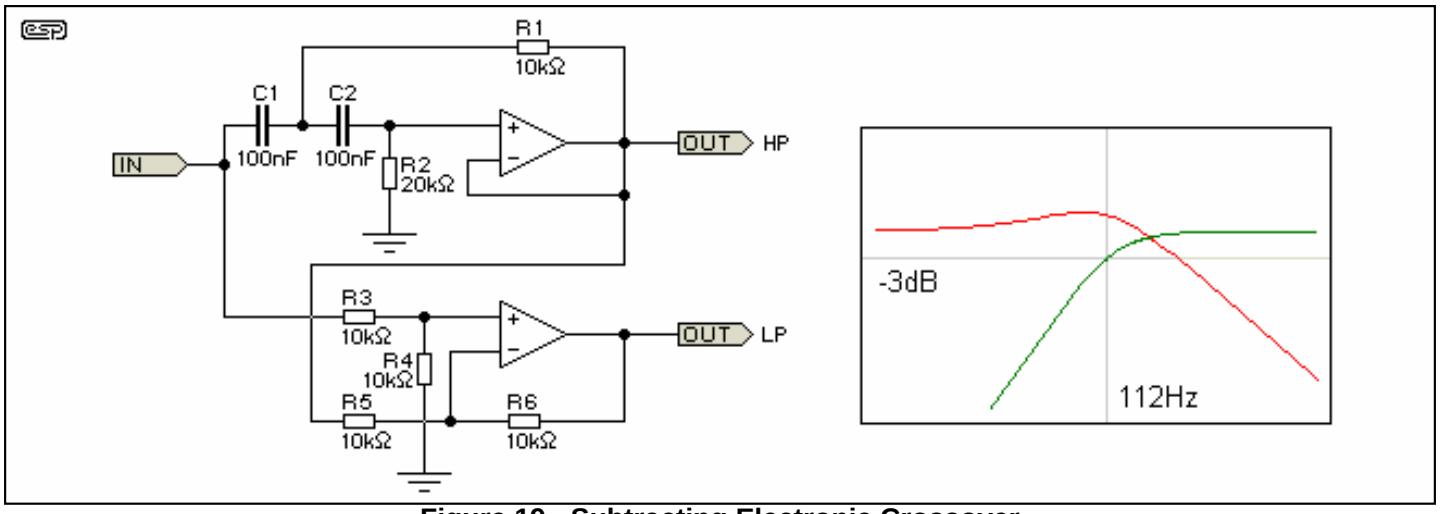


Figure 19 - Subtracting Electronic Crossover

Although these networks can be capable of excellent results, great care is needed to ensure that the driver getting the 6dB/octave rolloff can handle the increased power levels created by such a low rolloff rate. Notice that the LF rolloff peaks, and that the crossover frequency does not coincide with the actual -3dB frequency of the high pass section. Although this looks really bad, because of the phase differences the combined frequency response is completely flat. With loudspeaker drivers this will often not be the case, especially when listening off axis. Because of this, the subtracting crossover is really only useful at low frequencies (IMHO), where phase is less likely to cause major problems because the wavelength is so long. Personally I have never been a fan of this type of crossover, but the circuit is interesting, and shows the versatility of opamps.

Rather than show a multitude of crossover designs here, you can look at the Projects Pages to see a suitable sampling of crossover networks. Of these, the 24dB/octave Linkwitz-Riley is by far the best, and is highly recommended.

Notice the cunning way I introduced a new opamp configuration - the subtracting (or difference) amplifier. This (and other useful topologies) will be discussed in greater detail a little later.

Band Pass Filters

The band pass filter (as a single filter) is not normally very useful in audio reproduction. The frequency range passed is too narrow to be useful for anything other than equalisation circuits, or for test and analysis equipment. Having said this, there are many uses for bandpass filters that are very common in the *production* of music - the synthesiser, guitar wah-wah pedals, Vocoder, etc.

The range is far too great to cover in any real detail, but we can at least look at the fundamental principles, as these are common to all of the applications. The two most common parameters quoted for bandpass filters are frequency (f_o) and Q ("quality factor"). The latter may be inverted and referred to as damping.

A simple bandpass filter consisting of two reactive elements has an ultimate rolloff of 6dB/octave. When one looks at resonance, the slope appears to be much greater than this, which is fair and reasonable, since it is. Eventually, the high slope due to resonance effects cannot be maintained, and the final slope is at 6dB/octave, as shown in Figure 20. But wait! This is an article about opamp designs, and there is no opamp there, just a stupid inductor and capacitor. True, but we need to be able to understand the concept of resonance before delving into the opamp version of the circuit.

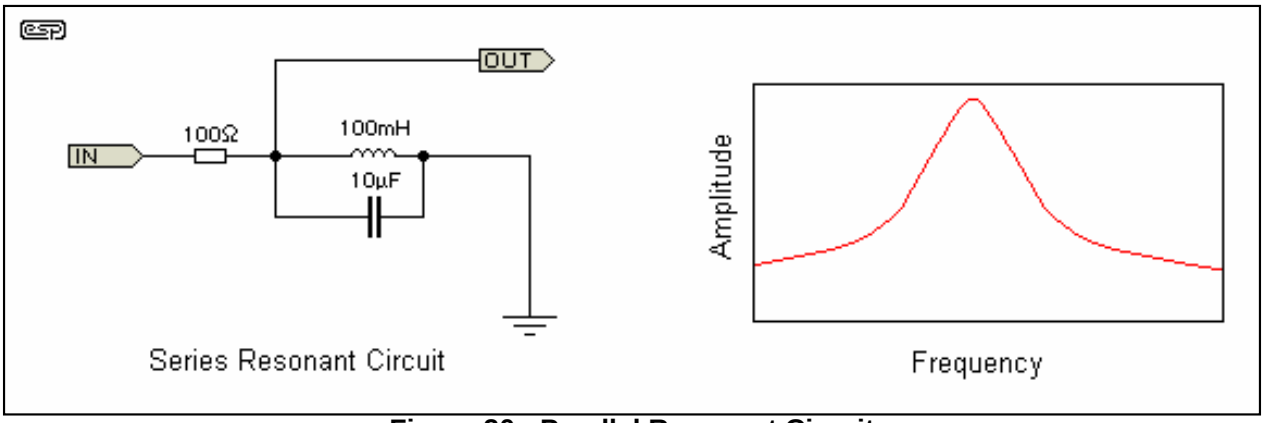


Figure 20 - Parallel Resonant Circuit

There are essentially two forms of resonant circuit - series and parallel. When only passive components are used (resistor, inductor and capacitor), the series resonant circuit has minimum impedance at resonance, and the parallel resonant circuit has maximum impedance. The formula is actually slightly different for each (the resistance of the coil changes the resonant frequency slightly in a parallel resonant circuit), but for all intents and purposes the following formula can be used with little error ....

$$f_o = 1 / (2 * \pi * \sqrt{LC})$$

Where f_o is resonant frequency, L is inductance in Henrys and C is capacitance in Farads. Any error caused

by the series resistance of the coil will generally be less than that caused by component tolerance.

In all cases, the resonant frequency is that where the capacitive and inductive reactance is equal. In the example of Figure 20, only when Xc (capacitive reactance) and Xl (inductive reactance) are the same will the circuit be at resonance. As the frequency is reduced, the inductive reactance decreases, and shunts more of the signal to earth. When the frequency is increased above resonance, the capacitor will be responsible for shunting the signal to earth. Since in each case there is the resistance and a single capacitor or inductor to bypass the signal, this is a single pole filter and will have 6dB/octave rolloff. As shown, the reactances at resonance are equal to the series resistance, so the rolloff slope is 6dB/octave. The initial slope can be increased by increasing the series resistance (and thus the Q of the circuit), but eventually the rolloff will go back to 6dB/octave.

Series resonance is especially interesting. At resonance, the circuit is almost a short circuit, so the input signal will be heavily loaded. At the junction of the inductor and capacitor there will be a voltage that is many times that at the input. This effect is described below. If enough current is available, incredibly high voltages may be obtained, and great care is needed to ensure that the voltage rating of the components is sufficiently high to withstand the voltage. Since the laws of physics and the taxman dictate that we cannot get something for nothing, the available current is very limited at the high voltage point. A circuit with a very high Q will generate much greater voltages than a circuit with low Q. The voltage magnification is equal to the Q of the circuit. With the component values as shown in Figure 21, the voltage at the output will be about 0.5 mV at resonance, but there will be 1V across both the inductor and capacitor. At resonance, the reactance of the capacitor and inductor are equal and opposite, so the circuit will appear to be almost a short circuit - input current is limited only by the wiring resistances and source impedance.

The maximum possible Q of an LC filter is dictated by the series resistance of the inductor. The Q may be reduced by adding resistance, but cannot be increased. The circuit Q is determined by the reactance of the cap or coil at resonance and the series resistance (for parallel resonance) or shunt resistance for series resonant circuits.

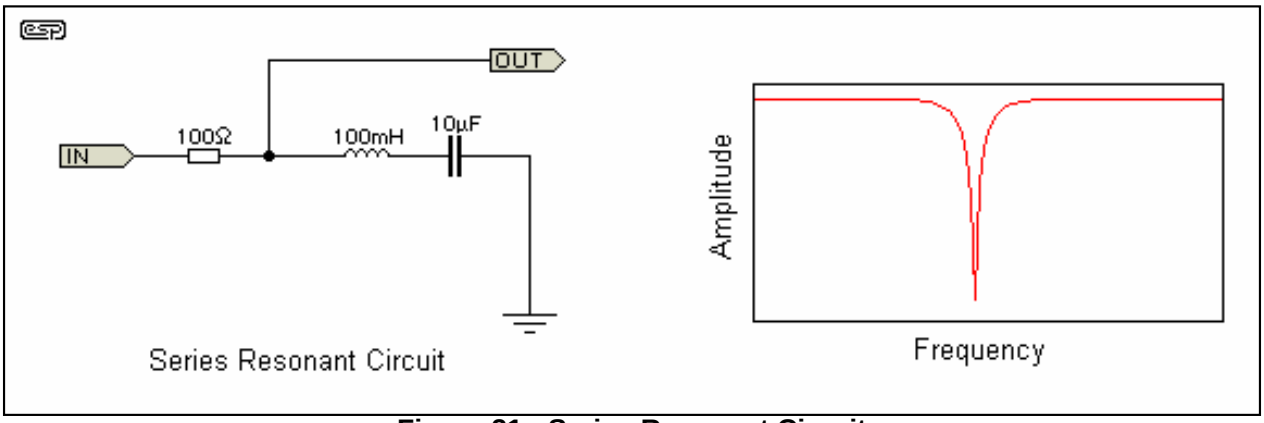


Figure 21 - Series Resonant Circuit

At very high frequencies, the so-called "skin effect" increases the apparent resistance of the inductor. The skin effect is where the electrons tend to want to occupy the outer section of the wire and are not evenly spread through the conductor. This gets progressively worse as the frequency increases. Likewise, the dielectric absorbtion of capacitors reduces their efficiency and lowers the overall Q. We shall not investigate these effects further, since they are unrelated to audio frequencies generally, and especially to opamps.

A quick word about Q. In both cases above, resonance is at 159Hz. The reactance of either the inductor (Xl) or capacitor (Xc) at this frequency is 100 ohms, so with a 1k series resistance (parallel resonance) or 10 ohm

shunt resistance (for series resonance), the circuit Q will be 10. The reactance can be calculated from ....

Xc = 1 / (2 * *pi; * f * C)
Xl = 2 * π* f * L

This basic understanding of the ratios is essential in the design of bandpass and band stop filters, and is also very much a part of the design process for passive loudspeaker crossover networks. The latter have absolutely nothing to do with opamps, but I just thought I'd mention it :-)

The Q of a resonant circuit determines its bandwidth. With a Q of 10, bandwidth of our filters above is 15.9Hz. This places the -3dB frequencies at 151Hz and 167Hz (near enough), so it is apparent that the range of frequencies allowed through is very limited. In practical audio work, this is far too narrow to be useful, so lower Qs are far more common. This is fortunate, because high Q opamp bandpass filters are difficult to design, and require opamps with very high bandwidth for proper operation at the upper end of the audio band.

Part 1 Part 3

References

I have used various references while compiling this article, with most coming from my own accumulated knowledge. Some of this accumulated knowledge is directly due to the following publications:

National Semiconductor Linear Applications (I and II), published by National Semiconductor
National Semiconductor Audio Handbook, published by National Semiconductor
IC Op-Amp Cookbook - Walter G Jung (1974), published by Howard W Sams & Co., Inc. ISBN 0-672-20969-1
Active Filter Cookbook - Don Lancaster (1979), published by Howard W Sams & Co., Inc. ISBN 0-672-21168-8
Data sheets from National Semiconductor, Texas Instruments, Burr-Brown, Analog Devices, Philips and many others.

Recommended Reading

Opamps For Everyone - by Ron Mancini, Editor in Chief, Texas Instruments, Sep 2001

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) Rod Elliott, 08 May 2000

Elliott Sound Products **Audio Designs With Opamps - 3**

Copyright © 2006 - Rod Elliott (ESP)

Articles Index

Main Index

Contents

- Part 1
- Part 2
- Common Mode Circuits
- Balanced Output Circuits
- Summing Amplifiers
- A Few More Filters
- Basic Power Amplifiers
- Basic Specifications
- Comparators
- Conclusion
- References
- Copyright Notice

Common Mode Circuits

Probably the best known common mode circuit is the single opamp balanced receiver circuit. While it has a number of perceived problems in real life, it is nonetheless a good place to start. The problems with the circuit are normally not a limitation. The schematic is shown below, and there are two circuits shown. The first shows the circuit the way it is normally used, with the input (source) connected in differential mode to the opamp inputs.

The second circuit shows how the circuit connects to the input for noise input - it is coupled (ideally) equally to both inputs at once.

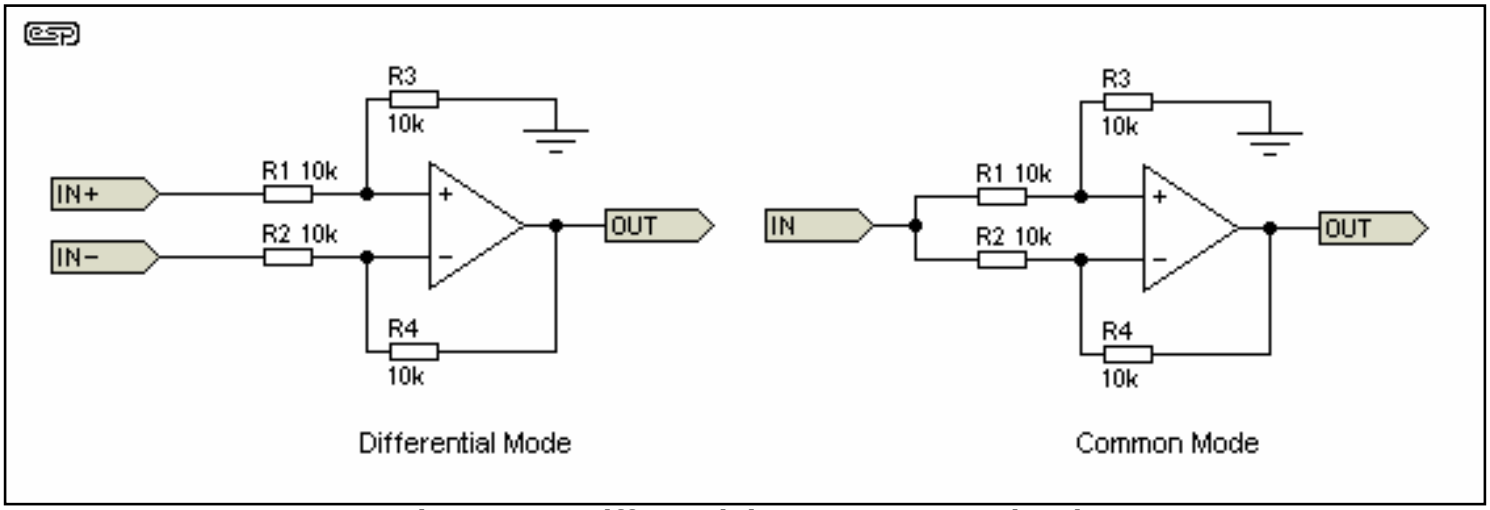


Figure 22 - Differential Input Opamp Circuit

While the wanted signal is passed directly through to the output (but is now a simple ground referenced signal at that point), any noise is presented to both inputs at equal level. This causes the cancellation of the noise, while allowing the signal to pass without alteration. This is shown in Figure 23, where we have a microphone with 10mV (differential) output, in the presence of a 100Hz 1V common mode noise signal. This is a ratio of 1:100 of signal to noise (or 100:1 noise to signal).

Although the results I obtained are simulated, the reality is not much different. At the output, the interference signal was measured at 27uV, while the signal level remains at 10mV. This means that the wanted signal is 51dB *greater* than the interference after the opamp, where the external noise level is 40dB greater than the signal at the input. The common mode rejection is therefore 91dB (1V common mode input, 27uV noise output). This is under ideal conditions, but in practice it is usually possible to get performance that exceeds the ability of the cable to maintain a perfect balance.

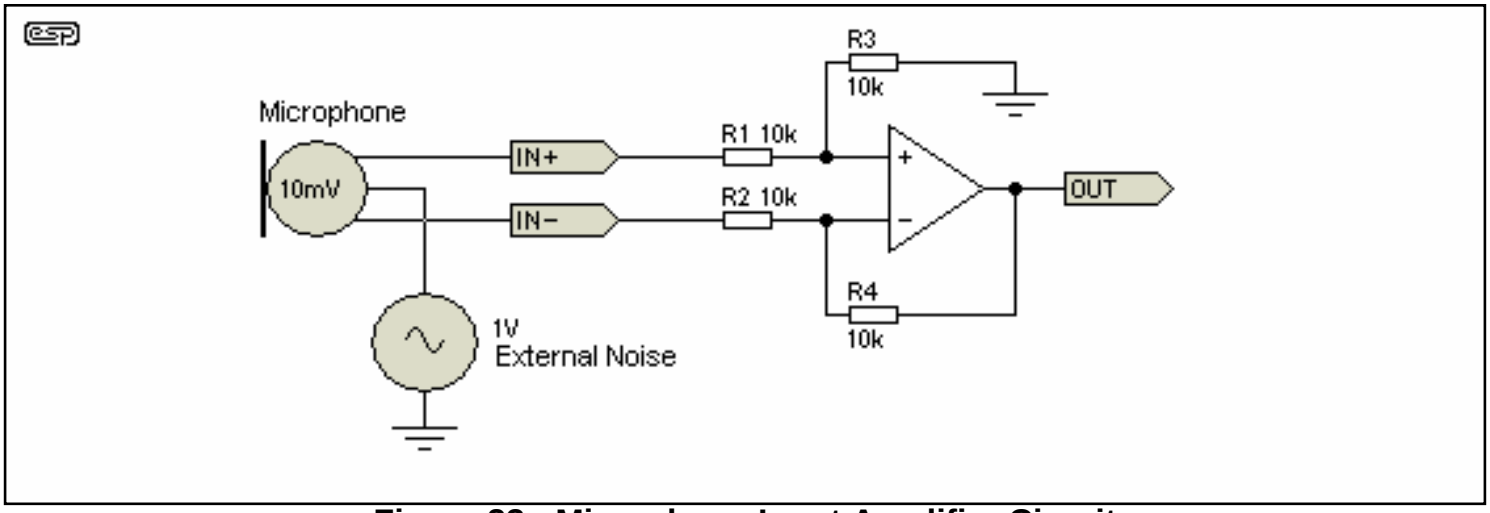


Figure 23 - Microphone Input Amplifier Circuit

Although there are a number of claimed problems with this circuit, in reality it works very well. There are better alternatives (especially for microphones), and these are well represented on the ESP site. The general principle of all balanced circuits remains the same as that shown above.

The biggest limitation of the Figure 23 circuit is that the input impedance for common mode (noise) signals is only equal if the signal applied to each input is equal. This is often claimed to cause major deficiencies in use, but in reality the common mode noise signal usually is very close to equal levels at each input, so the circuit works as described. The greatest limitation of this circuit for microphone use is opamp noise and reduced performance for high frequency common mode signals. For example, at 10kHz, the common mode rejection ratio is 6dB worse than at 100Hz.

Resistor tolerances, cable asymmetry and internal wiring will generally cause more error than the circuit limitations will impose. For optimum CMRR, the resistors should be 0.1% tolerance, and these may be selected from a batch of 1% components.

Balanced Output Circuits

Although these are discussed in depth elsewhere on the ESP site, a brief look at these essential circuits is worthwhile here. Since we have balanced input circuits, it makes sense to have a matching output circuit, allowing equipment to provide a balanced output to other (often remote) gear. The basic circuit shown below is the starting point, and is the basis of all other (often much more complex) circuits.

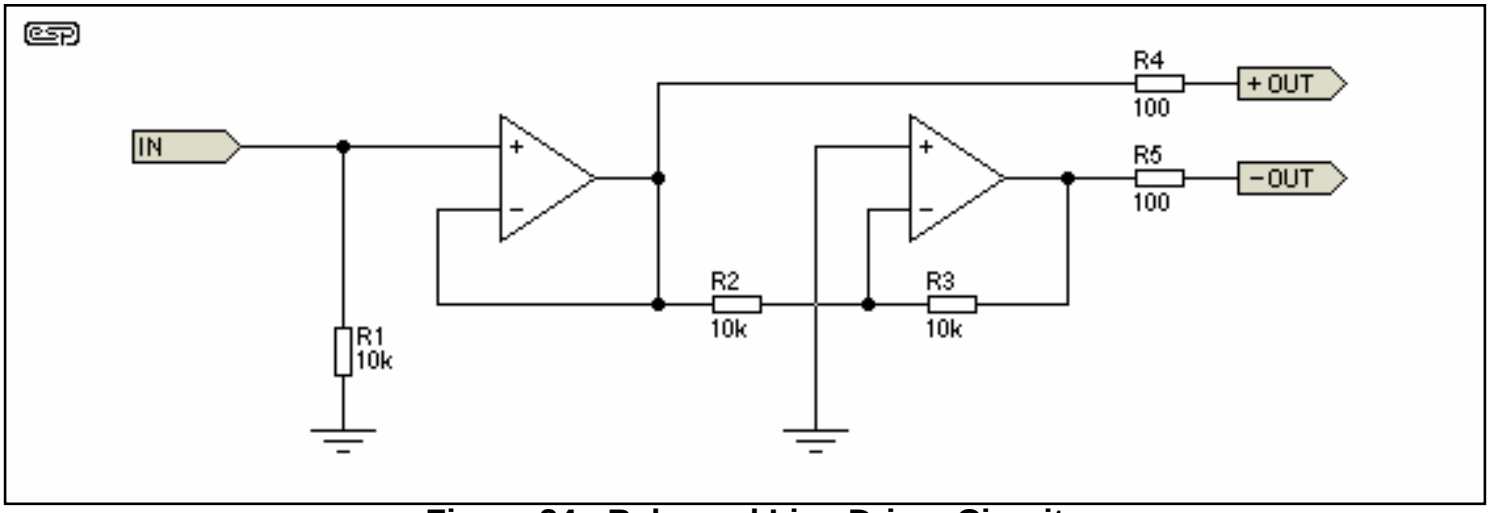


Figure 24 - Balanced Line Driver Circuit

The general idea is quite simple. A single-ended (unbalanced) signal is applied to the input, and it is buffered by the first opamp, and inverted by the second. You should recognise the inverting buffer from Part 1 of this series. The tolerance of the resistors around the inverting stage is again critical, and they should be as closely matched as possible to ensure that the two output signals are exactly the same, but with the signal from the second opamp inverted.

This is a true balanced output, but it has an inferred reference to ground. It behaves like a transformer with a grounded centre tap. While it is possible to approximate a fully floating output, in general it is not necessary to do so. It is usually essential to place a small resistance (typically around 100 ohms) in series with each output to prevent oscillation - R4 and R5. The cable connected to the opamp output acts as an unterminated transmission line at high frequencies, and this can cause the opamps to become unstable because of the reactive load. By including a resistance, the opamp's output is isolated from the reactive load and stability is usually unaffected, regardless of load.

While there is a small error caused by operating two opamps in series (therefore adding the propagation delay of each opamp), the circuit still maintains extremely good balance well above the audio band. Operating the inverting and non-inverting buffers in parallel (the inputs of both joined together), this gives a much lower input impedance and improves performance so marginally that it's not worth doing (IMO).

Summing Amplifiers

Summing amplifiers are based on the inverting buffer. Although we examined this earlier, one aspect of the circuit that was not covered at the time was just how it works. The inverting buffer is also called a virtual earth (ground) circuit, and is very common in mixing

consoles and analogue computers of old. When it was described earlier, we worked in terms of voltage, but the virtual earth amplifier is really a current to voltage converter.

Figure 25 shows the general idea. If a voltage of 1V (AC or DC) is applied to the input, that will cause a current of 0.1mA to flow through R1. Remembering the first Rule of opamps, the opamp itself will attempt to maintain both inputs at the same voltage. Since the positive input is earthed (ground, zero volts), the negative input has to stay at the same potential to satisfy the first Rule. This means the output will have the same voltage as the input, but with the opposite polarity. This is necessary because R2 must also pass exactly 0.1mA to maintain the +ve input at zero volts.

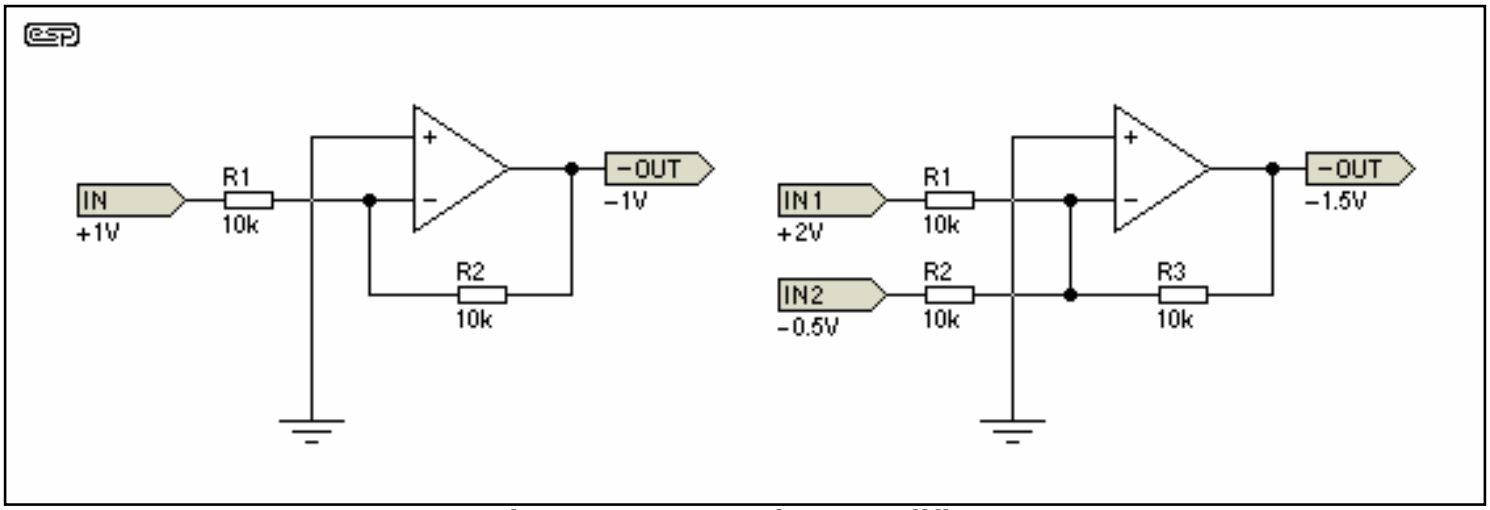


Figure 25 - Summing Amplifiers

By changing the value of R2 (relative to R1), we can modify the gain, making the output voltage larger or smaller in absolute magnitude than the input. It's all done with current - no smoke or mirrors required. Now, if we add another input as shown in the diagram to the right, we can apply another signal, and the opamp will give us a result that is the sum of the two input currents (or voltages, if fixed value resistors are used).

As shown, if one input has an instantaneous voltage of 2V, and the other is -0.5V, the output voltage will be the (inverted) sum of the two - in this case -1.5V. If both inputs were the same voltage and polarity, they are simply added together. At the other extreme, two input voltages of equal magnitude but opposite polarity will result in an output voltage of zero.

There is a drawback to this circuit though, and it is important to understand what happens when you have a large number of inputs. Think back to the inverting amp as originally described. The voltage gain is described by the formula ...

$$Av = R_{FB} / R_{IN}$$

We may decide to use several (let's assume 10 for the moment) inputs, all 10k, and all being fed with voltages from different sources (think in terms of a multi-channel mixer). For each input individually, the voltage gain is as described - i.e. -1 (unity, but inverted). What does the opamp see though? The total value of R_{IN} is 10 x 10k resistors in parallel ... 1k. The opamp therefore acts exactly as if it had a gain of 10, so input transistor noise is multiplied by 10, offset current is multiplied by 10, and bandwidth is reduced accordingly.

A small DC voltage shift caused by input offset current (the difference in current needed by each opamp input), is fine for audio. Any DC error is easily removed by adding a capacitor in series with the output. For instrumentation, the DC value may be critical, and this is why some opamps have 'offset null' pins. The designer can use a pot to adjust the offset to ensure that any DC error is 'nulled' out.

While 10 inputs is not going to cause a major problem in most cases, there is often a need for a great many more - a 32 channel mixer will need to be able to sum 32 channels, so the opamp will have a 'noise gain' of 32, even though the gain for each input individually is -1. Note that there is no polarity for noise gain - noise is random in nature, and not correlated to the input signal. Noise is noise.

The opamp also acts as if it were operating with a gain that is equal to the noise gain, even though each input individually has unity gain. As a result, the bandwidth may be (apparently) inexplicably limited, but by knowing the noise gain, we can treat the circuit as if it has a voltage gain that equals the noise gain - and indeed, this is exactly the case.

The inverting amplifier stage is actually noisier than a non-inverting stage with the same gain. For a non-inverting amplifier, the noise gain is equal to the voltage gain, but with an inverting stage, noise gain is equal to voltage gain + 1. When a large number of inputs is needed, a summing amplifier needs to have very low noise and wide bandwidth, or performance will not be as expected.

A Few More Filters

Because filters are so important in audio, it is necessary to examine a few more variations. Two additional functions will be examined - integration and differentiation. Although Part II did not make it clear, these are the building blocks of some of the most interesting filters that can be made using opamps. A simple integrator (low pass filter) and differentiator (high pass filter) are shown in Figure 26. These are conceptual - the real

world nature of opamps means that neither will work as a 'perfect' device, but must be limited to a defined frequency range. In reality, this is rarely an issue, since ideal versions of the circuits are not needed to cover the audio bandwidth.

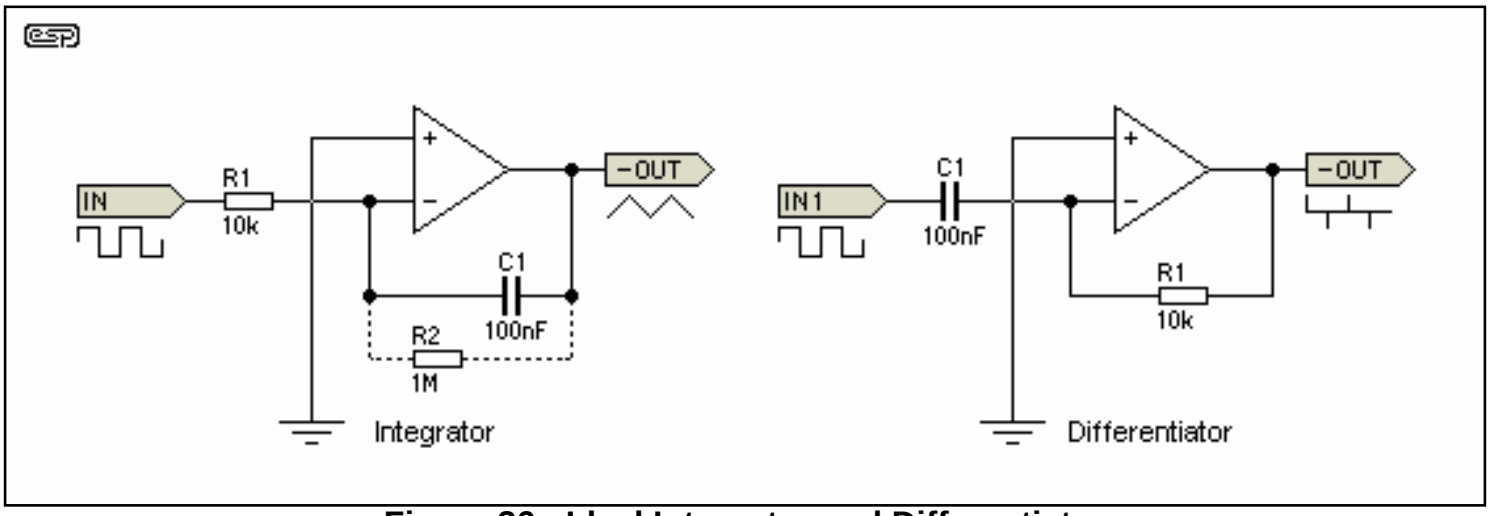


Figure 26 - Ideal Integrator and Differentiator

The differentiator will actually work as shown (albeit with less than ideal performance), but the integrator will not. The reason is simple - it has no DC feedback path, and will drift towards one supply or the other (depending on the opamp characteristics). To combat this, a resistor is needed in parallel with C1 (R2, shown dotted), with a value sufficiently low to provide DC feedback, but not so low as to cause the lowest frequency of interest to be affected. As always, compromises are a part of life.

With an input waveform of a square wave as shown, an integrator provides an output that is dependent on the input current (via R1) and the value of C1. Over a useful frequency range, the output of an integrator is an almost perfect sawtooth waveform with a squarewave input. The DC stability resistor (R2) ultimately limits the lowest usable frequency. While there are methods to get around this limitation, this level of complexity is not normally needed for audio.

The differentiator as shown is also supplied with a squarewave. The output is now based entirely on the rate of change of the input signal, so a squarewave with relatively slow rise and fall times will give a lower output than another with very fast transitions.

The integrator is a low pass filter, with a theoretical 6dB/octave rolloff starting from DC (where it has infinite gain). The 6dB/octave rolloff continues for as far as the opamp characteristics will allow. At an infinite frequency, gain is zero.

A differentiator is the exact opposite of an integrator. It has zero gain at DC, and an infinite gain at an infinite frequency. The filter slope is again 6dB/octave. With ideal

opamps, if the two circuits were connected together the output is a squarewave - identical in all respects to the applied signal. With the values shown, the unity gain frequency is 159Hz (1 / 2 π R1 C1).

To see integrators at work in a real circuit, you need look no further than the Project 48 subwoofer equaliser. Differentiators abound in all audio circuits, but in a rather crude form. Every capacitor that is used for signal coupling is a very basic differentiation circuit, in conjunction with the following load impedance. Note that the frequency where coupling circuits will actually function as a differentiator is well below the lowest frequency in audio.

State Variable Filter

The State Variable filter is probably one of the most interesting of all opamp filters. It uses nested feedback loops and a pair of integrators to define the filter Q, frequency and gain. Gain and Q are not independently adjustable in the version shown - changing one will change the other. To make gain and Q individually adjustable requires the use of two-opamp summing stage (the first opamp). A standardised version is shown in Figure 27, and with the values shown, passband gain is unity, Q is 0.707 and the centre frequency is 159Hz.

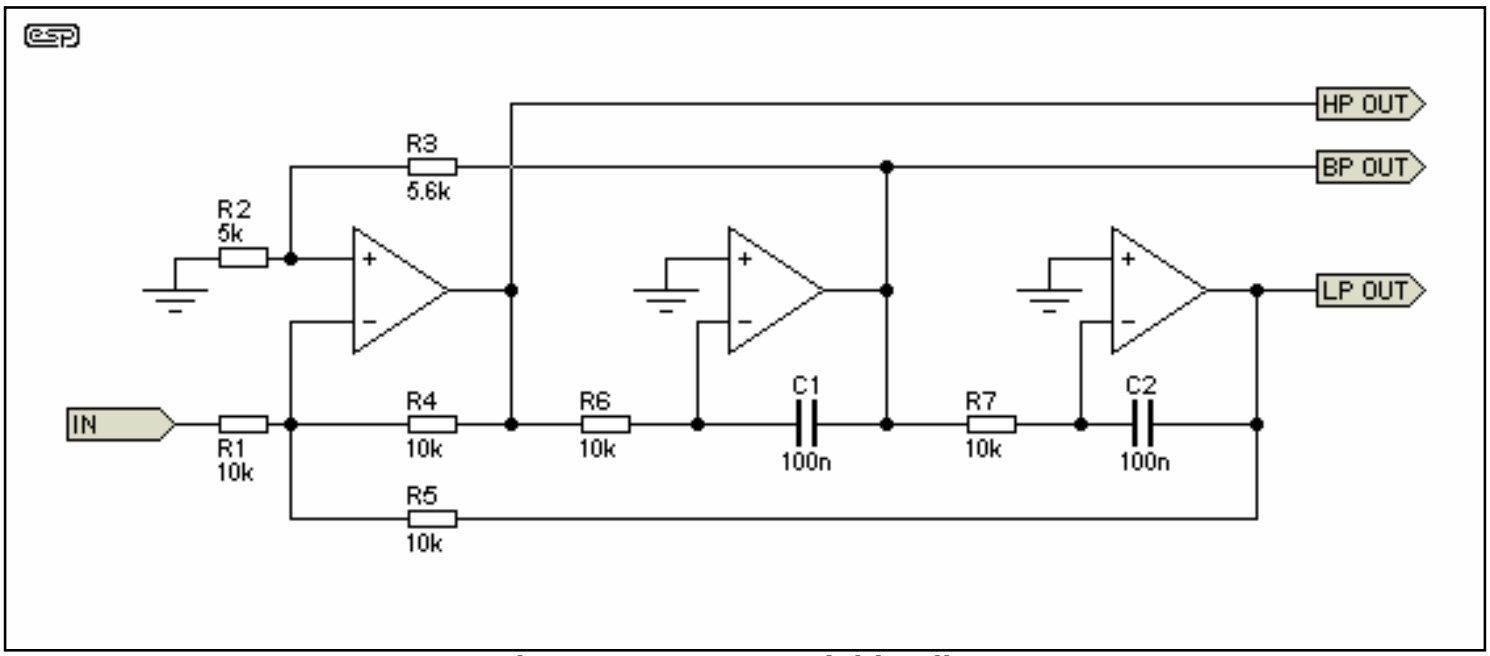


Figure 27 - State Variable Filter

As shown, the filter outputs low pass, band pass and high pass responses simultaneously. R3 sets the filter Q, and is selected as ...

$$R3 = (3 - d / d) * R2 \quad \text{Where } d = \text{damping } (1 / Q)$$

The damping is set at 1.414 in Figure 27, giving a Q of 0.707, so the filter response is

Butterworth (maximally flat amplitude). The frequency is changed by varying R6 and R7, or C1 and C2. The most common use for the state variable filter is in parametric equalisers, where a filter with variable frequency and Q is a requirement. it adding an opamp, the state variable filter allows the gain to be changed without affecting Q, so it is an ideal variable filter for audio use. Filter Q remains constant with frequency, so altering the frequency has no effect on the filter's overall (summed) response. Note that the high and low pass sections have opposite polarities, with the high pass shifted by -90° and low pass shifted +90° at the tuned frequency (f_o). The bandpass output is in phase with the input at f_o

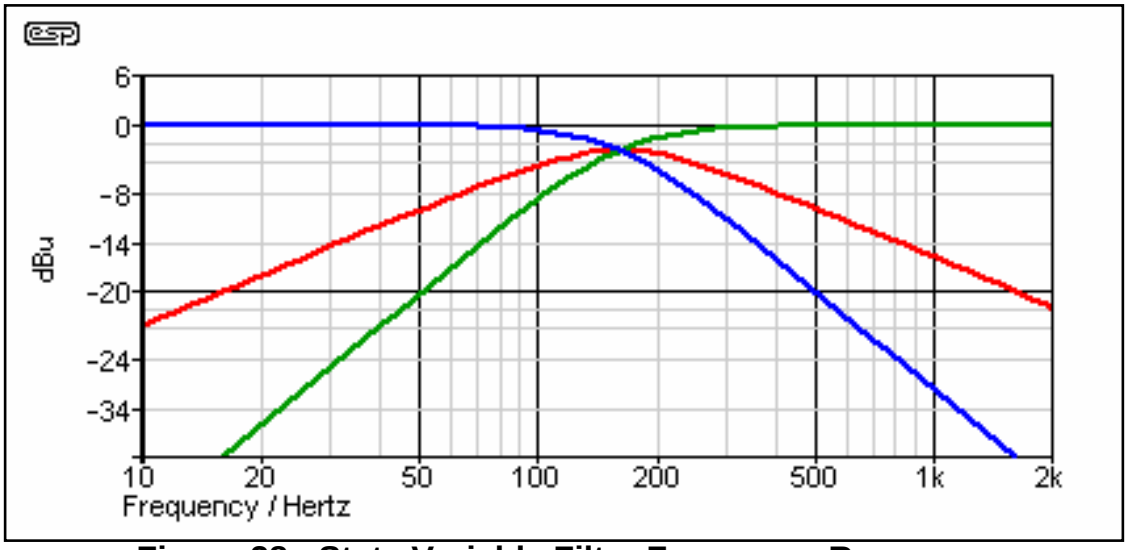


Figure 28 - State Variable Filter Frequency Response

The red trace is the band pass response, the blue trace is low pass, and green is the high pass. Remember that all of these are reproduced simultaneously, thus making the state variable filter on of the most interesting applications available. The standard state variable filter is second order, having rolloff slopes of 12dB/octave. It is also possible to make these filters with a fourth order (24dB/octave) response, but that is beyond the scope of this article.

A variation of the state variable is called a Bi-Quad (bi-quadratic). The difference between the two is subtle. With a bi-quad, as frequency changes, the bandwidth remains constant, which means that the Q must change. As you change the frequency, Q increases as frequency increases and vice versa.

Multiple Feedback Filter

The bandpass version of this class of filter has its own page on the ESP site - see Multiple Feedback Bandpass Filter for more information about this category. There are also low and high pass versions of the MFB filter, and these will be covered briefly here. The chief advantage of the MFB filter is that the opamp's gain bandwidth product (GBP) is relaxed

somewhat compared to the Sallen-Key topology. This is not normally a problem at audio frequencies and for audio applications, because very high Q values are rarely used.

Where the Sallen-Key filter requires a minimum open loop gain of 90Q² at the filter frequency, the MFB version requires only 20Q². To put this into perspective, a Sallen-Key filter with a Q of 0.707 at 20kHz requires an opamp open loop gain of at least 75 at 20kHz, while the equivalent MFB filter only needs a gain of 17 (close enough). While these figures are easy to obtain at low Q values, they become difficult or impossible if a high Q filter is needed.

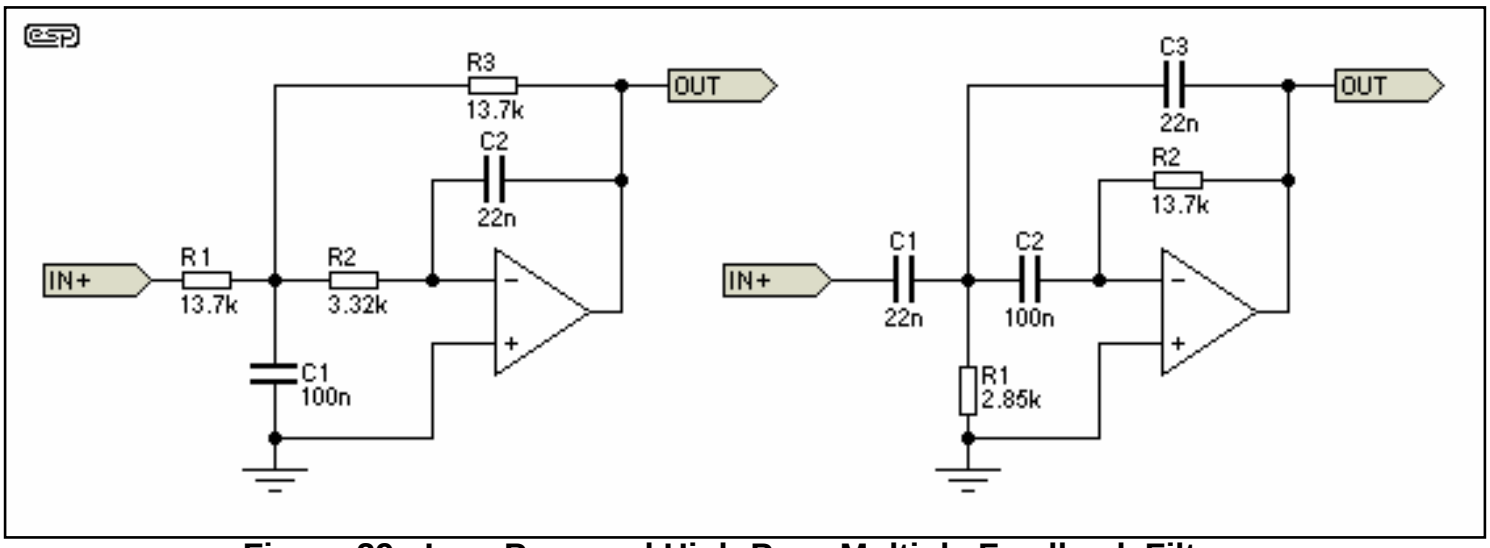


Figure 29 - Low-Pass and High-Pass Multiple Feedback Filters

The first thing to notice with these filters - the resistor (and to a lesser extent the capacitor) values are decidedly non-standard. The design formulae are also rather complex, and I eventually settled on the values shown based on a simulation. The calculations are tedious, and will invariably yield non-standard values. Simple parallel or series connections often cannot be used to get the values you need (based on others in the circuit).

There are several articles on the Web covering the low pass multiple feedback filter, but few that cover the high pass version. As a result, I didn't even try to calculate the values, but figured them out based on the low pass version. As shown, the low pass filter has a cutoff frequency of 503Hz, and the high pass filter has a cutoff frequency of 538Hz. Both have a Q of 0.707 (Butterworth response) and unity gain in the passband.

As filters, they function exactly the same as any other topology with the same cutoff frequency and Q, but as noted above are less demanding of the opamp performance. Whether this ever becomes a problem for most circuits is debatable. Note too that they are inverting, so if absolute phase is your goal, you will need to re-invert the outputs.

Notch Filter Based Bandpass

There is an almost infinite range of filter types, some are useful, others less so. An interesting idea was described in an online design website. The article, entitled "Bandpass filter features adjustable Q and constant maximum gain" shows an active notch filter followed by a difference amplifier (balanced input stage). The circuit detail is reproduced below.

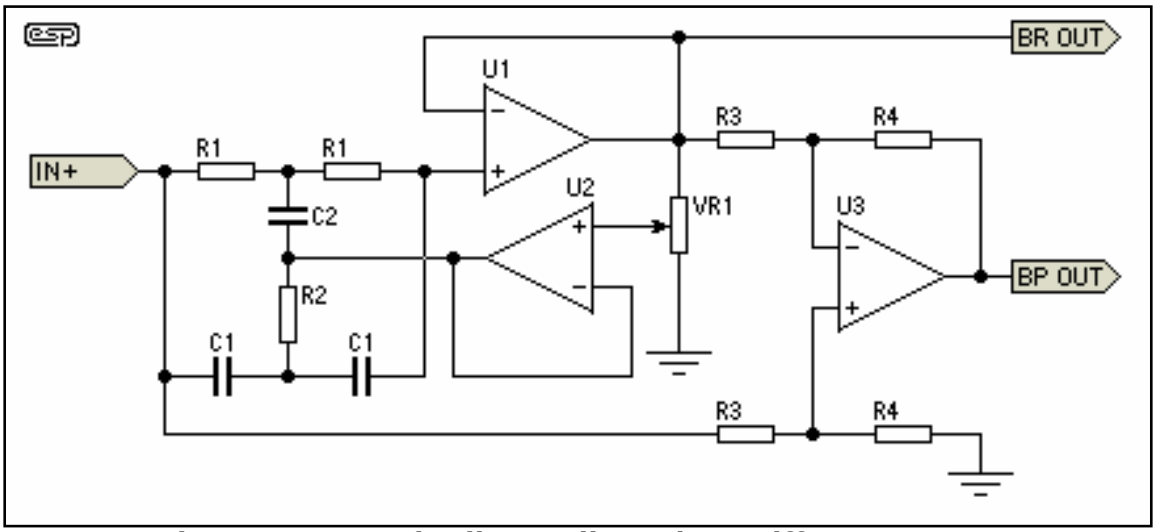


Figure 30 - Notch Filter Followed By Difference Amp

Note that all resistors and capacitors with the same designation are the same value. The 'BP' output is bandpass, and the 'BR' output is band reject (notch).

The notch filter is formed by R1, R2, C1 and C2, where R2 = R1 / 2 and C2 = C1 * 2. The output is buffered by U1, and U2 provides a low impedance feedback path to the notch filter, based on the level set by VR1. U3 is a conventional balanced amplifier, in this case being used as a difference amp. By subtracting the notch signal from the input signal, the result is a bandpass response. The output of the filter can never reach the same Q as the notch however, because phase shift reduces the attenuation of out of band signals.

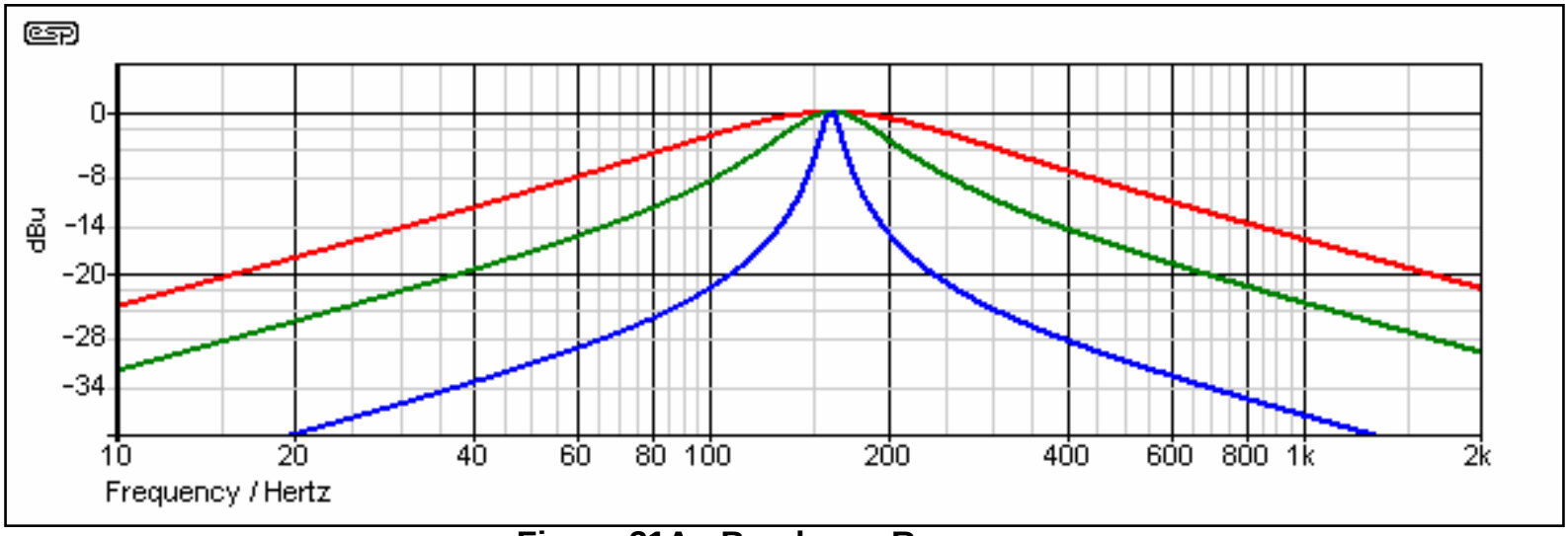


Figure 31A - Bandpass Response

The response with the pot (VR1) set at three levels is shown. The red trace shows the response with VR1 set to 80%, green at 90% and blue at 98%. Useful values of Q (for this type of filter) are only available with close to maximum feedback, but the circuit works as described.

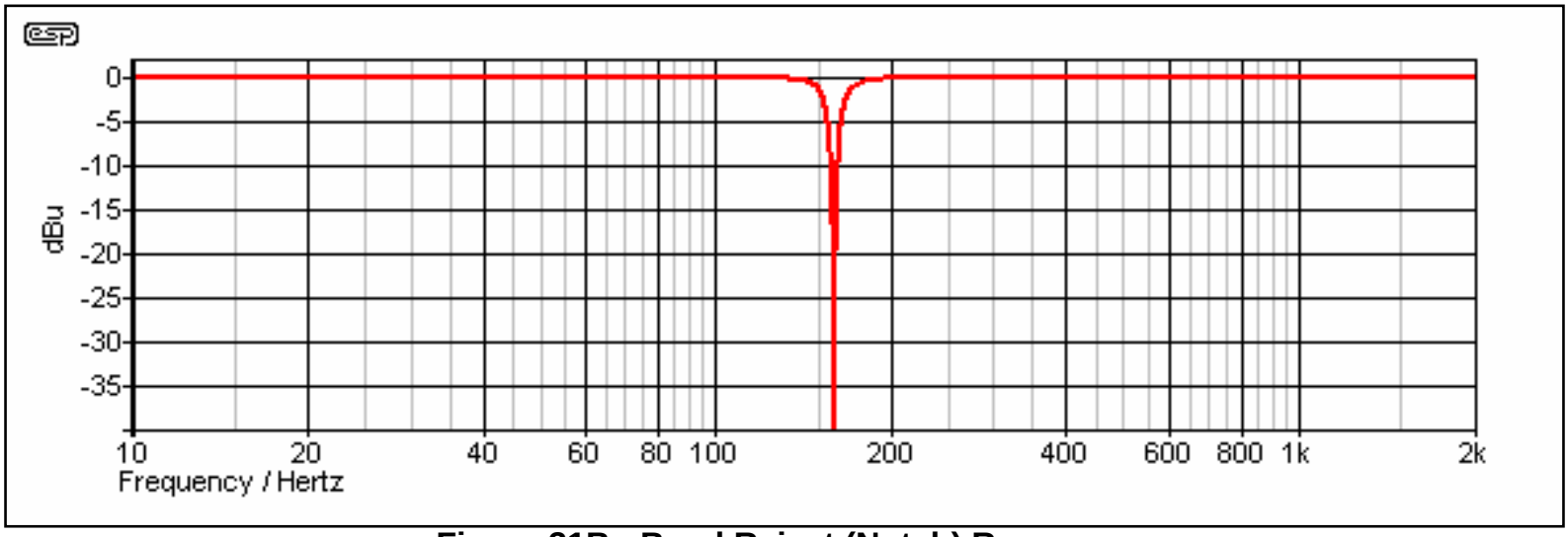


Figure 31B - Band Reject (Notch) Response

The notch filter response (shown in Figure 31B) is as one expects from the twin-tee circuit when feedback is applied. The notch itself is almost infinitely deep, and extremely narrow because of the feedback. The response shown was taken with VR1 set to 98% of its travel - almost maximum feedback applied. Applying more feedback is rather pointless, and the simulator decides that it cannot resolve the notch at all at higher pot settings.

This is another example of the huge range of possibilities for filters - there are obviously a great many more, but space (and possible value) preclude me from going any further in this direction. A web search will reveal many more, and there are also switched capacitor filters, digital filters and probably still others not yet invented.

Basic Power Amplifiers

Opamps have limited output current, usually only around 20mA or so. While this is usually more than enough for preamps and filters, a great deal more is needed to drive a loudspeaker. There are several power amplifiers that are configured in exactly the same way as an opamp, and these can be classified as 'power opamps'. While this is often a good way to get the extra current needed, in some cases it may not be considered appropriate or convenient.

Opamps also have a limited voltage swing, and operating them at the voltages typical for power amplifiers will cause failure. While there are already several opamp based amplifiers on the ESP site (as headphone amps, a small power amp, and two power opamps), the following design is different.

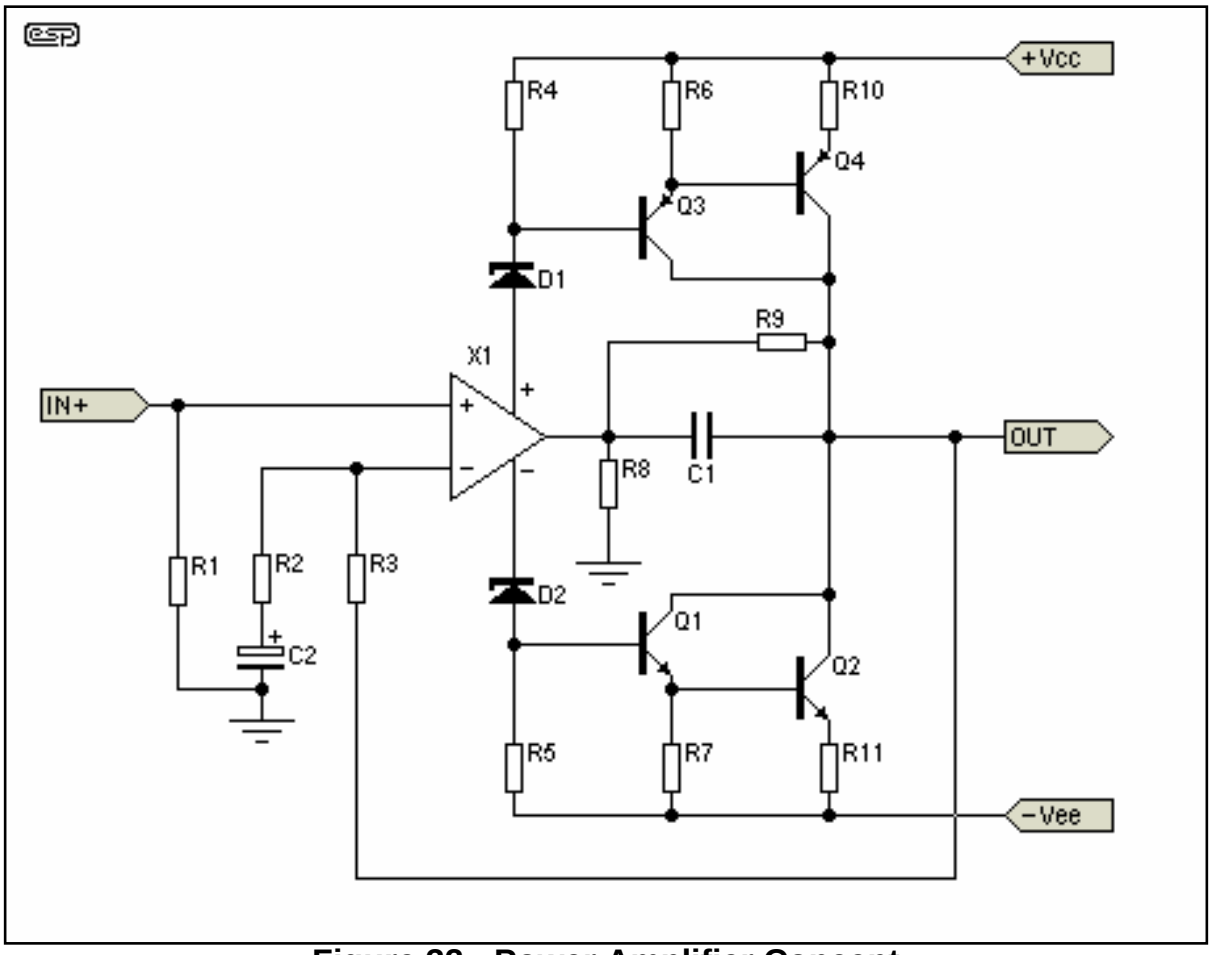


Figure 32 - Power Amplifier Concept

The circuit shown in Figure 32 is conceptional (that's why there are no component values). At one stage, this circuit was quite common, but had some major high frequency issues, as well as DC stability and a host of other problems. Although I do not recommend

that anyone even attempt to build it, the circuit is nonetheless interesting. By using the opamp supply current to modulate the transistors base current, the opamp could operate with voltages well above the normal allowed supply voltages, because the zener diodes reduced the voltage to within the allowable range. R9 was included because many years ago (when I actually contemplated using the design), it was found to work better with this in place.

Although my simulator has difficulties with the circuit, it is possible to get it to run with low output voltages. It shows low distortion, but I know from experience that this is dubious - measured (versus theoretical) distortion is higher than expected, and if I recall correctly, somewhat load dependent. Bias current stability can be very poor. This can lead to thermal runaway in the output stage, and it is very difficult to ensure good thermal stability without additional circuitry.

For other power amplifier designs that *do* work, just look through the ESP projects pages.

Basic Specifications

All in all, the specifications for even a fairly basic opamp can be daunting. There are so many terms used that it is difficult to understand what they all mean. The easiest are the absolute maximum values - these are simply a set of parameters that should never be exceeded. Supply voltage, input voltage, operating and storage temperatures are just a few of the figures quoted.

Let's have a look at the data for a TL072 ...

Abridged electrical characteristics, VCC = ±15 V, T_A = 25°C

Parameter	Test Conditions	Min	Typ	Max	Units
V_{IO} Input offset voltage	V_O = 0, Rs = 50Ω		3	10	mV
αV_{IO} Temperature coefficient of input offset voltage	V_O = 0, Rs = 50Ω		18		uV / °C
I_{IO} Input offset current	V_O=0		5	100	pA
I_B Input bias current	V_O = 0		65	200	pA

V_{ICR} Input common mode voltage		±11 -12 to 15	V
V_{OM} Maximum peak output voltage	R_L = 10kΩ	±12 ±13.5	V
A_{VD} Large signal voltage amplification	V_O = 10V, $R_L \geq$ 2kΩ	25 200	V / mV
B_1 Unity gain bandwidth	V_O = 10V, $R_L \geq$ 2kΩ	3	MHz
r_{IN} Input resistance		10^{12}	MΩ
CMRR Common mode rejection ratio	$V_{IC} = V_{ICRmin}$, V_O = 0, Rs = 50Ω	70 100	dB
k_{SVR} Power supply rejection ratio	$V_{IC} = V_{ICRmin}$, V_O = 0, Rs = 50Ω	70 100	dB
I_{CC} Supply current (each amplifier)	V_O = 0, No load	1.4 2.5	mA
V_{O1}/ V_{O2} Crosstalk attenuation	A_{VD} = 100	120	dB
SR Slew rate at unity gain	V_I = 10V, C_L = 100 pF, R_L = 2 kΩ	5 13	V/us
t_r Rise time overshoot factor	V_I = 20mV, C_L = 100 pF, R_L = 2 kΩ	20%	
V_n Equivalent input noise voltage	R_S = 20Ω, f = 10Hz - 10kHz	4	uV
THD Total harmonic distortion	V_{Irms} = 6V, A_{VD} = 1, $R_L \geq$ 2kΩ, $R_S \leq$ 1k&Omega, f = 1kHz	0.003%	

Well, it certainly does look rather daunting, so we'll look at each parameter in turn to see what it means. It must be understood that there are many different ways to specify an opamp, and the table above is intended to be representative only.

V_{IO} Input Offset Voltage

This is a measure of the typical voltage difference that may exist between the inputs when the first Rule is applied. So, while an ideal opamp will try to make both inputs *exactly* the same voltage, in a real opamp it may differ by this amount. A TL072 connected as a unity

gain buffer with zero input could have between 3 and 10mV between the two inputs (plus or minus). This is measured with zero volts at the output, and a source resistance of 50 ohms.

αV_{IO} Temperature Coefficient of Input Offset Voltage

All real devices are affected by temperature, so the input offset voltage may vary by the amount shown as the operating temperature changes. This figure can be ignored for most audio applications. It is important for instrumentation amplifiers, high gain DC amplifiers and other critical applications.

I_{IO} Input Offset Current

Even though the TL072 is a FET input opamp, it has some input current, primarily caused by gate leakage. The value is very small, having a maximum value of only 100pA.

V_{ICR} Input Common Mode Voltage

This is often an important parameter, and is very much so with the TL07x series. If exceeded, the output state becomes undefined. In the case of the TL07x devices, there can be a change of state of the output voltage if the input common mode voltage is exceeded. A signal that should just clip, has a sudden transition to the opposite polarity during the period where the common mode voltage is exceeded.

V_{OM} Maximum Peak Output Voltage

Few opamps can swing their outputs to the supply rails, and the amount of current drawn from the output affects this further. This limitation is only apparent when using low impedance loads, although there is still some loss with no load at all. By using the highest (sensible) supply voltage, this is not normally a problem.

A_{VD} Large Signal Voltage Amplification

This is usually the open loop (no feedback) condition, and is a measure of the maximum gain of the opamp for high level outputs. At a minimum of 25V / mV, this represents a gain of 25 / 0.001 = 25,000 (it is typically as high as 200,000 or 106dB at low frequencies). This is rarely a limiting value in any audio circuit.

B_1 Unity Gain Bandwidth

The frequency at which the opamp's open loop gain falls to unity is the unity gain bandwidth. When feedback is applied, it is usually desirable to have at least 10 times the gain that you specify with the feedback components. This represents an error of ~10% at the upper frequency. Unity gain bandwidth limits the maximum gain you can use for a given upper frequency.

r_{IN} Input Resistance

Nothing more complex than the input voltage divided by the input current (Ohm's law). So with an input voltage of 10V and an input current of 65pA, the input resistance is 153GΩ. The exact derivation of the value claimed in the table is unclear, since the measurement conditions are unspecified.

CMRR Common Mode Rejection Ratio

A measure of how well the opamp rejects signals applied to both inputs simultaneously. See the description of the balanced input stage for more information. Common mode rejection depends on the available open loop gain, and deteriorates at higher frequencies. This is graphed in most data sheets.

k_{SVR} Power Supply Rejection Ratio

All opamps can have some noise on the supply lines without serious degradation of the signal. PSRR is a measure of how well the opamp rejects (ignores) the supply noise or other unwanted signal(s) that may be carried by the supply buses.

I_{CC} Supply Current (Each Amplifier)

Needless to say, some current is drawn by all opamps. This is simply the typical current you expect to draw from the supply for each opamp in the package. At 1.4mA per opamp, a dual (TL072) will typically draw 2.8mA with 15V supplies.

V_{O1}/ V_{O2} Crosstalk attenuation

When there are two or more opamps in a package, it is inevitable that some signal will pass from one to the other. As specified, this will be -120dB if the amplifiers are operating with a gain of 100. In general, PCB layout causes far more crosstalk than the IC itself.

SR Slew rate at unity gain

The slew rate is simply how fast the output voltage can change. The specification says that this is measured with the opamp connected as a non-inverting unity gain buffer. In this case, the opamps is specified for a typical slew rate of 13V/us, meaning that in one microsecond, the output voltage can change by 13V. There are faster and slower opamps of course, but for audio work it is actually difficult to exceed the slew rate of any but the slowest opamps.

t_r Rise time overshoot factor

When any electronic circuit is subjected to a very sudden change of input voltage, the opamp will often not be fast enough to maintain control via the feedback loop. Once

control is lost, there is a finite time before the opamp 'catches up' to the input signal. This causes the output to overshoot the steady state level for a brief period, and the overshoot is measured as a percentage of the voltage change.

V_n Equivalent input noise voltage

This is a theoretical noise voltage that lives at the opamp's input. This noise is amplified by the gain that is set for the circuit. In this case, if the amp is designed to have a gain of 10, the output noise will be 40uV, over the frequency range of 10Hz to 10kHz. Noise is also expressed as nV√Hz (see separate article).

THD Total Harmonic Distortion

This is pretty much self explanatory. The test conditions are not always representative of real world application, but in this case appear to be reasonably sensible. The use of a unity gain amplifier is not the choice I'd make, but I didn't design the test specification.

The data above is based on National Semiconductor's data sheet and terminology. Other manufacturers may choose to use different terms for the parameters, use different test methods, or specify different operating conditions for the same test. The only way you will get to understand the terminology used is to read it - don't just look at it as gobbledygook and ignore it - you will never learn anything that way.

Comparators

The poor comparator is the (almost) forgotten first cousin of the opamp. Although comparators are very similar to opamps, they are designed based on the fact that they will never have negative feedback applied (although *positive* feedback is not uncommon!). Consequently, there are no constraints set by the necessity for stability with gain reduction caused by applied feedback. Although opamps can be used as comparators in low frequency applications, they are totally unsuitable at high frequencies. This is because of the frequency compensation applied in opamps, necessary to maintain closed loop (negative feedback) stability.

Because this restriction is removed for true comparators, they are able to be much faster than opamps, although it can still sometimes be a challenge to find a comparator that is fast enough if the operating frequency is high. Class-D amplifiers rely on a fast comparator to convert the analogue input signal into a pulse width modulated switching signal, and they need to be very fast indeed if timing errors are to be avoided.

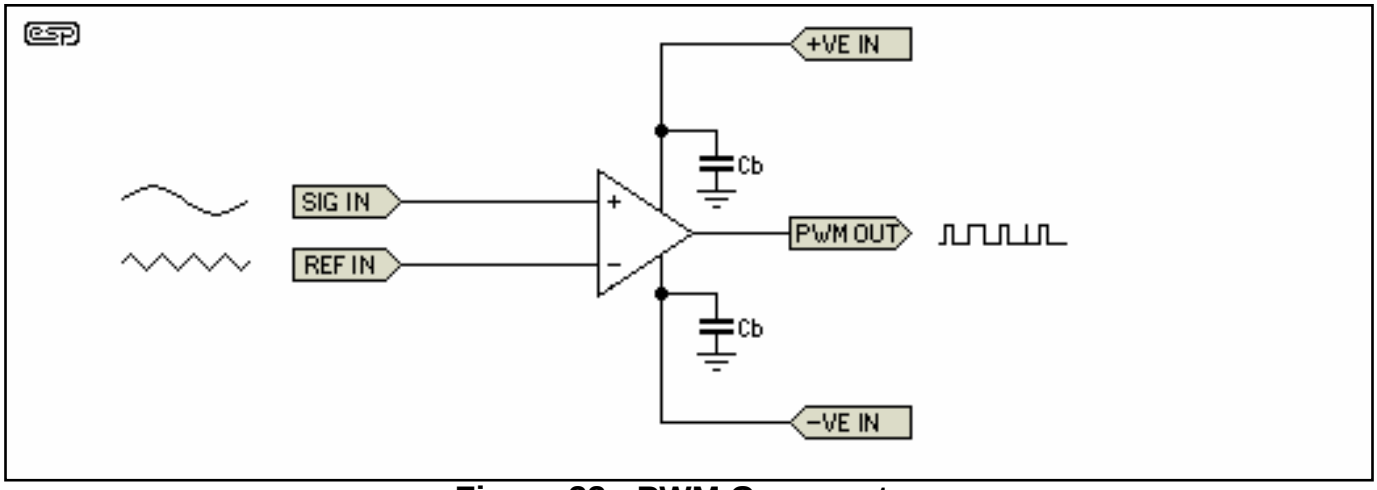


Figure 33 - PWM Comparator

Figure 33 shows the general idea for a pulse width modulator. The input sinewave is compared to the reference signal - usually a very linear sawtooth waveform. Since typical PWM amplifiers operate at a switching frequency of 250kHz or more, even a few nano-seconds switching delay becomes significant, but there are other applications where even faster operation is desirable.

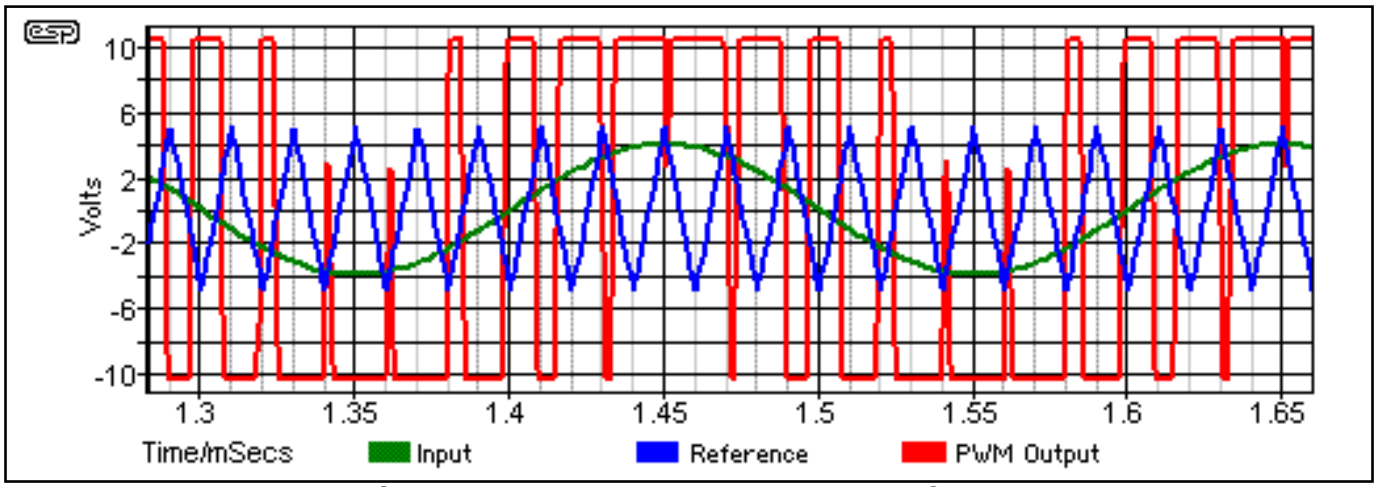


Figure 34 - PWM Comparator Waveforms

The PWM waveforms are shown above. A perfect example of the speed limitation is visible as the input signal approaches the peak value of the reference signal. The comparator is simply not fast enough to switch from one extreme to the other, resulting in the loss of pulses. This is a real phenomenon, and occurs with all PWM amplifiers as they approach clipping.

Achieving very high speed is a compromise, because high speed means that all circuit impedances must be low, thus increasing supply current. This means that the parts will get hot, unless the supply voltage is limited. For exactly the same reasons, microprocessor ICs are now using 3.3V instead of 5V, and high speed digital logic chips are all limited to 5V supplies. We can expect to see the voltage decrease even further as

the speed of digital systems increases.

Needless to say, PWM amps are not the only place comparators are used in audio. LED meter ICs have a string of comparators, they may be used for basic (or precision) timing applications, clipping detectors, etc. All analogue to digital converters use comparators as well.

Because they have no feedback, comparators *always* obey the second rule of opamps ... the output takes the polarity of the more positive input. Look carefully at the PWM waveform shown above, and you will see exactly that. Look closely - it is not immediately apparent, but it is visible.

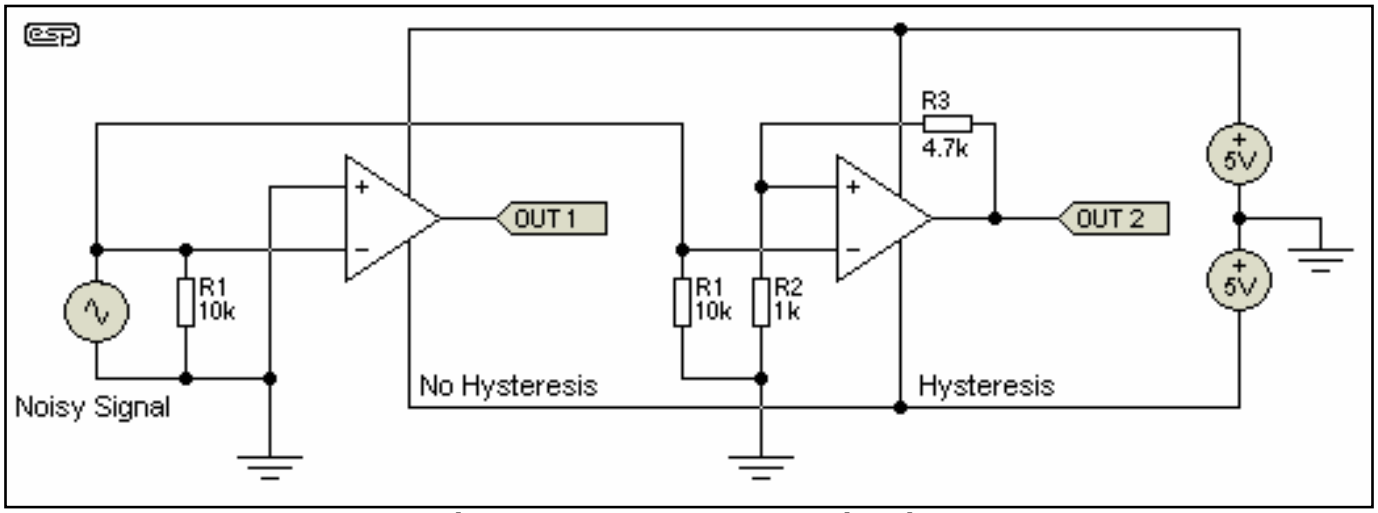


Figure 35 - Comparator Circuits

Comparators can be absolute, meaning that the output will change state whenever the signal passes the threshold. While this is needed for analogue to digital conversion and many other applications, it is often preferable to arrange the circuit so it has 'hysteresis'. This is a rather odd concept, but means that once the signal has caused an output change, it needs to change further to cause a reversal. Hysteresis can be likened to the snap-action of most switches, and indeed, this is mechanical hysteresis.

Both circuits are shown above. The waveforms are shown below. Note that both comparators are shown as inverting, because this connection provides the highest input impedance when hysteresis is added. If a non-inverting connection is used the input would be applied to the +ve input, via R2 for the version with hysteresis, and the -ve input connected to the reference voltage - in both cases this is ground (zero Volts). In a non-inverting connection, the positive feedback will create distortion on the input waveform - this can be a problem if the signal is intended to be used as an analogue waveform elsewhere in the circuit.

The polarity of any comparator can be reversed with a digital logic inverter, another comparator, or the input can be buffered to prevent positive feedback artefacts from being added to the signal. As always, there are many, many ways to achieve the same result, and the final circuit depends on your specific needs.

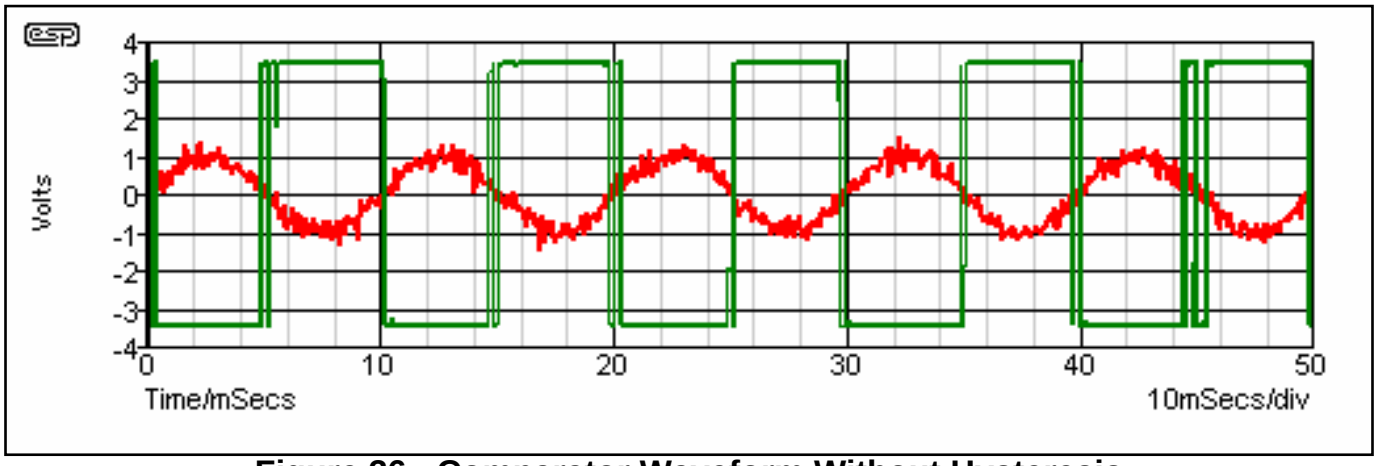


Figure 36 - Comparator Waveform Without Hysteresis

As you can see from the above diagram, if a noisy signal is applied to the input of an absolute (no hysteresis) comparator, the output shows multiple state changes as the input signal passes through zero. This happens because the noise amplitude is enough to cause the instantaneous input amplitude to pass through zero multiple times. This is usually undesirable, because comparators are commonly used to convert input waveforms into a digital representation, based on zero crossings of the input. Multiple triggerings as shown will cause an erroneous output.

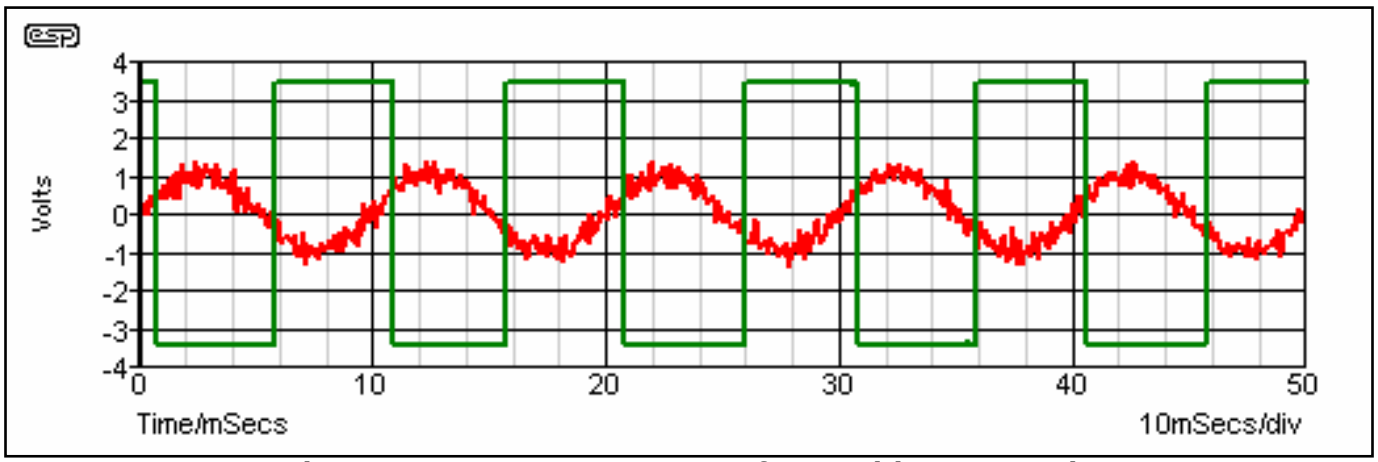


Figure 37 - Comparator Waveform With Hysteresis

Figure 37 shows the input and output waveforms of the comparator with hysteresis. The noise is ignored, because once the comparator changes state, the signal must go negative by more than the hysteresis voltage before it will change state again. The amount of hysteresis is determined by R2 and R3, and may be calculated to give a specific (and exact) level before the output will change state. This circuit is also known as

a Schmidt Trigger, and is a very common circuit. It is used in the PCB version of P39 (power transformer soft-start) to ensure accurate timing without any possibility of relay chatter as the timing voltage reaches the trigger point.

Calculating the trigger voltages for the inverting case is easy (see below), but is somewhat more irksome for the non-inverting configuration because the input and output voltages interact. Only the inverting is described here. R2 and R3 form a simple voltage divider. When the output is high (+3.5V), the voltage at the +ve input is ...

Vd = R3 / R2 + 1 = 4.7 / 1 +1 = 5.7 (where Vd is voltage division)
Vin = Vout / Vd = 3.5 / 5.7 = 614mV

The signal therefore must exceed +614mV before the output will swing negative. When it does, the input then has to exceed -614mV before the output will change state again, because the circuit is symmetrical. Any signal that does not reach the ±614mV thresholds will not cause the output to change state - such signals are completely ignored.

It is also possible (and not uncommon) to make the trigger voltages asymmetrical, and I shall leave this as an exercise for the reader to work out how this can be done. (Hint - diodes are commonly used to do just this).

This is but a small sample of comparator applications. They may well be the almost forgotten first cousin of opamps for the beginner or novice, but are extensively used in all kinds of circuitry - not necessarily audio, but they are very common there too. Most comparator applications are easy enough to understand once you have the basics, and this section is intended to provide just that - the basics.

Conclusion

Although the three articles in this series have only scratched the surface, hopefully you will have sufficient information as to how opamps work to be able to analyse any new circuit you come across. There is no doubt that there will be some applications that will cause pain, and it is completely impossible to make it otherwise.

Even though there are three fairly large pages devoted to the topic here, there are countless other applications for opamps - not only in audio, but in instrumentation, medical applications, and any number of industrial processing systems. There is almost no analogue application these days that does not use opamps, although in many cases you may not be aware they are there. Opamps are embedded in many other devices,

from analogue to digital converters and their opposite numbers, digital signal processing ICs, switchmode power supply controllers - the list is endless.

Part 1 Part 2

References

I have used various references while compiling this article, with most coming from my own accumulated knowledge. Some of this accumulated knowledge is directly due to the following publications:

National Semiconductor Linear Applications (I and II), published by National Semiconductor
National Semiconductor Audio Handbook, published by National Semiconductor
IC Op-Amp Cookbook - Walter G Jung (1974), published by Howard W Sams & Co., Inc. ISBN 0-672-20969-1
Active Filter Cookbook - Don Lancaster (1979), published by Howard W Sams & Co., Inc. ISBN 0-672-21168-8
Data sheets from National Semiconductor, Texas Instruments, Burr-Brown, Analog Devices, Philips and many others.
The TL07x data sheets from National Semiconductor was extensively referenced in the basic specifications section.
Cirrus Logic, Application Note 48
Bandpass filter features adjustable Q and constant maximum gain - EDN

Recommended Reading

Opamps For Everyone - by Ron Mancini, Editor in Chief, Texas Instruments, Sep 2001

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) Rod Elliott, 02 Mar 2006./ Published 11 May 2006

The Audio Pages

Elliott Sound Products **Components - Part II**

Copyright © 2004 - Rod Elliott (ESP)
Page Created 15 Feb 2004

Articles Index

Main Index

Contents

- 1 - Introduction
- 2 - E12, 24, 48 & E96 Series
- 3 - Capacitors
- 4 - Zener Diode Listing
- 5 - Conclusion

---

1 - Introduction

The EIA (Electronic Industries Association) and other authorities worldwide specify standard values for resistors, commonly referred to as the "preferred value" system. This system has its origins in the early period of electronics, at a time when most resistors were carbon composition with poor manufacturing tolerances. The idea is simple - select values for components based on the tolerances to which they are able to be made.

For example, based on 10% tolerance devices, we will work with a preferred value of 100Ω. It makes no sense to produce a 105Ω resistor, because 105 ohms falls within the 10% tolerance range of the 100Ω resistor. The next reasonable value is 120Ω, because a 10% 100Ω resistor will have a value somewhere between 90 and 110 ohms.

A 10% 120Ω resistor has a value ranging between 108 and 132Ω. Following this logic, the preferred values for 10% tolerance resistors between 100 and 1,000Ω is a roughly logarithmic sequence of 100, 120, 150, 180, 220, 270, 330 and so on (rounded to the closest sensible value). This is the E12 series shown in the first table below. The values in any decade can be derived by multiplying or dividing the table entries by powers of 10.

In each series, values may start from as low as 0.1 ohm, and may extend to several megohms, depending on the type of resistor and its intended purpose (and cost, of course).

The highest number of values is the E192 series, but these are normally only required where extreme accuracy is needed. For the odd occasion where a highly specific resistance is needed, it is usually simpler to use 2 or more resistors in series or parallel to obtain the needed resistance. Few suppliers stock the E96 or E192 series, so they will be difficult for most people to obtain.

---

2 - E12, 24, 48 & E96 Series

E12	1.0	1.2	1.5	1.8	2.2	2.7	3.3	3.9	4.7	5.6	6.8	8.2

12 Values per Decade

E24	1.0	1.1	1.2	1.3	1.5	1.6	1.8	2.0	2.2	2.4	2.7	3.0
	3.3	3.6	3.9	4.3	4.7	5.1	5.6	6.2	6.8	7.5	8.2	9.1

24 Values per Decade

E48	1.00	1.05	1.10	1.15	1.21	1.27	1.33	1.40	1.47	1.54	1.62	1.69
	1.78	1.87	1.96	2.05	2.15	2.26	2.37	2.49	2.61	2.74	2.87	3.01
	3.16	3.32	3.48	3.65	3.83	4.02	4.22	4.42	4.64	4.87	5.11	5.36
	5.62	5.90	6.19	6.49	6.81	7.15	7.50	7.87	8.25	8.66	9.09	9.53

48 Values per Decade

E96	1.00	1.02	1.05	1.07	1.10	1.13	1.15	1.18	1.21	1.24	1.27	1.30

1.33	1.37	1.40	1.43	1.47	1.50	1.54	1.58	1.62	1.65	1.69	1.74
1.78	1.82	1.87	1.91	1.96	2.00	2.05	2.10	2.15	2.21	2.26	2.32
2.37	2.43	2.49	2.55	2.61	2.67	2.74	2.80	2.87	2.94	3.01	3.09
3.16	3.24	3.32	3.40	3.48	3.57	3.65	3.74	3.83	3.92	4.02	4.12
4.22	4.32	4.42	4.53	4.64	4.75	4.87	4.99	5.11	5.23	5.36	5.49
5.62	5.76	5.90	6.04	6.19	6.34	6.49	6.65	6.81	6.98	7.15	7.32
7.50	7.68	7.87	8.06	8.25	8.45	8.66	8.87	9.09	9.31	9.53	9.76

96 Values per Decade

---

3 - Capacitors

Film capacitor values normally follow the E12 series. In some cases, suppliers will decide (based on what criteria I really don't know) that some of the available values are "not needed", and they can be hard to find.

Electrolytics usually follow a limited range of the E12 series, and in the case of larger types, may not follow any particular series at all. For example, 8,000uF caps are quite common, but don't fit into any of the above tables.

Electrolytics also have rather broad claimed tolerance (up to +20% -50%), but in reality, most are remarkably close to the marked value.

Although the E6 series was used with resistors (or so I believe - I've never seen it), it is still common with electrolytic capacitors. Although the range seems very limited, it is normally quite sufficient for the typical uses of electros - power supply decoupling, input caps, etc.

E6	1.0	1.5	2.2	3.3	4.7	6.8

6 Values per Decade

---

4 - Zener Diodes

The following is far from a complete listing, but gives a reasonable range of voltages and power dissipation. Personally, I prefer the European designations, such as BZV85C6V8

- you can tell instantly that the voltage is 6.8V from the number. Unfortunately, these are not always easy to obtain, and the 1N series are more common (in non-European countries at least). Please note that due to the amount of data in the table, it is almost a certainty that I have made one (or more) mistakes in the translation, so *always* check the data sheet before committing yourself to a particular device. Also bear in mind that many of the devices listed will be extremely difficult to get. 1W zeners are the most commonly available, and a method is shown below to use these at much higher power levels.

	Power (Watts)							
Voltage	**0.25**	**0.4**	**0.5**	**1.0**	**1.5**	**5.0**	**10.0**	**50.0**
1.8	1N4614							
2.0	1N4615							
2.2	1N4616							
2.4	1N4617	1N4370						
2.7	1N4618	1N4370						
3.0	1N4619	1N4372	1N5987					
3.3	1N4620	1N5518	1N5988	1N4728	1N5913	1N5333		
3.6	1N4621	1N5519	1N5989	1N4729	1N5914	1N5334		
3.9	1N4622	1N5520	1N5844	1N4730	1N5915	1N5335	1N3993	1N4549
4.7	1N4624	1N5522	1N5846	1N4732	1N5917	1N5337	1N3995	1N4551
5.6	1N4626	1N5524	1N5848	1N4734	1N5919	1N5339	1N3997	1N4553
6.2	1N4627	1N5525	1N5850	1N4735		1N5341		1N4553
7.5	1N4100	1N5527	1N5997	1N4737	1N3786	1N5343	1N4000	1N4556
10.0	1N4104	1N5531	1N6000	1N4740	1N3789	1N5347	1N2974	1N2808
12.0	1N4106	1N5532	1N6002	1N4742	1N3791	1N5349	1N2976	1N2810
14.0	1N4108	1N5534	1N5860			1N5351	1N2978	1N2812
16.0	1N4110	1N5536	1N5862	1N4745	1N3794	1N5353	1N2980	1N2814
20	1N4114	1N5540	1N5866	1N4747	1N3796	1N5357	1N2984	1N2818
24	1N4116	1N5542	1N6009	1N4749	1N3798	1N5359	1N2986	1N2820
28	1N4119	1N5544	1N5871			1N5362		
60	1N4128		1N5264			1N5371		
100	1N4135	1N985		1N4764	1N3813	1N5378	1N3005	

120		1N987	1N6026	1N3046	1N5951	1N5380	1N3008	1N2841
Voltage	**0.25W**	**0.4W**	**0.5W**	**1.0W**	**1.5W**	**5.0W**	**10.0W**	**50.0W**

Voltages that do not appear are easily made up by connecting zeners in series. If possible, keep the voltages of the two (or more) zeners as close as possible, or their current handling capabilities will be different, possibly leading to overheating of the higher voltage (and therefore higher dissipation) device(s).

Zeners should always be operated at between 10% and up to a maximum of 80% of rated power to obtain the best (most stable) reference voltage. To determine the optimum current (say 25% of maximum to keep dissipation reasonable), use the following simple formula ...

I = (P / V) / 4

where I is current, V is rated voltage, and P is rated power. For example, a 27V 1W zener should be operated at around

I = (1 / 27) / 4 = 0.00926A = 9.26mA (power is 250mW)

In many cases, the preferred current will be too high and will cause excessive heating, but it must be understood that the regulation of the zener is not very good until it is operating at the lower limit of 10% rated power.

There will be times when you really do need a high power zener, but find that the one you need is either not available or very expensive. There is generally no real need to use high powered zeners unless space is at a premium, since the simple circuit below will work in most cases.

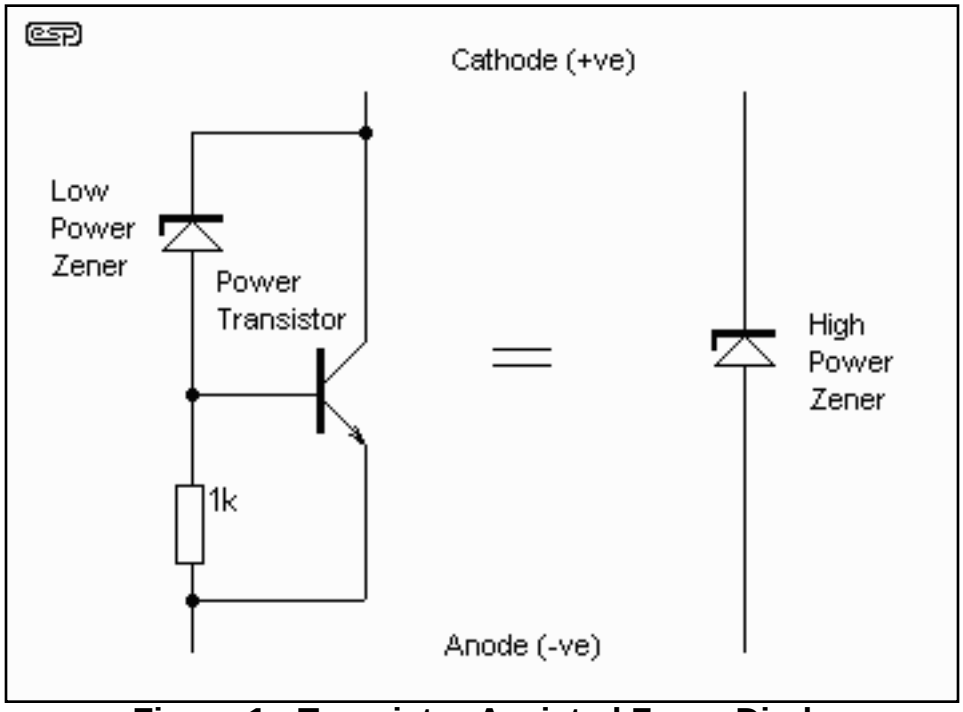


Figure 1 - Transistor Assisted Zener Diode

The circuit works by simply amplifying the zener current, the majority of which is fed directly into the base of the transistor. If the voltage attempts to rise, more zener current flows, thus more base current to the transistor. This causes the transistor to turn on until a state of equilibrium is reached, where the voltage across the 'composite zener' is held at the correct value (+ 0.65V base-emitter voltage, of course).

Since zeners are typically classified by their dynamic resistance (or impedance), it is worthwhile looking at the assisted version to see what sort of performance we can expect. Figure 2 shows the dynamic impedance of a zener by itself, and that of the assisted version.

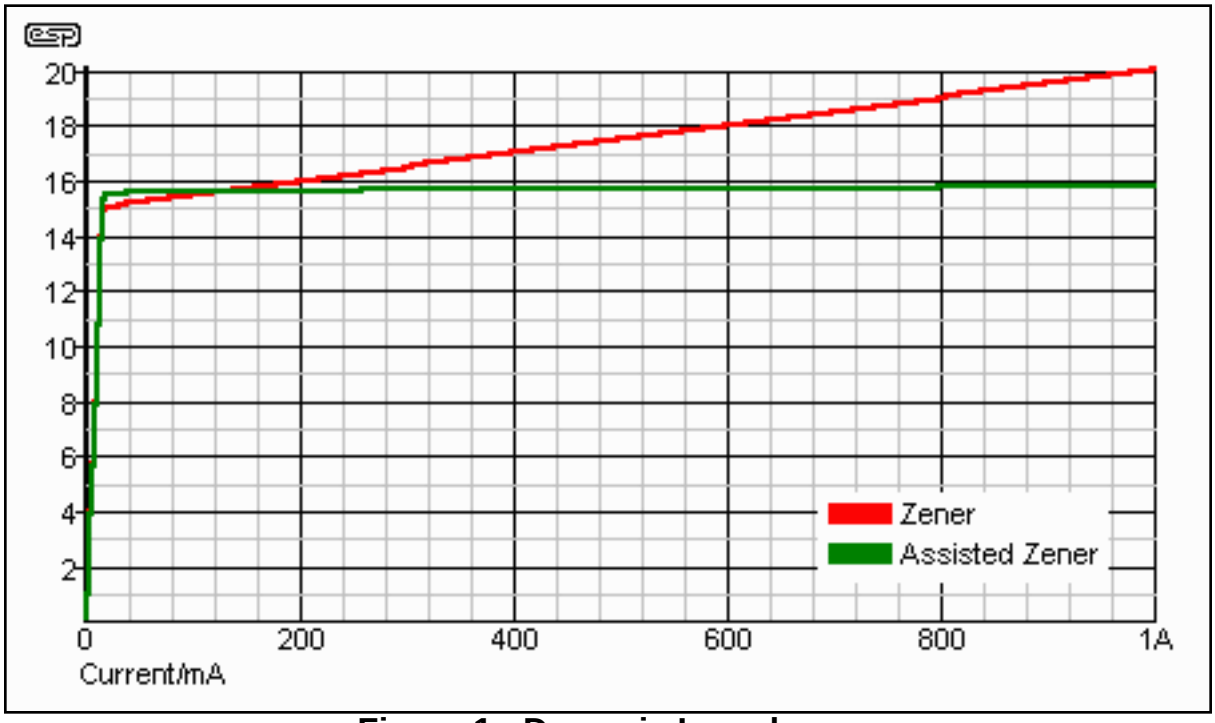


Figure 1 - Dynamic Impedance

The dynamic impedance is measured by changing the current by a known amount, and measuring the voltage change. In the case of the zener, the current was varied by 100mA, and the voltage changed by 0.4975V, therefore ...

$$R = V / I = 0.4975 / 0.1 = 4.975\Omega$$

The transistor assisted zener is a great deal better (as can be seen from the flatter curve). In this case, a current change of 100mA only achieved a voltage change of 25mV, so dynamic resistance/impedance is ...

$$R = V / I = 0.025 / 0.1 = 0.25\Omega$$

It is possible to improve this further, but I'd suggest that if you need better performance than an assisted zener can provide, then you'd be a lot better off with a proper regulator.

Note that the transistor must be chosen so that it is operating in its continuous safe operating area, and must be appropriately derated for temperature. Expect a 100W transistor assisted zener (for example) to dissipate *a lot* of heat, so don't skimp on the heatsink.

For those who have looked at a lot of the ESP site, you will notice that the P37 (DoZ preamp) power supply uses an assisted zener as the main power supply regulator.

5 - Conclusion

The main purpose of this section is to tie up a few "loose ends", and add some of the harder to find (or just plain frustrating) information that you will need from time to time.

It is probable that more loose ends will turn up in time, and this will be added to the information here as found or needed. Unfortunately (or perhaps fortunately) the range of components is so diverse that it is not possible to cover everything.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 15 Feb 2004

The Audio Pages

Elliott Sound Products **Meters, Multipliers & Shunts**

Copyright © 2006 - Rod Elliott (ESP)
Page Published 06 May 2006

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Basic Meter Movements
  - 2.1 Movement Specifications
- 3.0 Voltage Multiplier
- 4.0 Current Shunt
- 5.0 Expanded Scale Voltmeter
- 6.0 Digital Panel Meters
  - 6.1 Digital Voltmeters
  - 6.2 Digital Ammeters
- 7.0 Make Your Own Multimeter
  - 7.1 Measuring Resistance
- 8.0 Conclusion
- 9.0 References

1.0 - Introduction

The moving coil meter movement (also known as a galvanometer) was invented by the French physicist and physician, Jacques-Arsène D' Arsonval in 1882. It is the basis for all modern meter movements, and the basic design principles remain the same after all this

time. The actual construction can differ quite widely, but upon examination it is obvious that there are simply different ways to achieve the same outcome.

Meters are common in audio. They are sometimes used as 'eye candy' to impress - especially on power amplifiers, but they have many real uses as well. Meters are used to display the level from mixing desks, either as a VU (volume unit) or PPM (Peak Programme Meter) display, and while LED meters save space and can be very fast acting, they have neither the coolness of an analogue movement nor the retro appeal. To many people, an analogue movement provides a better sense of what is happening, even though they lack the immediacy of a LED display. In some cases, the two may even be combined to give the best of both worlds.

Meters are also used on power supplies and many other pieces of test equipment, and although it is assumed that digital is more accurate (you can see the exact voltage displayed), this is not always the case. Although digital meters *appear* accurate, this is often an illusion (read the specifications ... 1% ±1 digit is common, and that last digit can make a big difference sometimes).

In addition, there are some applications where digital is essentially useless. If a voltage (or current) is continually changing, the readout from a digital meter is impossible to interpret accurately. With analogue, you can see peaks and dips, and it is easy to see a trend (or average) just by looking at the pointer. Analogue is far from dead, and to this day I still use many analogue meters on millivolt meters, distortion analysers, power supplies, etc.

Although many of the techniques shown in this article are aimed at analogue applications, they are equally at home with digital meters - DPMs (Digital Panel Meters) are commonly available for about the same price as their analogue counterparts. This makes them very attractive for some applications - especially since good moving coil meter movements are now quite expensive and may be hard to get. Some applications are also shown for DPMs.

2.0 - Basic Meter Movements

The basic analogue meter movement is the moving coil type. These have been the mainstay of most metering applications for a very long time, but there are others that are common in other industries. Moving iron meters are often used for mains applications (especially in switchboards and the like), and although they are non-linear this is not a limitation for the intended applications. The latter are interesting, but will not be covered

because of limited availability and lack of usefulness for audio applications. Another interesting meter uses electrostatics to display the voltage. These are restricted to very high voltage applications and apply virtually no circuit loading. Like the moving iron movements, they are not useful for general workshop use because they are too specialised. A photo of a very ordinary moving coil meter movement is shown in Figure 1.

Figure 1 - Moving Coil Meter Movement

Figure 2 shows the essential sections - yes, it is different from Figure 1. The drawing shows the way that moving coil movements were commonly constructed many years ago, which is somewhat easier to draw than more modern types. The essential parts are labelled so you get an idea of the construction of these meters. Nearly all moving coil meters are low voltage, low current devices, and the multipliers and shunts referred to in the title are used to convert the movement to read higher voltages and currents than it was designed for. This versatility is the reason that moving coil meters have stayed with us for so long. They can be made to read up to thousands of volts (or amps), AC voltage and current (with the addition of rectifiers), audio levels, or anything else where a physical quantity can be converted to an electric current.

The beauty of the analogue scale is that a plant operator (for example) can tell at a glance if the reading is normal, whereas it is necessary to actually read the displayed value of a digital meter. You don't need to read a value on an analogue meter to see if it

is normal. Look at the meter on a battery tester - it is simply labelled 'Replace' and 'Good' or similar - the exact value is unimportant, but you still see a linear scale so you can estimate 'Marginal' without even thinking about it.

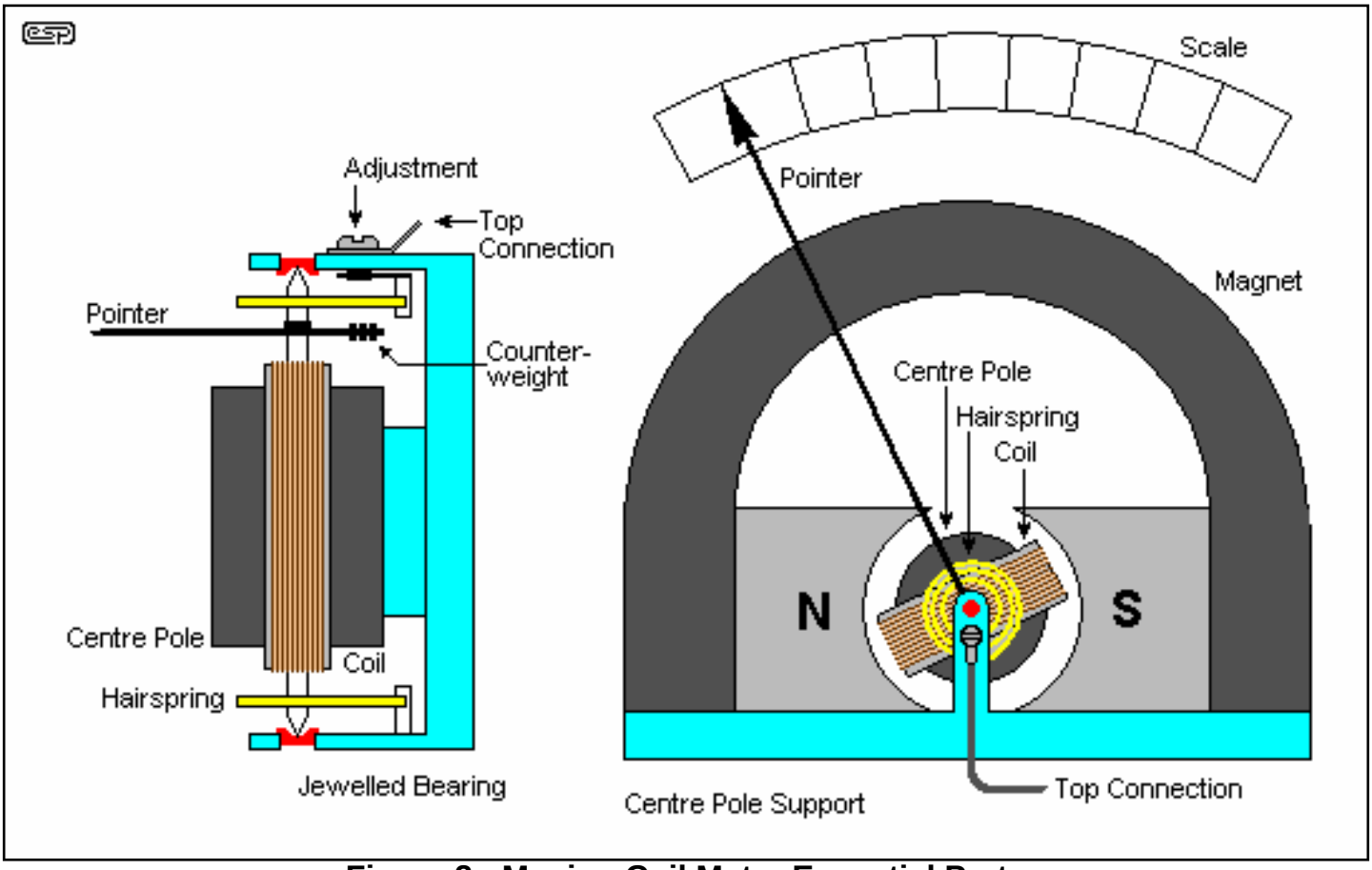


Figure 2 - Moving Coil Meter Essential Parts

The moving coil movement uses a coil former of aluminium, around a centre pole and 'immersed' in a strong magnetic field. The coil is most commonly supported by jewelled bearings (although taut-band suspension is a much better arrangement, IMHO). The coil is maintained at the zero position by the tension of the hairsprings, and one of these (almost always the top) is made adjustable from outside the meter case. This allows the user to zero the pointer. Current to the coil is carried by the hairsprings.

Taut band suspension uses no bearings, but supports the coil on a tiny flat spring (a flat wire) at each end. The flat spring acts as both suspension and restoring force, as well as providing current to the coil itself. Unfortunately, taut band movements are not very common, possibly because they are sometimes not as mechanically rugged as the traditional jewelled pivot suspension, and are very difficult to repair if the suspension breaks (personal experience!). A major advantage is that they have very low (virtually zero) hysteresis - this is caused in jewelled movements if the pivot sticks slightly because of contamination or damage.

The aluminium former is almost invariably made so that it forms a shorted turn around the centre pole. This provides electrical damping, preventing excessive pointer velocity. There is a lot more to the analogue meter movement than meets the eye, but we shall leave the topic now, so that the usage of these devices can be covered.

2.1 - Movement Specifications

All moving coil meters have a rated current for FSD (Full Scale Deflection), and this parameter is of primary importance. The FSD current determines how much load the meter will place on any drive circuitry, or for a voltmeter, how much current it will draw from the voltage source. This may or may not be important, depending on application.

Most commonly available meters are readily available with a sensitivity of between 50uA and 1mA FSD. More sensitive meters are available, but the cost goes up with increasing sensitivity. The most sensitive meter I have heard of was used by Sanwa in an analogue multimeter - 2uA FSD, taut band movement!

All meter movements have resistance, because the coil uses many turns of fine wire. The resistance varies from perhaps 200 ohms or so (1mA movement) up to around 3.5k for a 50uA movement. These figures can vary quite widely though, depending on the exact technique used by the manufacturer.

Normally, moving coil meter movements are suitable for DC only. Some (such as VU meters for audio) have an internal rectifier so that AC may be measured, but accuracy is generally rather poor, especially with low voltages.

To obtain good AC performance requires the use of external circuitry. The project pages have a design for an AC millivoltmeter, and there is an interesting array of precision rectifier circuits in the application notes section of the ESP site.

Some movements have a mirrored scale, where a band of highly polished metal is just behind the scale itself. This is used to eliminate parallax errors as you read the meter, and can improve reading accuracy dramatically. None of this is useful if the meter is poorly calibrated or non-linear. Moving coil meters *can* be non-linear if the magnetic path is not adjusted correctly - such adjustments are not recommended for anyone not trained or used to working on very delicate equipment. It also helps if you know exactly what to do, a topic that is well outside the scope of this article.

3.0 - Voltage Multiplier

When a meter is to be used as a voltmeter, a series resistor is used to limit the current to the specified FSD with the maximum applied voltage that you want to measure. This is a very easy calculation to make, since it involves nothing more advanced than Ohm's law.

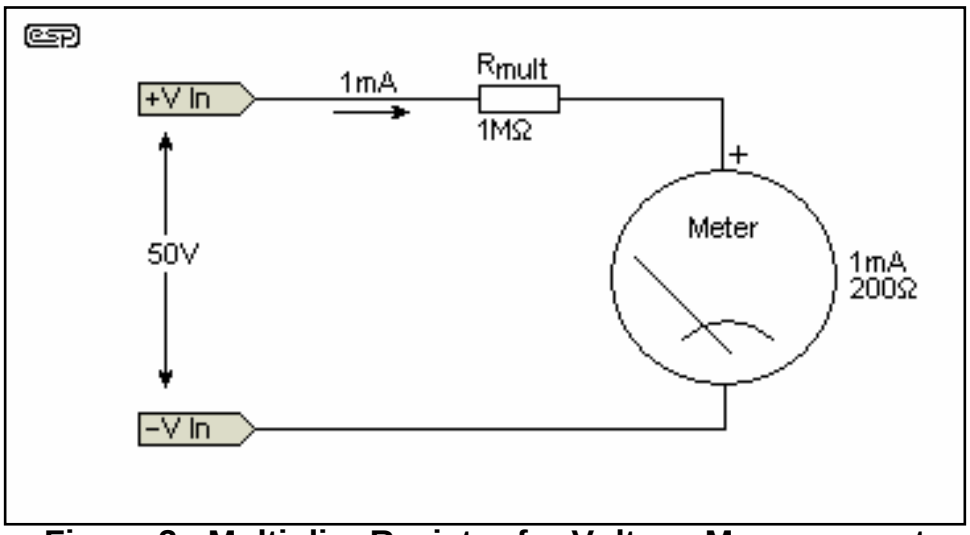


Figure 3 - Multiplier Resistor for Voltage Measurement

For example, we want to measure the voltage from a power supply, and have a 1mA meter movement available, having a coil resistance of 200 ohms. If the maximum supply voltage is 50V, then the meter should read from 0-50V. The total resistance needed will limit the current through the meter to 1mA with 50V applied, so ...

$$R_{total} = V / I = 50 / 1mA = 50k\Omega$$

Since the meter has 200 ohms resistance, the series resistor will be

$$R_{mult} = 50k - 200 \text{ ohms} = 49{,}800 \text{ ohms}$$

This is not a standard value, so will need to be made up using series / parallel resistors. Of course, one can always cheat and use a 47k resistor in series with a 5k pot, thus enabling the meter to be calibrated to a high accuracy. We do need to check the resistor power rating, because it is easy to forget that the multiplier resistor can dissipate a significant power - especially at high voltages. The resistor power is given by ...

$$P = I^2 * R = 1mA^2 * 47000 = 53mW$$

The power dissipation is well within limits for even the lowest power resistor. Be very careful when determining the multiplier resistance for high voltages. Although the power

rating may be quite low, the gradient voltage across the resistor may exceed its ratings. It is imperative that resistors are not operated above the maximum rated voltage for the particular type of resistor. This specification is not often given, so it is best to assume the worst case, and limit the voltage across any 0.5W resistor to no more than around 150V - less for 0.25W resistors.

It is generally preferable to use the most sensitive meter you can get within your price range, so in this case, a 50uA movement would be a far better proposition. Less current is drawn from the measured voltage source, so there is less loading on potentially sensitive circuits. This was always a problem when measuring voltages in valve amplifiers, because typical cheap analogue multimeters often used relatively high current movements, and this loaded the voltage under test giving incorrect readings. Analogue multimeters usually had a rating of 'Ohms/Volt' - the 1mA movement described above uses 50k total resistance to measure up to 50V, so that would be rated at 1kΩ/ Volt. The better multimeters of yesteryear were rated at a minimum of 20kΩ/V up to 100kΩ/Volt (the Sanwa meter mentioned above was 500kΩ/Volt!). To obtain even higher measurement impedance, the better equipped workshops and laboratories back then used a VTVM (Vacuum Tube Volt Meter), offering an input impedance of around 10-100MΩ. These were followed by FET input transistorised units, and finally displaced by digital multimeters. Despite their popularity, digital multimeters are still very bad at some measurements, and are often not as accurate as we tend to think they are.

Using a 50uA movement, the multiplier resistor needs to be ...

$$R_{mult} = V / I = 50 / 50uA = 1M\Omega - 3500 \text{ (meter resistance)} = 996{,}500 \text{ ohms}$$

... which works out to be 20kΩ/ Volt. Again, this resistance can be made up by series connection of different values, but a 1MΩ resistor is perfectly ok. The error is much smaller than the tolerance of the resistor or the meter movement, at 0.35%. If you need greater accuracy you will need to use a trimpot with a series resistor as described above for the 1mA movement.

That's all there is to multipliers - as stated in the beginning of this section, they are very easy to work out.

4.0 - Current Shunt

The situation is a little more complex when calculating a shunt for current measurement. Not so much because the calculations are difficult, but because you will be working with

very low resistance values. It is also important to ensure that the meter is connected directly to the shunt - even a small length of wire in series may make readings uselessly inaccurate. The schematic diagram below shows not only the electrical connection, but also the physical connection to the shunt.

In most cases, it is easier to calculate (or measure) the voltage across the meter movement for FSD. If you don't know the resistance, it can be measured with a digital multimeter. The current from most digital multimeters is low enough not to cause damage to the meter, but the pointer may swing rather violently. Connect with reverse polarity to minimise the risk of bending the pointer.

Unless you are measuring low currents (less than 1A or so), the shunt resistance can be worked out using Ohm's law, and will be accurate enough for most purposes. This is covered below.

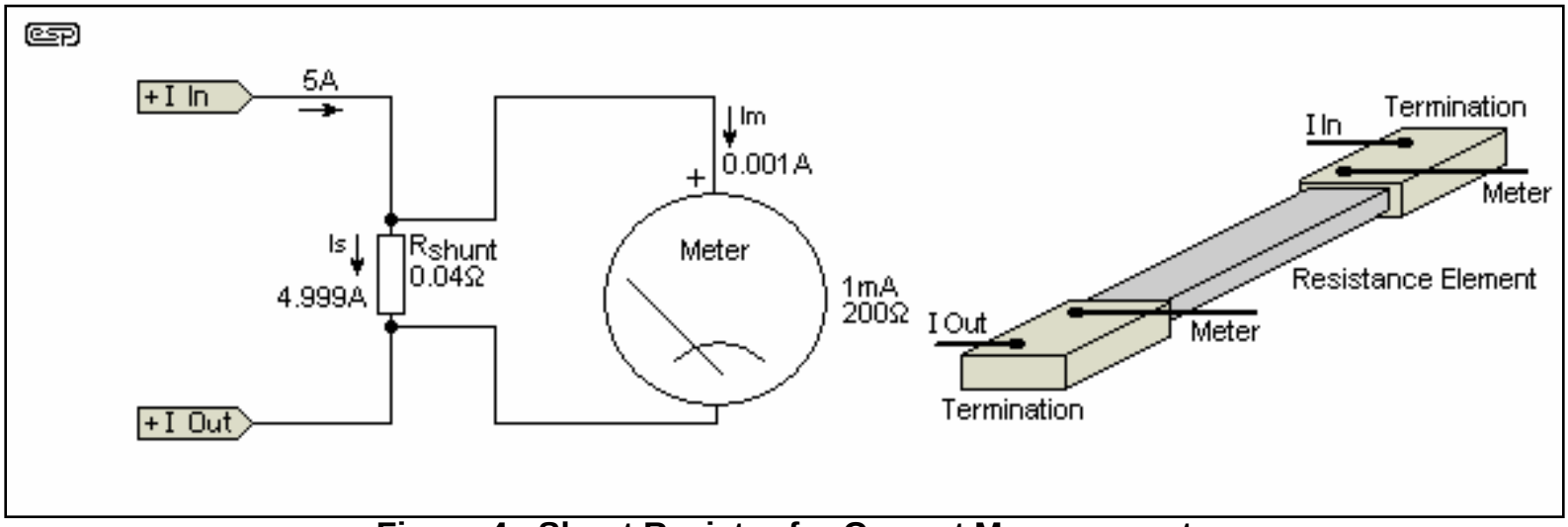


Figure 4 - Shunt Resistor for Current Measurement

Assuming a 1mA movement with an internal resistance of 200 ohms, as an example we wish to measure 5A. This means that 4.999A must pass through the shunt, with the remaining 1mA passed by the meter movement. The shunt resistance can be found with the following formula ...

> Rs = Rm / (Is / Im) where Rs is the shunt resistance, Rm is the meter resistance, Is = shunt current, Im = meter current

So for our example,

> Rs = 200 / (5A / 1mA) = 0.04Ω

If we use only Ohm's law (having determined that there will be 200mV across the movement - 1mA and 200 ohms), the shunt can be calculated as ...

Rs = Vm / I where Rs is shunt resistance, Vm is meter voltage at FSD, and I is the current
Rs = 0.2 / 5 = 0.04Ω

This method will work to within 1% accuracy provided the measured maximum current is more than 100 times the meter current. One thing we have to be careful of with shunts is that the voltage 'lost' across them is not excessive. This will reduce the voltage supplied to the load, and can result in significant errors, especially at low currents. For example, if we only need to measure 1mA, we can use the meter directly, but we lose 200mV across the meter. In the case of the 0.04 ohm shunt calculated above, we lose

V = R * I = 0.04 * 5A = 200mV

... exactly the same voltage loss! It's not a great deal, but can be critical in some exacting tests or at very low voltages. 200mV is almost nothing with a 50V supply (0.4%), but is very significant if the applied voltage is only 1V (a full 20% loss). The voltage drop can be reduced slightly by using a more sensitive movement. For a 50uA movement with 3,500 ohms resistance, the loss is

V = R * I = 3500 * 50uA = 175mV

... not much of a gain, but not many alternatives. Current measurement will always lose some voltage, so it is important that the voltmeter is always connected *after* the ammeter, so that the 'lost' voltage is taken into consideration. Where extremely low voltage drop is important, one must resort to amplification. An opamp can be used to amplify the voltage across a much smaller value shunt, but at the expense of circuit complexity and temperature drift. Digital panel meters are better than analogue movements for current measurements.

The idea of a shunt is all well and good, but where does one obtain an 0.04 ohm resistor? It can be made up of a number of wirewound or metal film resistors in parallel, or a dedicated shunt may be available. Obtaining high accuracy at such low resistances is very difficult though, and shunts are generally cut, machined or filed to remove small amounts of metal until the exact value needed is achieved. The shunt must be made from metal having a low temperature coefficient of resistance to prevent the reading being affected by changes in temperature - either ambient, or caused by the load current heating the shunt..

There is an easier way, as shown in Figure 5. The voltage drop will be a bit higher than it should be, but you only need a few millivolts extra to be able to use the technique.

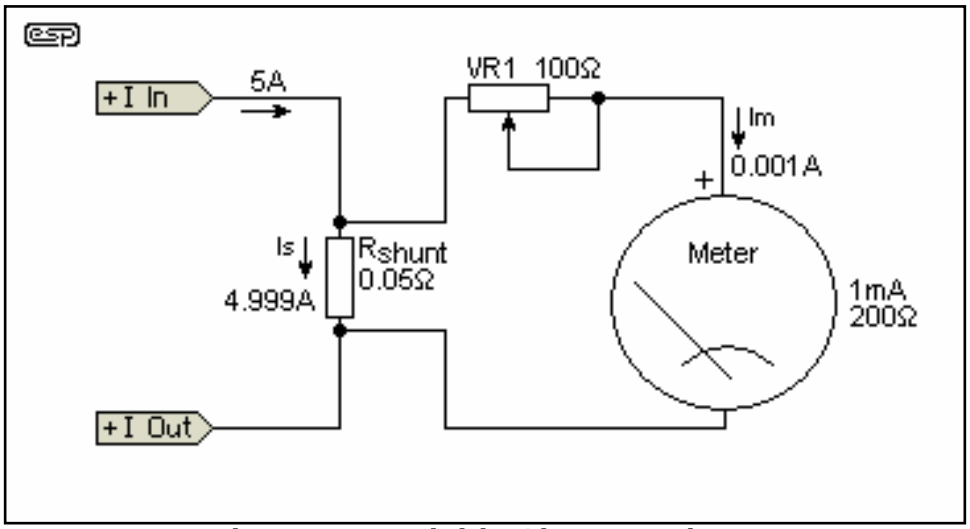


Figure 5 - Variable Shunt Resistor

Now it is possible to use 2 x 0.1 ohm resistors in parallel, giving 0.05 ohm. The voltage drop at 5A will be 250mV, but you have the advantage of being able to use standard tolerance resistors, which can represent a significant saving. The power is only 1.25W at full current, so a pair of 5W resistors will barely get warm. The trimpot can be adjusted to give an accurate reading, without having to resort to close tolerance resistors with impossible values. As an example for the above 5A meter, we could use a 100 ohm trimpot in series with the meter. The value is not particularly important, but needs to be within a sensible range.

What is 'sensible' in this context? Easy. We already know that the meter needs 200mV for full scale and that we will get 250mV across a 0.05Ω shunt, so we need a resistance that will drop 50mV at 1mA.

$$R = V / I = 0.05 / 0.001 = 50 \text{ ohms}$$

Since we are using a pot, it is advisable to centre the wiper under ideal conditions to give maximum adjustment range (to allow for worst case tolerance), so a 100 ohm pot is ideal.

5.0 - Expanded Scale Voltmeter

You may have seen expanded scale voltmeters used in cars to monitor the battery voltage. Since no-one is interested if the battery measures less than 10V (it's dead flat!), and it should never exceed 15V, a meter that measures from 10V to 15V is nice to have. This is surprisingly easy to do, and although absolute accuracy is not wonderful in a

simple application, it is more than acceptable for the purpose.

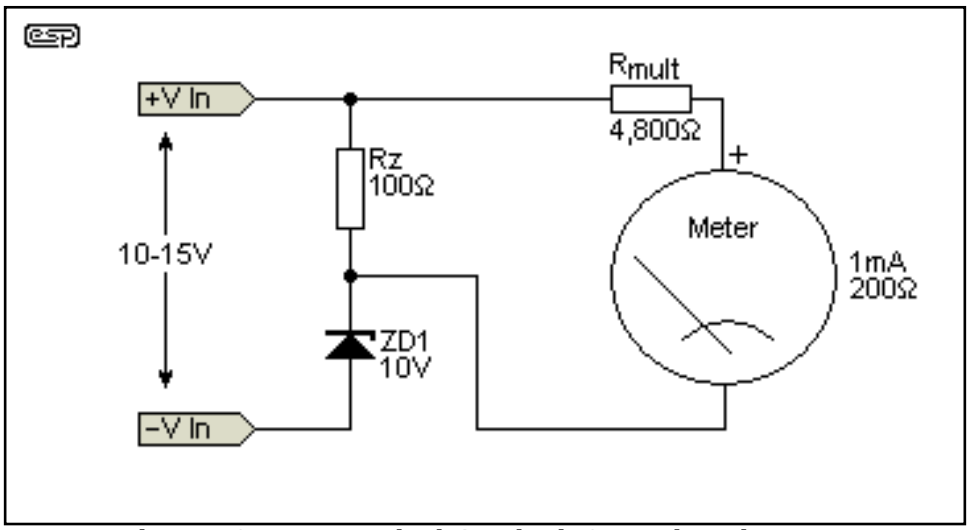


Figure 6 - Expanded Scale (10-15V) Voltmeter

By using a zener diode, a base reference is established, and the meter only measures between the reference and actual battery voltage. We will use a 1mA movement again (as shown above). This scheme can be adapted for any desired voltage. The voltmeter only needs to measure the voltage drop across the zener feed resistor, which is needed to ensure that an acceptable current flows in the zener diode. The 1mA drawn by the meter is not enough to obtain a stable voltage.

The multiplier is worked out in the same way as before ...

$$R_{total} = V / I = 5 / 1mA = 5k\Omega$$

Because the multiplier resistance is much smaller than before, we must take the meter resistance of 200 ohms into consideration.

$$R_{mult} = R_{total} - R_{meter} = 5000 - 200 = 4800\Omega$$

A 4.7k resistor will introduce a small error, but a 3.9k resistor in series with a 2k trimpot will allow the meter to be set very accurately. The zener feed resistor value is not critical, but should ensure that the zener current is between 10% and 50% of the maximum for the device. Assuming a 10V 1W zener, the maximum current is ...

$$Iz_{max} = P / V = 1 / 10 = 0.1 = 100mA$$

Using Ohm's law, we get a resistance value of 100 ohms for a zener current of 50mA at

15V. This will fall as the voltage is reduced, but extreme accuracy at 10V is not possible, so this arrangement should work fine.

6.0 - Digital Panel Meters

DPMs (Digital Panel Meters) are often very attractive, not just for their perceived accuracy, but because they can often be obtained for the same or less than a good analogue meter movement. They also have better linearity than most of the cheap movements, so there are some real benefits. Most are available with a quoted sensitivity of 200mV (199.9mV full scale), so are comparable to analogue meters in terms of voltage drop for current measurement. They have the great advantage of a (typical) 100MΩ input impedance, so voltage loading is extremely low. In addition, they will measure positive and negative voltage or current - this is available with a centre zero analogue meter, but they are hard to find.

Most DPMs are classified as 3½ digit, meaning that they display up to a maximum of 199.9mV. The most significant digit can only be blank or 1, and the other 'half' is used to display a negative sign to indicate that the input is negative with respect to the common or ground terminal. This often means that much of the range is wasted if you want to display a range other than 0-1999. Note that most DPMs do not automatically select the decimal point, and there are extra pins to allow the user to select the position of the decimal point (or to ignore it completely). Analogue meters have no such limitation, because the scale can be calibrated with any units you wish, and covering any range.

6.1 - Digital Voltmeter

Measuring voltage with a DPM is easy - most even come with instructions that show you how to do it. You do need to be careful to ensure that possibly destructive voltages cannot be coupled to the inputs. Like all ICs, the ADC (Analogue to Digital Converter) used is sensitive to excess voltage, and the IC can be destroyed.

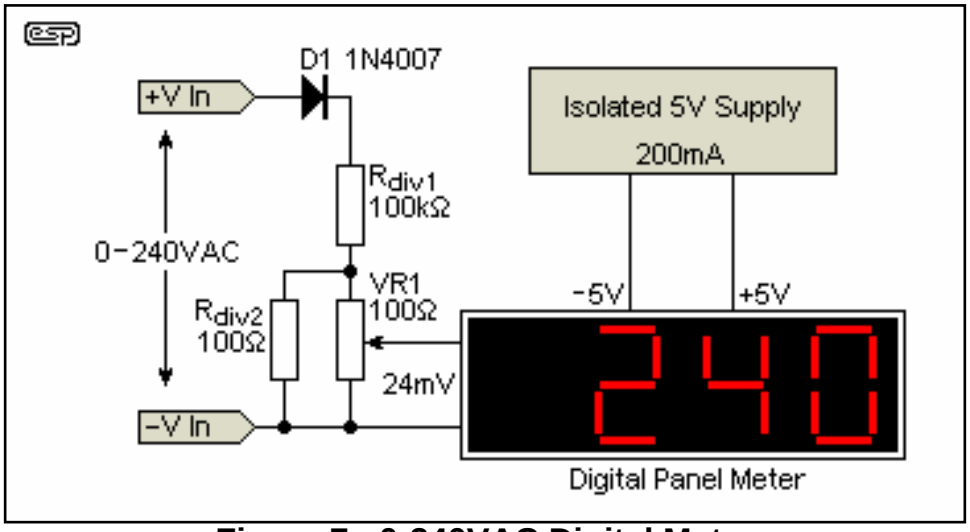


Figure 7 - 0-240VAC Digital Meter

Figure 7 shows the circuit of a DPM voltmeter I built recently. This is designed to monitor the output from my workshop Variac (variable transformer). To ensure an adequate voltage rating for R_{div1} 4 x 100k 1W resistors were used in series parallel, maintaining the peak voltage across each to 170V (the peak of 240V AC is 340V). 1W resistors were not used for their power rating, but to have a large resistance section, maintaining a relatively low voltage gradient across the resistor surface. Because it is only used for 0-240V, the voltage to the DPM will be 0-24mV, and this is a half-wave rectified signal. The meter averages the applied voltage. Note that the 5V supply *must* be isolated, because it could have the full mains potential on all terminals if the active (live) and neutral conductors are ever swapped around. This is critically important - the entire circuit (including power supply) must be considered as being at mains potential.

To obtain the (approximate) average value of ½ wave rectified AC, you divide the peak voltage by 3.12. Based on this and for an average signal of 24mV, the average input voltage is 109V (340 / 3.12), so the voltage divider needs a ratio of

$$Vdiv = Vin / Vout = 109 / 24mV = 4542$$

For all reasonably high voltages, the division ratio is so high as to cause significant errors, and the use of a trimpot to adjust the value is strongly recommended. Since I used 100k for Rdiv1 (because I had 100k/1W resistors handy), the parallel combination of Rdiv2 and VR1 needs to be slightly more than

$$Rdiv2 \cong (Rdiv1 / Vdiv) - 1 \cong 22 \text{ ohms}$$

50 ohms (as used) allows VR1 to be roughly centred, and there is plenty of adjustment range. Needless to say, exactly the same technique can be applied to an analogue meter

as well, but you need to allow for the much lower input impedance (perhaps 100 ohms rather than 100M for the DPM that I used). As it turns out, with an average voltage of 108V and a resistance of 100k, the current is 1.08mA, so the meter can be driven directly (leaving out Rdiv2 and VR1). You will need to readjust the resistance though, because the (in)accuracy is 8% - much better results can be obtained. A pot is highly recommended because the AC waveform is not very predictable, and severe errors may result from waveform distortion.

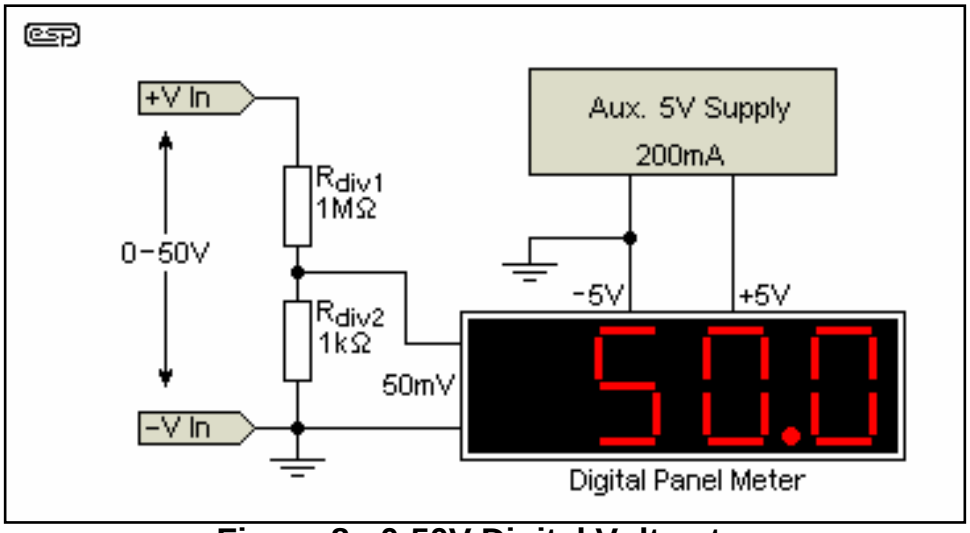


Figure 8 - 0-50V Digital Voltmeter

For a more conventional application, Figure 8 shows a basic 0-50V digital meter. The resistor values are fixed in this case. Because of the high input impedance of the DPM, we can use 1M for the upper divider resistor. The division ratio is determined the same way as before ...

Vdiv = Vin / Vout = 50 / 50mV = 1000
Rdiv2 = Rdiv1 / Vdiv = (1M / 1000) - 1 = 999Ω

Using a 1k resistor is not an issue, because the resistor tolerance is much wider than the 1 ohm difference in the calculated values. The same result can be achieved using 10k and 100 ohms, but there is not normally any need to aim for very low impedances. You may find that the meter displays 'rubbish' values in the least significant digit - this means that noise is being picked up, and a lower impedance will reduce that. If you need the circuit to be accurate, then you will need to use 0.1% resistors or add a pot so it can be adjusted. A pot is a lot cheaper and easier to get than 0.01% resistors, especially if you end up with odd values.

6.2 - Digital Ammeter

DPMs have a benefit as ammeters, but usually only if you don't need the full scale. Since the typical sensitivity is 200mV, by using only a part of the maximum reading, you can use lower shunt resistances than with analogue movements.

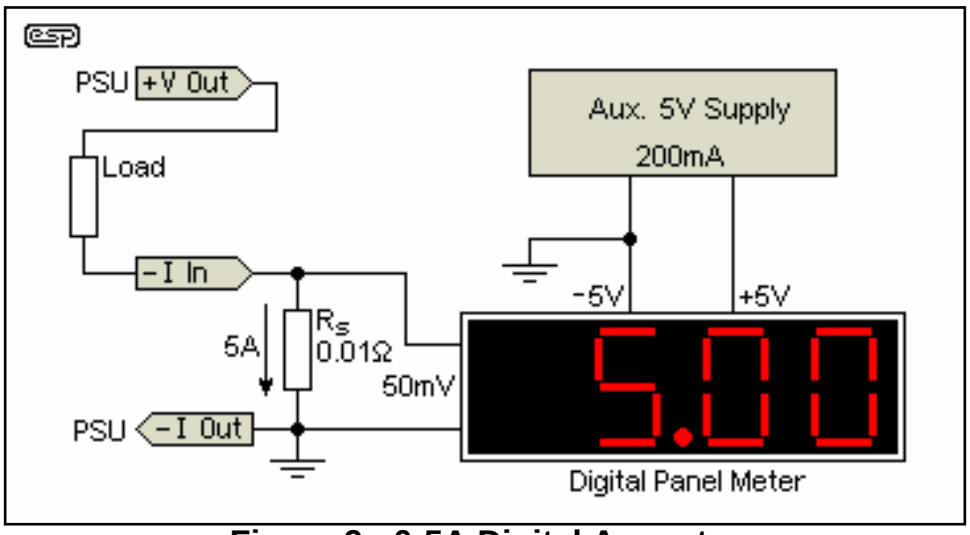


Figure 9 - 0-5A Digital Ammeter

The procedure for calculating the shunt is exactly the same as for an analogue meter, except that there is no meter current. You simply need to calculate the shunt based on the meter voltage for the desired current reading ...

$$R_s = V_s / I = 50mV / 5A = 0.01\Omega$$

This gives a much lower shunt resistance, because only 50mV is needed at the meter input. The circuit shown will work up to 20A (19.99A to be exact) with the same 0.01Ω shunt resistor. Note that the input is shown on the negative supply, with the +ve input going to the positive supply via the load. If the input and power supply -ve terminals are not at the same potential, then the supply for the meter *must* be floating - it cannot be grounded. If you wanted to monitor the current in the positive supply lead for example, you need a floating auxiliary supply.

7.0 - Make Your Own Multimeter

In general, this would have to be considered a silly topic. After all, one can buy a multimeter quite cheaply, and the switching is a nightmare. For specialised applications though, there may be perfectly good reasons for making a multi-range meter. Bear in mind that the circuit shown below does not include protection for the DPM, so if 2kV were applied when the 200mV range was selected, the meter will be destroyed.

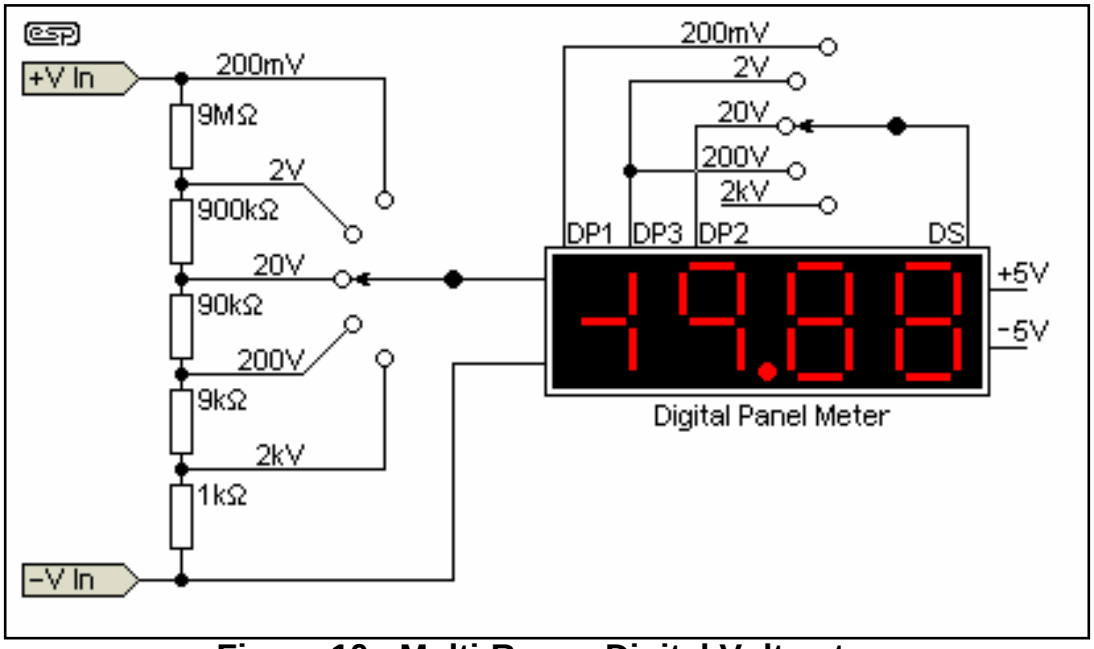


Figure 10 - Multi-Range Digital Voltmeter

You need a 2-pole 5-position rotary switch, and the insulation must be sufficient for the maximum voltage. Any protection circuit that you add must not load the external circuit, otherwise the meter may appear as a short circuit to high voltages. As noted, this is basically a silly idea, but it may be useful (even essential) for some applications where a conventional multimeter would be inappropriate. No, I can't think of such a situation either .

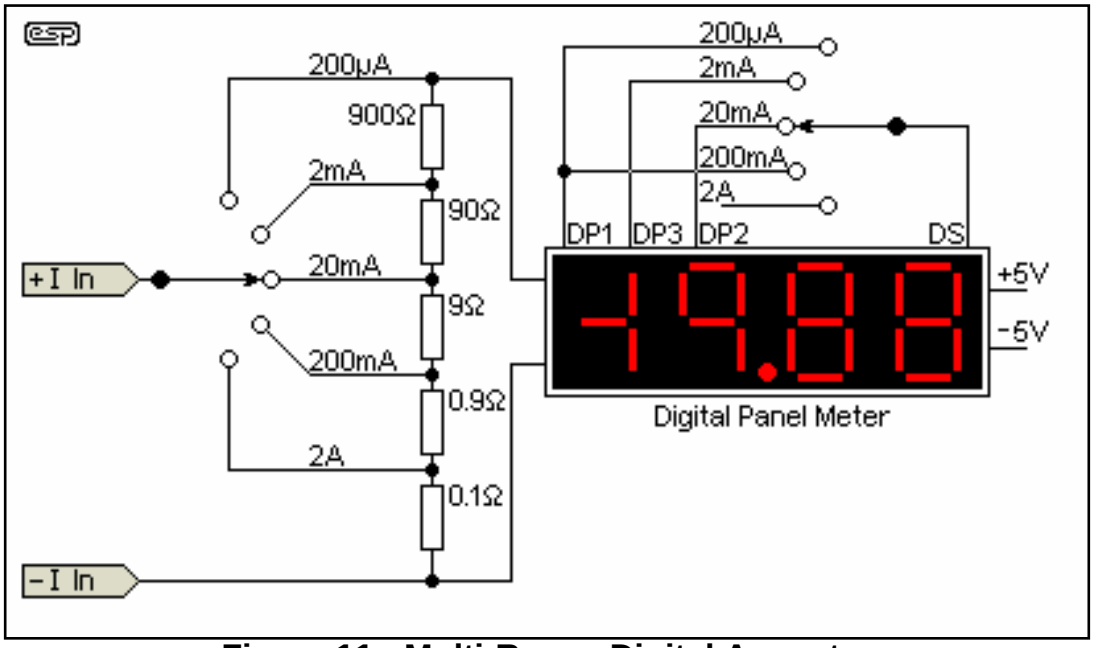


Figure 11 - Multi-Range Digital Ammeter

Similar comments apply to the ammeter. In this case, the resistors and switch must be capable of handling the current, although this only becomes an issue on the highest current range. Like the multi-range voltmeter, the usefulness of Figure 11 is somewhat

dubious, although it would be nice on a laboratory power supply. The ranges can be expanded or moved - for example you may find that ranges from 2mA to 20A suit your needs. Simply reduce all resistance values by a factor of 10, and that's what you have.

It almost looks like this section is pretty useless, but the final application allows you to do things that no normal multimeter will - measure very low resistances.

7.1 - Measuring Resistance

There are many reasons one may want to measure very low resistance values. Transformer windings, loudspeaker crossover inductors (assuming you are actually interested in passive crossovers), or perhaps you need to be able to measure current shunts .

For very low resistance values you have two choices - either use a very sensitive voltmeter, or a high measurement current. Both methods have disadvantages. High sensitivity is difficult for DC amplifiers because of drift. Changes in temperature cause opamp offset voltage and current to change, and that affects the readings. While there are methods to (almost) eliminate drift, they are beyond the scope of this article.

High measurement current can cause the device under test (DUT) to heat, and that will affect the resistance. Some things that have low resistance may not be able to even tolerate the kind of current that you may need to be able to measure them. In general, a maximum current of around 1A will allow most low resistance measurements without too many risks, but naturally the current source can be made variable, with switched ranges to provide a wide measurement range.

With a measurement current of 1A you will get a meter that can measure 0.2Ω full scale, so very low resistances can be measured. Needless to say, battery operation is not recommended if you aim to make a resistance meter that will provide 1A.

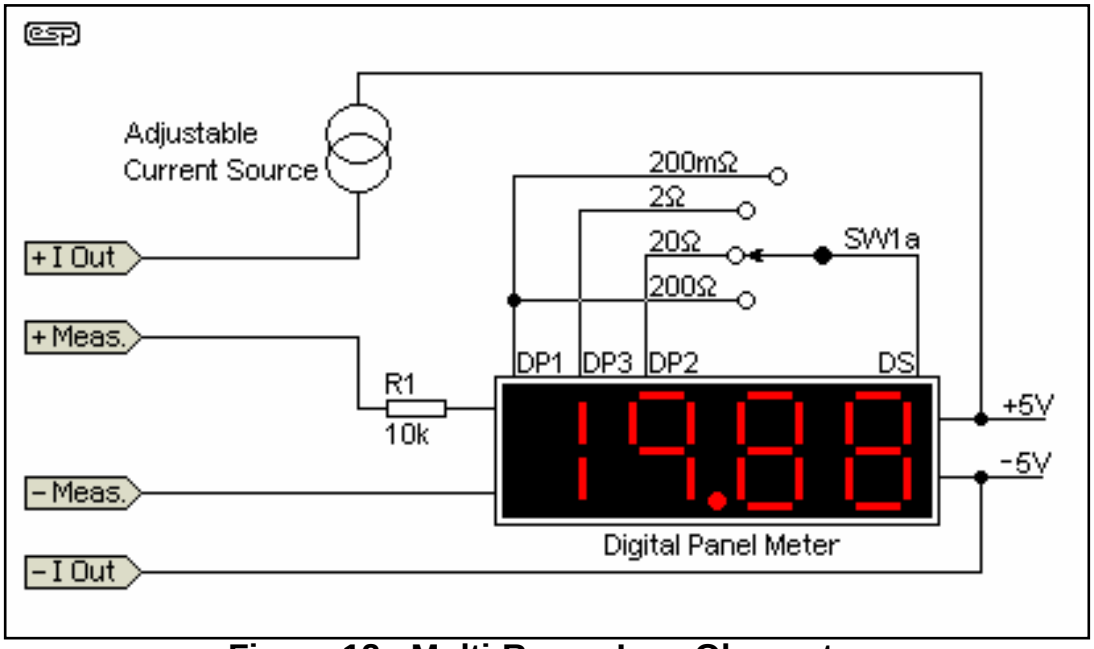


Figure 12 - Multi-Range Low Ohmmeter

The use of 4 terminals is essential for very low resistances. Two wires carry the current to the DUT, and the two measurement leads are then connected as close as possible to the device itself, with a component lead length equal to what will be used when the component is installed. This technique avoids errors caused by lead and connection resistances. While it is possible to null out the lead resistance, connection resistance tends to be variable, and can cause substantial measurement errors. This method is very common for this type of instrument. R1 is used to prevent possible damage to the DPM if it is subjected to an over-voltage condition.

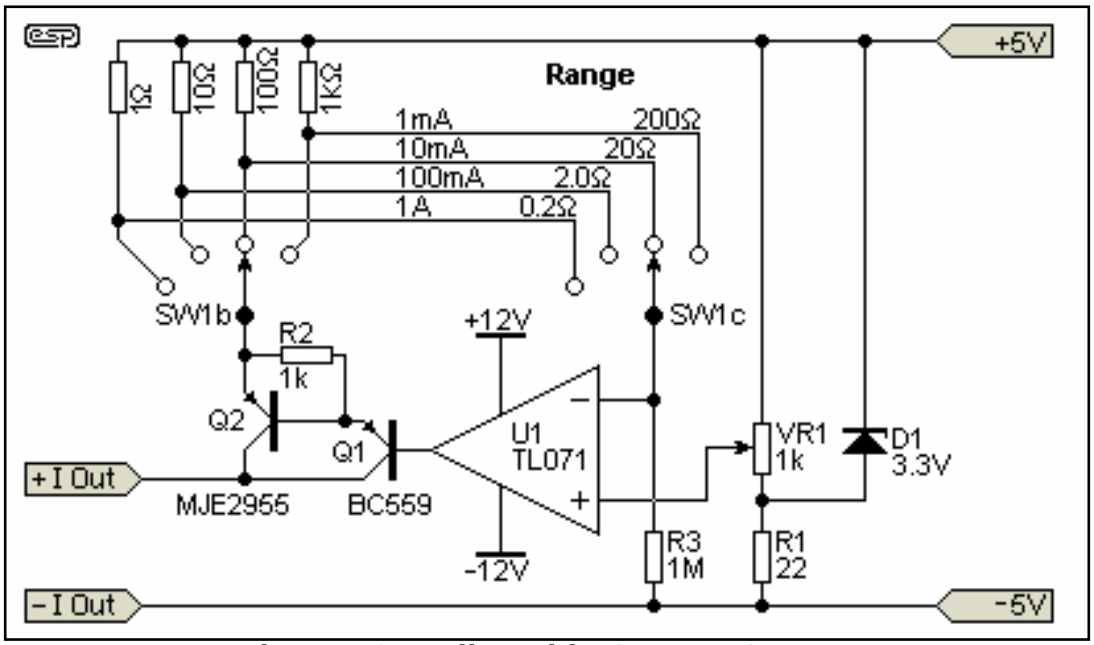


Figure 13 - Adjustable Current Source

The adjustable current source requires accurate calibration, and will be as good as your

construction and choice of components allows. Temperature drift is always a problem with precision circuits like this one, but the circuit as shown will be quite accurate within the normal ambient temperature range. The current setting resistors (those connected to SW1b) need to be as accurate as possible.

The greatest difficulty is the switch used to select current ranges. Even the smallest amount of resistance will cause large errors. By switching both the resistor and the measurement point (the opamp's inverting input), the error is minimised because the switch resistance does not form part of the measurement circuit. R3 is included to ensure that the current source is switched off as you change ranges.

VR1 is adjusted so there is exactly 1V between the opamp's positive input and the 5V supply. When exactly the same voltage (1V) is developed across any of the current setting resistors, the current through it must be as specified. A tiny error is introduced because the base current of Q1 is added to the total, but this should amount to less than 0.1%. The 5V supply needs to be well regulated, and capable of at least 1.5A without any appreciable change of voltage. If the 0.2 ohms range is not needed, you can leave out the 1 ohm resistor and simplify the switching accordingly. Q2 can then be changed to a BD140.

Although a zener is not the most ideal voltage reference, they are easy to obtain. Precision voltage reference diodes are available, but they are relatively expensive and only stocked by a few major parts suppliers. The zener is deliberately operated at a relatively high current (about 100mA) so that it will get reasonably hot. This helps to stabilise it against ambient temperature variations, so the circuit will take a few minutes to settle down after power is applied.

This circuit can also be used as a stand-alone low ohms adaptor. It obviously needs the power supplies, but you can use your multimeter to measure the voltage across the DUT. The resistance is read as a voltage (the same way that your meter does it internally), with the appropriate conversion based on the current source setting.

The second section of the low ohm meter circuit can be used in conjunction with an analogue movement if you prefer. You will need to apply your own multiplier to the scale and add any necessary extra resistance for calibration, but it will work just as well. You will have to make your own scale - see conclusion, below.

8.0 - Conclusion

The metering systems described here should be considered a guideline, rather than usable circuits in their own right. By following the information shown, you will be able to create a meter for almost any measurement for which meters can be used. If AC metering is needed, then I suggest that you look at the various meter circuits in the Projects pages.

Although it may seem unlikely, this article has only covered the basics. Metering is widely used for many different applications, and it is impossible to cover every possibility in a short article. It is hoped that the information proves useful to anyone who has been wondering exactly how to go about adding a meter to their latest power supply project, or who has a real need to measure low resistances.

One final point - scales. It is often difficult (or impossible) to get a meter scale that is calibrated with the units you want. The resolution of modern printers is more than acceptable to allow you to create your own scale, which can then be printed. Ink-jet photo printing paper gives an excellent finish, and after you have cut the scale to fit, it can be attached over the existing scale with spray adhesive. Make sure that there is sufficient clearance for the pointer, and avoid 'whiskers' of paper that can cause the pointer to stick. While the meter is dismantled, be careful to ensure that no magnetic materials (iron filings, etc.) are allowed to enter the gap, as these will cause the meter to stick and are a real pain to remove (personal experience).

9.0 - References

1. Shunts and Multipliers - Jaycar Electronics
2. Moving Coil Meters - HyperPhysics

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2006. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 03 May 2006

The Audio Pages

Elliott Sound Products **Beginners' Guide to Potentiometers**

Copyright © 2001 - Rod Elliott (ESP)
Page Created 22 Jan 2002

Articles Index

Main Index

Contents

- Introduction
- Basic Pots and Knobs
  - Knobs
  - Trimpots
- Potentiometer Tapers
  - Pot Markings
  - Power and Voltage Ratings
- Putting Your Pot to Use
- Changing the Law of a Pot

Introduction

The humble potentiometer (or pot, as it is more commonly known) is a simple electro-mechanical transducer. It converts rotary or linear motion from the operator into a change of resistance, and this change is (or can be) used to control anything from the volume of a hi-fi system to the direction of a huge container ship.

The pot as we know it was originally known as a rheostat (or reostat in some texts) - essentially a variable wirewound resistor. The array of different types is now quite astonishing, and it can be very difficult for the beginner (in particular) to work out which type is suitable for a given task. The fact that quite a few different pot types can all be used for the same task makes the job that much harder - freedom of choice is at best confusing when you don't know what the choices actually are, or why you should make them. This article is not about to cover every aspect of pots, but is an introduction to the subject. For anyone wanting to know more, visit manufacturers' web sites, and have a look at the specifications and available types.

The very first variable resistors were either a block of carbon (or some other resistive material) with a sliding contact, or a box full of carbon granules, with a threaded screw to compress the granules. More compression leads to lower resistance, and vice versa. These are rare in modern equipment, so we shall limit ourselves to the more common types :-)

Basic Pots and Knobs

It is worthwhile to have a look at a few of the common pot types that are available. Figure 1 shows an array of conventional pots - both PCB and panel mounting.

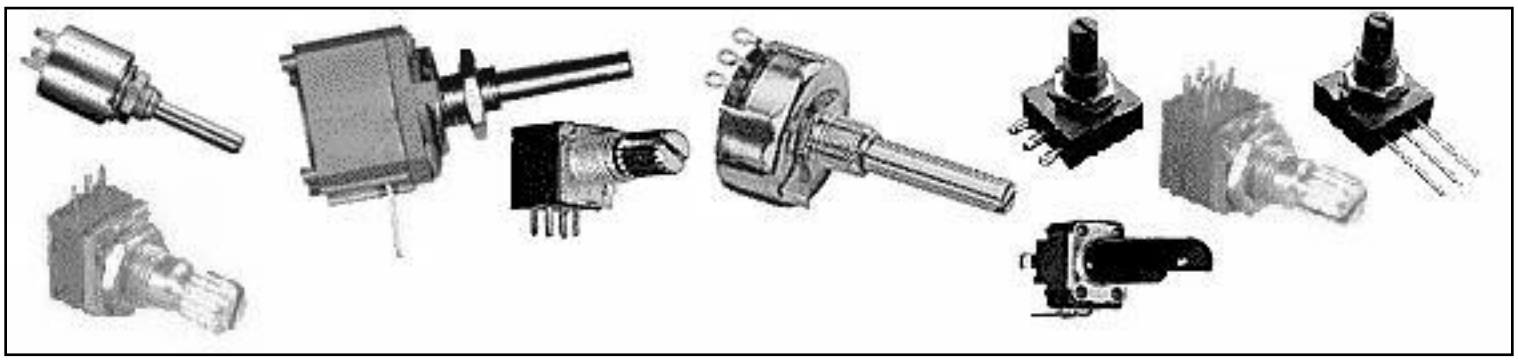


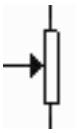
Figure 1 - Some Examples of Pots

Note that these are not to scale, although the relative sizes are passably close. Apart from the different body shapes and sizes, there are also many "standard" mounting hole and shaft sizes. Probably the most common of all is the one in the centre of the picture. A panel mount, 25 millimetre (1") diameter pot. This uses a 10mm (3/8") mounting hole, and has a 6.35mm (1/4") shaft. These pots have been with us almost unchanged for 40 years or more.

The remainder show a few of the many variations available. The fluted shaft types are commonly referred to as "metric", but will accept a standard 1/4" knob - albeit with a little

play (it is less than a perfect fit, but is acceptable if the grub screw is tight enough). Metric pots are also available in 16mm round and 25mm round formats.

Most rotary pots have 270 degrees of rotation from one extreme to the other. A "single turn" pot is therefore really only a 3/4 turn device, despite the name. There are some other rotary types with only 200 degrees or so, and some specialty types may have less than that again.



The standard schematic symbol for a pot is shown to the left (although some people insist on using zig-zag lines for resistors and pots, I don't, so don't expect me to start now :-). A little later, we shall look at the many ways a standard pot may be wired, as well as some further explanations of the different "law" or taper used. Project 01 has been on this site for a long time now, and is a simple and effective way to create an almost logarithmic taper from a linear pot - but I am getting ahead of myself here.

First, we need to continue with the examination of the basic types (and you thought the above small sample was enough :-) Well, as they say ... "You ain't seen nothin' yet!"

Knobs

Before we look at other pot types, a quick sample of knobs. Yes, I know that everyone has seen knobs, but a dissertation on pots would be less than complete if I didn't include these.

Figure 2 - Some Examples of Knobs

Of these, only one deserves special mention - the one on the left. This is a multi turn vernier readout (analogue in this case) for a standard pot. Typically used with precision wirewound or conductive plastic pots, these used to be common on equipment where very accurate (and repeatable) settings were required. They are expensive, but in their day were almost indispensable. Now, a digital panel meter is cheaper, and considered much more "high tech" - such is progress, but at the expense of the "olde worlde" charm of a mechanical contrivance. And yes, you can still get them!

The remainder are perfectly ordinary knobs, and again, are but a very small sample of those available from a multiplicity of manufacturers. Most cheap knobs are plastic, but they are available with brass inserts, in solid aluminium (brushed, anodised, etc.), plastic innards with a thin aluminium outer shell or just an insert. The list is endless, but I shall end it here.

Trimpots

Then of course, there are trimpots (aka trimmers) - pots designed for "set and forget" applications. They are used for "trimming" the value of a resistor, and are commonly used for calibrating instruments, setting the bias current on power amplifiers, and a host of other areas where a circuit cannot be relied upon to give an exact gain, output voltage, or current. Naturally, a normal panel pot can be used, but these are very much bigger, and any calibration or setup control should not be made available for everyone to fiddle with as they please.

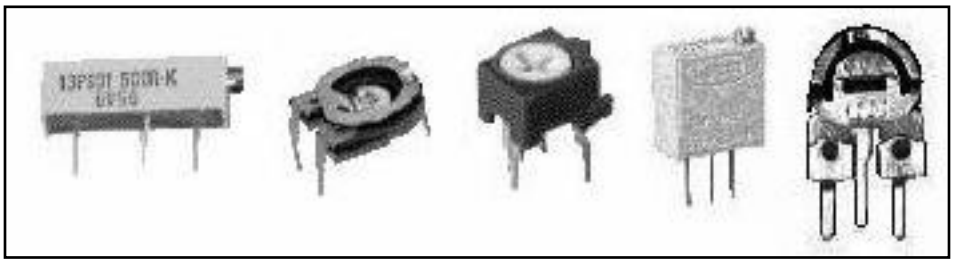


Figure 3 - Some Trimpot Styles

This is a very small sample of those available. The first and fourth are multi-turn types, and these should be used when a very precise setting is required. Because they are sealed, they are relatively immune from contamination, and for all but the most trivial application, should be used instead of the open types (#2 and #5). Trimpots (as shown) are generally available as vertical or horizontal - the choice is usually made based on ease of adjustment of the final circuit.

Potentiometer Tapers

The taper (also called "law") of a pot is important. We need not worry with trimpots, since they are almost always linear, and I do not know of a supplier of anything other than linear trimpots. For all panel pots, we must be aware of the use the pot will have, and select the correct type accordingly.

The most common use of a pot in audio is as a volume control. Since our hearing has a

logarithmic response to sound pressure, it is important that the volume control should provide a smooth variation from soft to loud, such that a given change in position of the pot causes the same sensation of volume change at all levels.

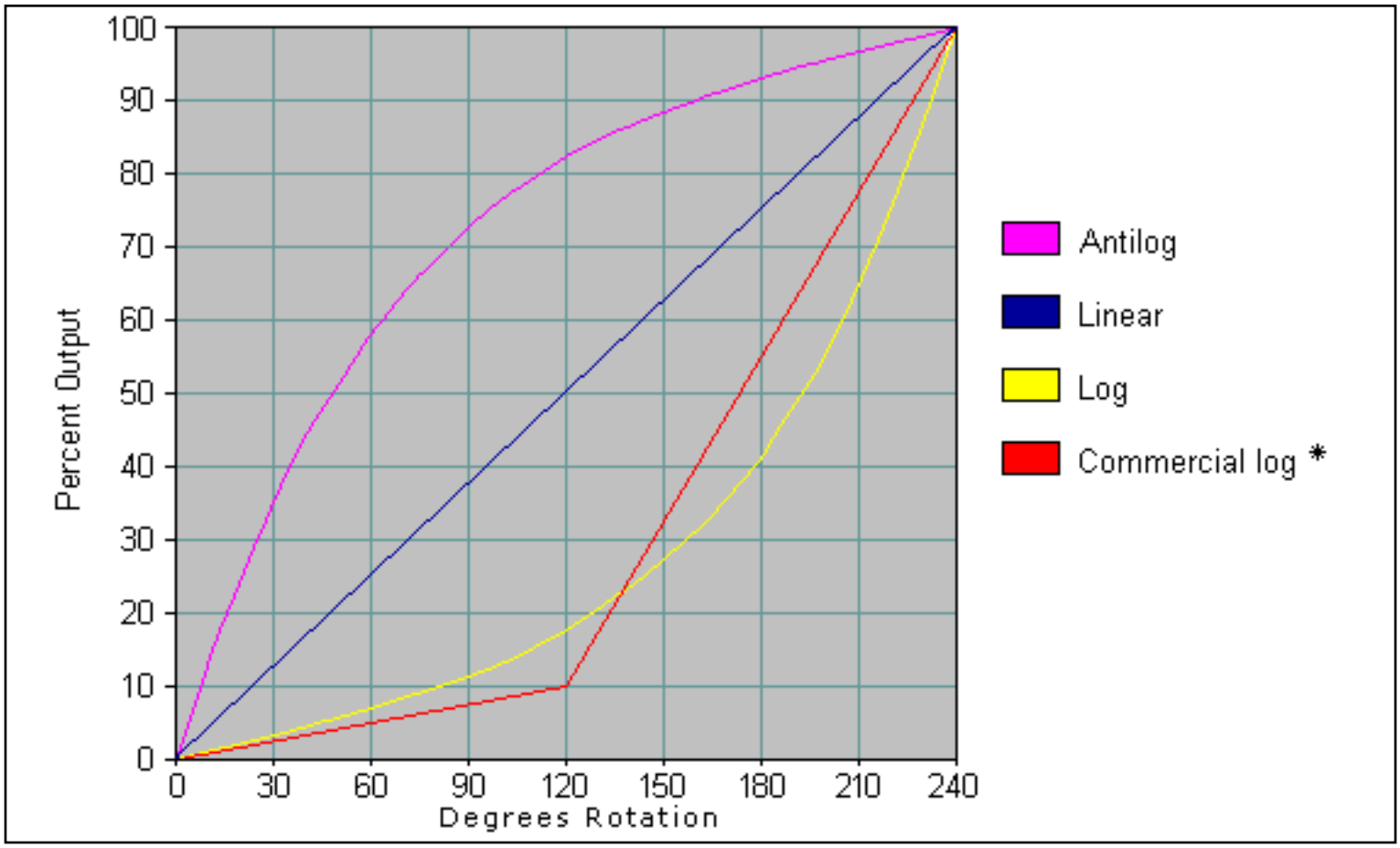


Figure 4 - Potentiometer Tapers

First, the term "taper" needs some explanation. In the early days, when an audio taper (logarithmic, or just log) was needed, the resistance element was indeed tapered, so that it provided a different resistivity at different settings. By changing the physical taper, it was possible to make a pot provide the exact gradient of resistance needed. By definition, a linear pot has no taper as such (the resistance element is parallel sided), but the term has stuck, so we might as well get used to it.

The violet curve in Figure 4 shows an antilog or reverse audio taper pot. These are quite uncommon, but used to be used for balance controls using a log/antilog dual section (commonly called dual gang) pot. It is shown on the graph mainly for its interest value, but they are generally an historical component now.

All this tapering proved a rather expensive exercise, so manufacturers economised ("they" won't notice the difference!), and worked out a method of using two resistance elements of differing resistivity, and joining them to create what I referred to as the

"Commercial log" taper. In short, it doesn't work (not properly, anyway), and the discontinuity where the two sections join is almost always audible with cheap "log" pots.

Project 01 showed how this can be fixed, and I will explain the logic and maths a little more as we progress. In the meantime, I suggest that you get an old pot and dismantle it so that you can see exactly what is inside. I could show you some photos, but there is nothing like doing it yourself to really get to know the subject.

Pot Markings

Now, this should be dead easy - a simple code to indicate the resistance and law of a pot should cause no grief to anyone, right? Wrong! It wouldn't have been so bad if someone hadn't decided to change it, and even then, it wouldn't have been so bad if there was no overlap between the "old" and "new" "standards" ... I think you can see where this is headed by now.

Taper	Old Code	New Code	Alternate
Linear	A	B	LIN
Log (Audio)	C	A	LOG
Antilog	F	N/A	N/A

Wasn't that a nice thing to do? It is obviously important to check before you make assumptions, or you can easily get the wrong type - especially if working on older equipment.

At least the resistance marking is usually sensible, so a 100k pot will be marked as 100K - but not always. The coding system used for capacitors is sometimes used as well (especially on small trimpots), so a 100k pot could also be marked as 104 - 10, followed by 4 zeros, or 100000 (100k) ohms.

Because they are variable, there is a much smaller range of potentiometer values, almost always in a 1, 2.5, 5 sequence. Common values for panel pots are 1k, 5k, 10k, 25k, 50k, 100k, 500k and 1M - 2.5k and 250k went missing along the way, and these are not stocked by very many distributors. 25k pots are becoming harder to get as well. Not all values are available in log and linear, and in some cases you may even find that for a particular type, you can get them in any value you want, as long as it's 100k (for example).

Trimpots suffer a similar fate. The only way to know what you can get from your local

supplier is to check their catalogue. In reality, everything is available, but you may have to go a very long way to get it.

Power and Voltage Ratings

For most audio applications, these are of little on no consequence. In many other applications however, exceeding the specified ratings could lead to the destruction of the pot or yourself! Neither can be considered a good thing.

Power - A pot with a power rating of (say) 0.5W will have a maximum voltage that can exist across the pot before the rating is exceeded. All power ratings are with the entire resistance element in circuit, so maximum dissipation reduces as the resistance is reduced (assuming series or "two terminal" rheostat wiring). Let's look at the 0.5W pot, and 10k is a good value to start with for explanation.

If the maximum dissipation is 0.5W and the resistance is 10k, then the maximum current that may flow through the entire resistance element is determined by ...

> **$P = I^2 * R$** ... therefore
> **$I = \sqrt{P / R}$** ... so I = 7mA

In fact, 7mA is the maximum current that can flow in any part of the resistance element, so if the 10k pot were set to a resistance of 1k, current is still 7mA, and maximum power is now only 50mW, and not the 500mW we had before.

Voltage - Two separate issues here. One is directly related (in part, at least) to the power rating, and is important to ensure that the life of the pot is not reduced. Knowing about the other might save ***your*** life.

Voltage across resistance element - The maximum voltage across the example pot from above is 7mA * 10k, or 70V. This will rarely (if ever) be achieved in an audio system, but is easy with many other designs. As the resistance increases, so does the voltage - a 0.5W 1M pot will pass only 700uA at maximum power rating, but the voltage needed to create this current is 700V. Unless the pot is actually rated to withstand 700V across the resistance element (rather unlikely), it will fail - maybe not today, or tomorrow, but it ***will*** fail eventually. Special pots are made (custom jobs, of course) for high voltages, and standard pots should never be used beyond their rating - assuming that you can find out what the rating is, of course.

Dielectric Voltage - The dielectric (insulation of pot "guts" to the body) rating is especially important if the pot is connected to mains operated, non-isolated equipment. Wall mounted lamp dimmers and such are typical examples. This is not commonly specified, but for safety, should be at least 2.5kV. A common way to achieve this is to use a plastic shaft, with the body of the pot insulated from the chassis, and inaccessible by the user (*even if the knob falls off or is removed!*) This point cannot be stressed highly enough.

Most standard pots will safely withstand (maybe) 100V or so between the resistance element and terminals, and the body and shaft. Miniature types will usually be less than this. Never, *ever*, use a standard pot with a metal shaft to control direct mains operated equipment.

Potentiometer Types

"But we already covered that, didn't we?" Not really - I merely glossed over the basics. Now, we shall look at a few examples of pots you may come across. Firstly, there is the resistive material and some typical characteristics ...

Material	Manufacturing Method	Common uses	Power (Typ)
Carbon	Deposited as a carbon composition ink on an insulating (usually a phenolic resin) body	Most common material, especially for cheap to average quality pots. Has a reasonable life, and noise level is quite acceptable in most cases. (DC should not be allowed to flow through any pot used for audio control)	0.1 to 0.5W
Cermet	Ceramic/metal composite, using a metallic resistance element on a ceramic substrate	High quality trimpots, and some conventional panel mount types (not very common). Low noise, and high stability. Relatively limited life (200 operations typical for trimpots)	0.25 to 2W (or more)
Conductive Plastic	Special impregnated plastic material with well controlled resistance characteristics	High quality (audiophile and professional) pots, both rotary and linear (slide). Excellent life, low noise and very good mechanical feel	0.25 to 0.5W

Wire wound	Insulating former, with resistance wire wound around it, and bound with adhesive to prevent movement	High power and almost indefinite life. Resistance is "granular", with discrete small steps rather than a completely smooth transition from one resistance winding to the next. Low noise, usually a rough mechanical feel.	5 to 50W (or more)

Bear in mind that the above list is a rough guide only, and is not intended to be be the "last word" on the different resistive types or their characteristics. In all cases, if you really want to know the full details about any one of those listed, get the manufacturers' data for the pot - it will be a lot more accurate (and specific) than the brief explanations above.

In addition to the resistive materials, there is also the physical type of pot. I am not going to describe the size and shape, but how the pot is configured mechanically and electrically.

Actuator	**Configuration**	**Type**	**Typical uses**
Rotary	Single gang	Single turn	Single channel controls for monoblock amplifiers, guitar amps, or anywhere that a single control is sufficient for the application.
Rotary	Single gang	Multi turn	Precision trimpots for critical applications. The resistance range is covered in anything from 10 to 25 turns of the screwdriver slotted actuator. There are some multi turn panel pots, but these are quite rare and expensive. Multi turn dual pots are also very uncommon.
Rotary	Dual gang	Single turn	Stereo applications, or anywhere it is desirable to change two separate resistances at once. Nearly all dual gang pots have equal resistances and tapers, but it may be possible to rebuild a dual gang pot using intestines removed from another pot of the same make and type (not needed very often though)
Rotary	Dual concentric	Single turn	Commonly used in (cheap) car radios and some consumer goods. These feature dual concentric shafts, allowing a single pot position to provide (for example) volume and "tone". The knobs are designed to fit the separate shafts (which are usually of different diameters). Almost impossible to buy from retail outlets or manufacturers in small quantities. (Usually to special order)

Linear	Single gang	Slide	Commonly used as "faders", unless they are of high quality, best just called slide pots. They are available in a variety of lengths, from 30mm to 100mm or more of linear travel. True faders will normally be relatively long, and generally are conductive plastic (and rather expensive :-(
Linear	Dual gang	Slide	As above, but for stereo mixers. Otherwise identical comments apply.

Again, this is a simplified listing. If you are willing to pay for 10,000 units, most pot makers will quite happily build you a triple gang pot with unequal resistances and different tapers, or an eight gang pot so you can build a variable stereo crossover network. In fact, almost any configuration is possible, but for various reasons may not be plausible or sensible.

Nearly all manufacturers and distributors have settled on a limited range of "standard" values and types, based on the most common uses for their products. That other configurations used to be available but were withdrawn due to lack of consistent sales is a lamentable fact, brought about by "economic rationalisation", which basically means that if they don't sell them in good quantities, they will be neither made nor stocked by anyone (unless you are happy to pay through the nose, of course).

Most problems of this type can be solved by throwing money at them until the problem disappears, but few of us can afford this approach - besides, I think the military establishments of the world have a patent on that method :-)

A standard single gang pot is shown in Figure 5. The important external bits are shown so you can refer to them as needed. I have (somewhat arbitrarily) numbered the terminals as 1, 2 and 3. Terminal 2 is the wiper. For a "standard" volume control application, 1 is normally connected to ground, the input is applied to 3, and the output taken from 2 (wiper) allowing the output to be varied from ground (no signal) to input (maximum signal).

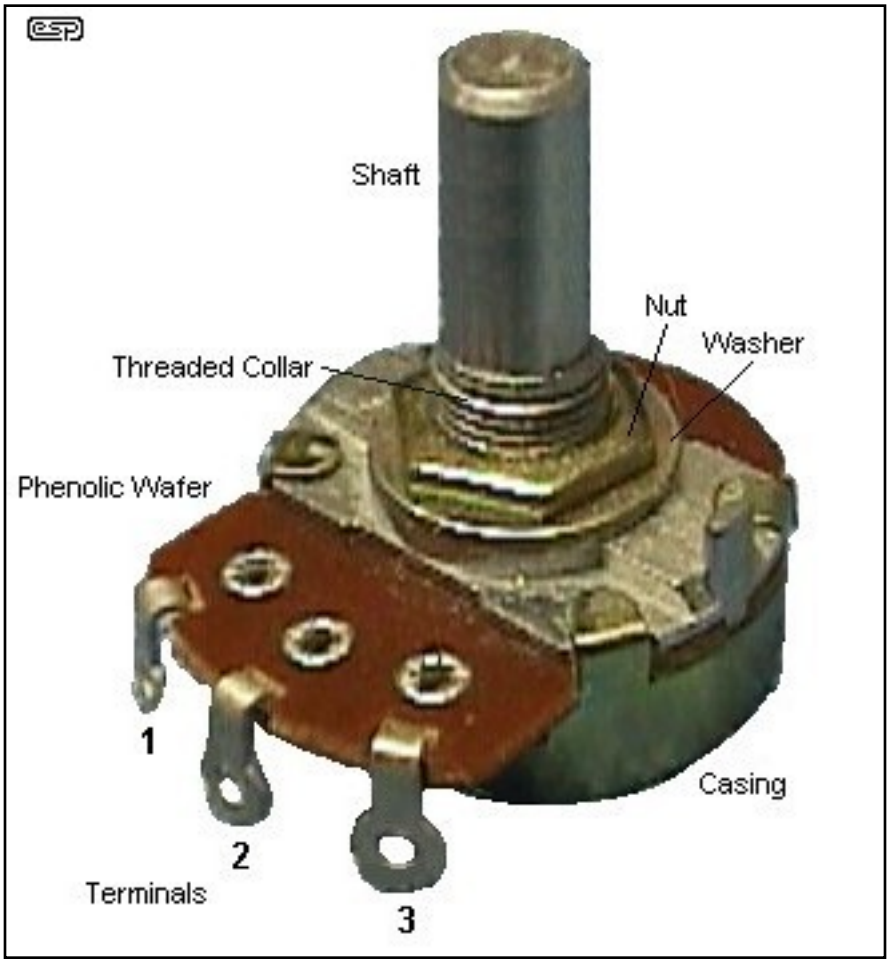


Figure 5 - Single Gang Pot Detail

In addition, there are a few odd-ball additions to the list. These include pots with integral switches (as used in small transistor radios - hint as to where to get one if you need it badly enough). The switches may be rotary, so in the minimum volume position, the switch is off, or they may use a push-pull switch. Older car radios often use a combination switch and dual concentric pot, so that power, volume and tone can all be controlled with one knob complex.

As before, the possibilities are almost endless, limited only by imagination and budget. There are few mechanical constraints that will prevent a special design from being feasible, although expecting accurate tracking on a 100 gang pot might be asking too much :-) Could it be made though? But of course - leave your large sack of money at the door, sir, and come on in.

Oh yes, I almost forgot. Motorised pots. Standard (or high quality) rotary or slide pots that are driven by small DC motors to allow remote control. Even a cheap pot will usually outperform an expensive "digital" volume control, with the added advantage that it can be operated by hand or with the remote. These are quite common, and even some of the (relatively cheap) Chinese made subwoofer "plate" amps use them for

remote control.

Ah! Another one ... Most pot "gangs" are 3 terminal types, but there are some with a tapping partway along the resistance element. This was used in the bad old days to create a "loudness" control, where the bass and treble are increased at low levels to compensate for the way our hearing reacts to different levels. Because there was rarely (if ever) any attempt to match the acoustic power levels, the loudness control was always wrong. To get it right requires source, preamp, power amp and speakers to have a known gain/ sensitivity, and ideally a preset control would have been incorporated to ensure the system could be calibrated. This was never done by the vast majority of manufacturers - Yamaha appears to be the *only* maker who even made an attempt (I don't know how good it was, never having seen a system that used it).

Putting Your Pot to Use

Well, that part is simple, isn't it? Judging from the number of e-mails I get asking about how to wire pots, the answer is obviously "no". Being 3 terminal devices (for a single gang), there are quite a few different ways that they can be wired. Connection to a single terminal is rather pointless, so at least that eliminates three "possibilities". At this point, a diagram is needed ...

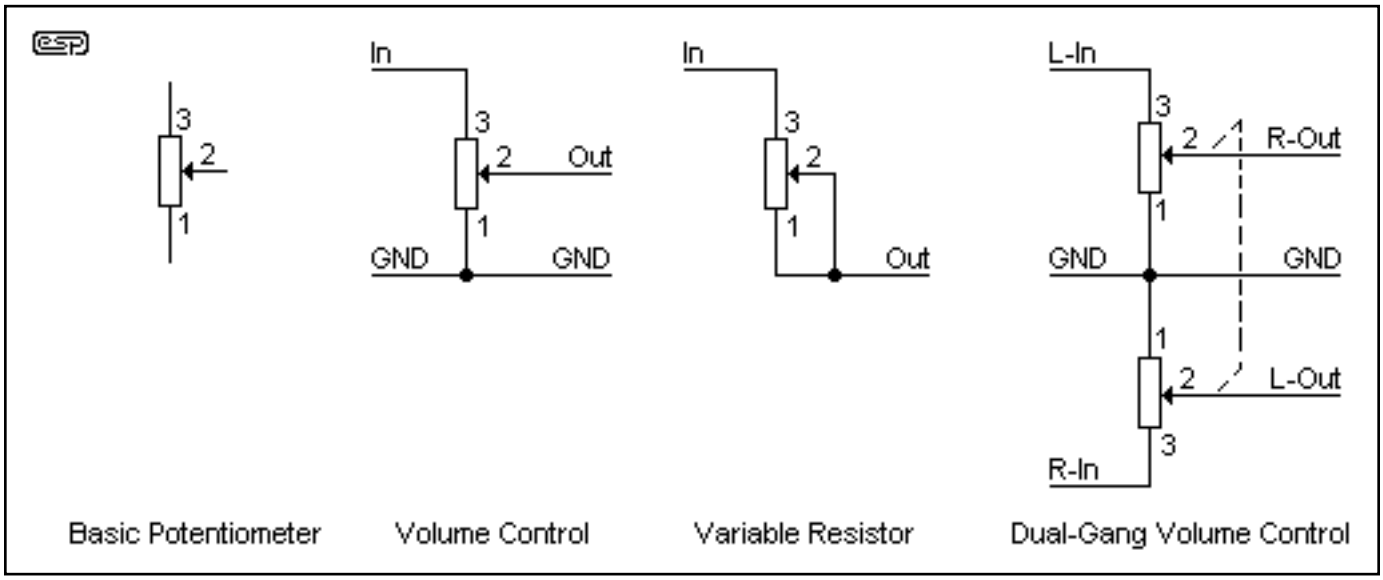


Figure 6 - Potentiometer Terminals and Connections

As shown in Figure 6, a pot is usually wired using all three terminals, and I have used the same numbering scheme as in Figure 5. One terminal (1) is earthed (grounded) for use as a volume control - the most common usage. This allows the wiper to be turned all the way to zero signal for maximum attenuation. Note that if the earth terminal were

to be left disconnected, all we have is a variable series resistance, whose effectiveness will be minimal in typical circuitry. This is still a common usage however, but for different reasons (see below).

Turning the shaft clockwise (CW - by convention, to move pin 2 physically closer to pin 3, and increase (for example) volume) will select a different point along the resistance element, and forms a voltage divider, so the attenuation of the signal is proportional to the rotation of the shaft. At the fully clockwise position, there is close to zero ohms in series with the signal, and the full resistance of the pot to earth. Attenuation at this setting is zero (assuming a zero or low impedance source - this is often overlooked!), and this is full volume (maximum signal level).

The source impedance should normally be no greater than 1/10th (0.1) of the pot's stated resistance. Further, the load resistance or impedance should be 10 times the pot's resistance to prevent the taper from being adversely affected. You may (of course) be deliberately loading the pot as described below, but the following stage must still present a high impedance unless its impedance has been included in your calculations.

The second form of connection is a variable resistor. Not usable as a volume control, but still extensively used for other applications. It is common (and preferable) to join two of the leads together - the wiper, and one end or the other. The actual connection depends on what you are trying to achieve, and since there are so many possibilities, I won't even try to explain them all. When used in this mode it is most commonly refererred to as a variable resistance or resistor - the word rheostat is somewhat dated (to put it mildly) and is not a term that I use in any of my articles.

To get an idea of the different configurations that are in common use, have a look at the ESP Projects pages, and those on other web sites. The number of possibilities is actually not that great, but people use different conventions as well. For example, in Australia, we use the term "anti-clockwise" or ACW. In the US, this is "counter-clockwise" or CCW. At least the term clockwise seems to be common to both countries :-) Naturally enough, these are only two conventions, and I am unsure of the terminology in other countries - especially if they don't use English (and why would they, if they already have a language of their own).

As a completely irrelevant side issue, the Web is changing this quite quickly, as the majority of web sites are in English.

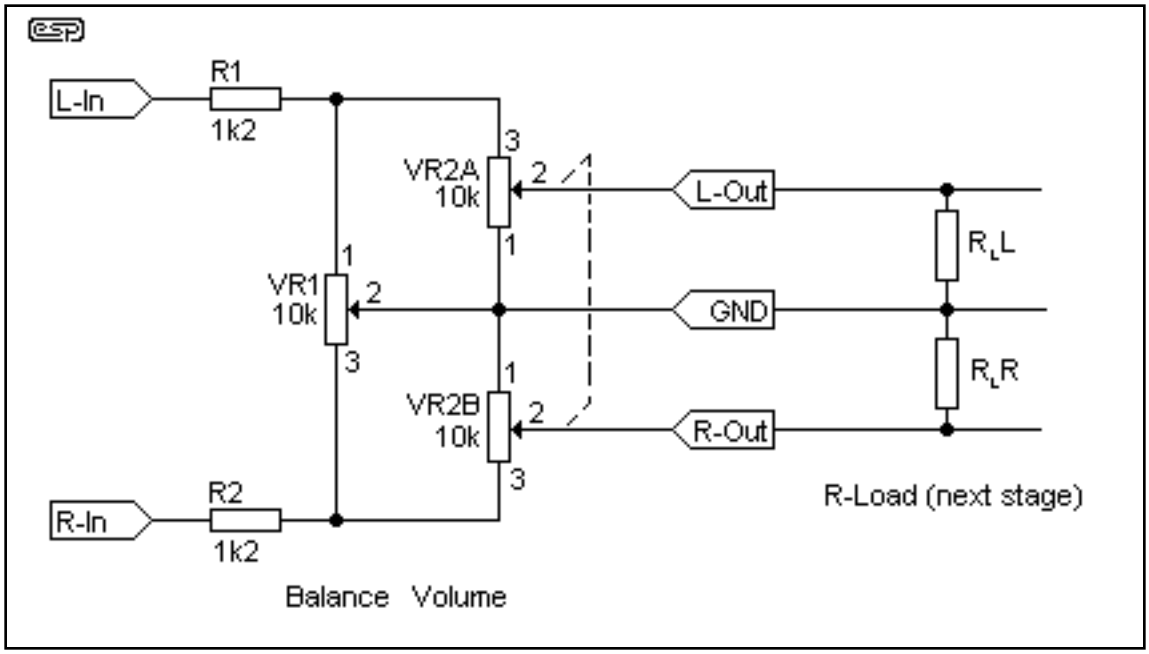


Figure 7 - Volume and Balance Controls

Figure 7 assumes the use of a log pot for volume. The balance control can be done in many different ways, with that shown being but one. Quite a lot of Japanese equipment uses a dual gang pot for balance, but the resistance element only goes for half the travel. When set in the centre position, there is no loss at all, and rotation in either direction attenuates the appropriate channel, but leaves the other unaffected. This is yet another type of custom pot, made for a specific purpose. I know of no manufacturer that sells such an item through the normal distribution channels, so home builders have to come up with different ways to achieve the same (or similar) things. The balance control as shown above (with the values shown) will give a response very similar to the more complex version described in the next section.

Changing the Law of a Pot

Using pots can be done in the conventional way, or you can get adventurous and achieve a lot more. A good example is the "Better Volume Control" shown in Project 01. The other ideas presented also show how you can make modifications to the way a pot behaves, just by adding a resistor (R). The "ideal" value by calculation is 22k for a 100k pot, and this gives a maximum deviation of +1.58 and -1.7dB from a real log curve. This is contrast to the original article, where 15k was suggested, and although the error is greater (+2.89dB and -1.12dB), the overall behaviour is almost ideal in listening tests.

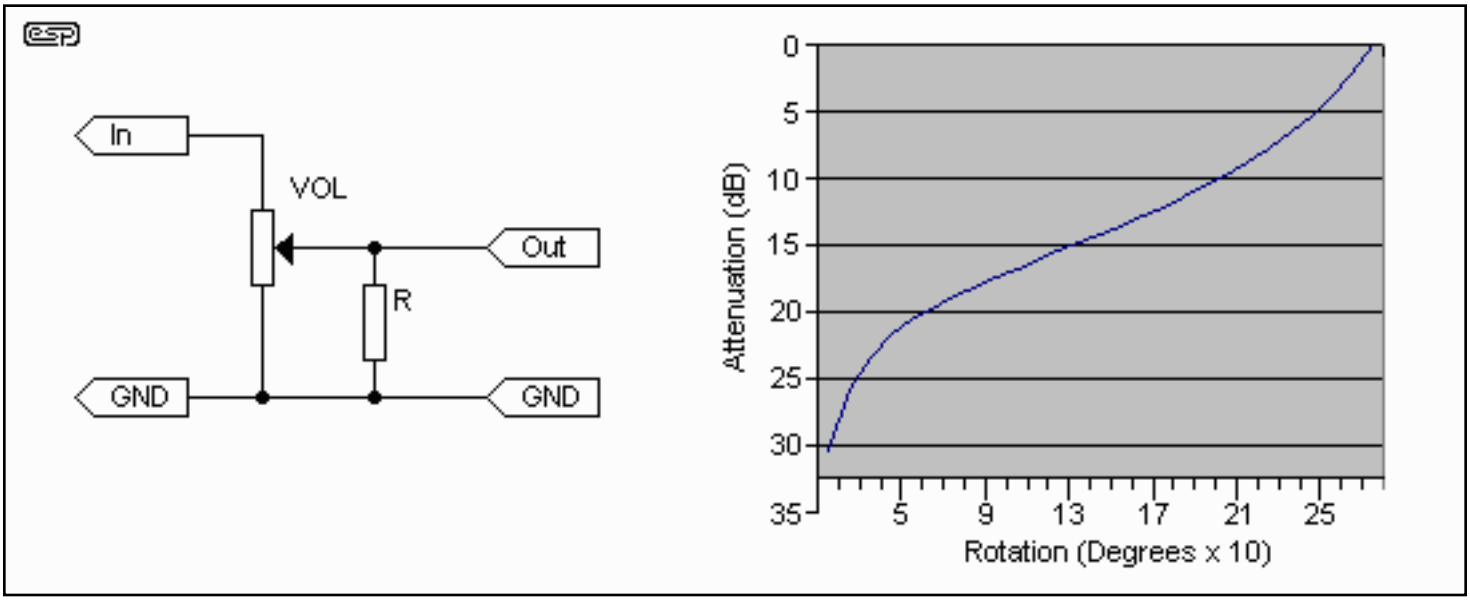


Figure 8 - A Better Volume Control

Take a look at the balance control (below) as an example. The conventional balance control requires either a log/antilog pot (virtually impossible to obtain), or one of the special types commonly used in Japanese consumer hi-fi gear. About the only way you'll get one of those is to remove it from the equipment - again, they are virtually impossible to get from normal hobbyist suppliers.

Add a couple of resistors to a dual gang linear pot, and the problem is solved. Not only is the pot heavily "centre weighted", but will also maintain a relatively constant sound level as the balance is changed from full Left to full Right. The centre weighting means that for most of the pot's travel, the balance is shifted subtly, so it provides a very fine resolution around the central position - there is little requirement for only one channel (other than testing), but that is still available. In short, lots of benefits, and few drawbacks.

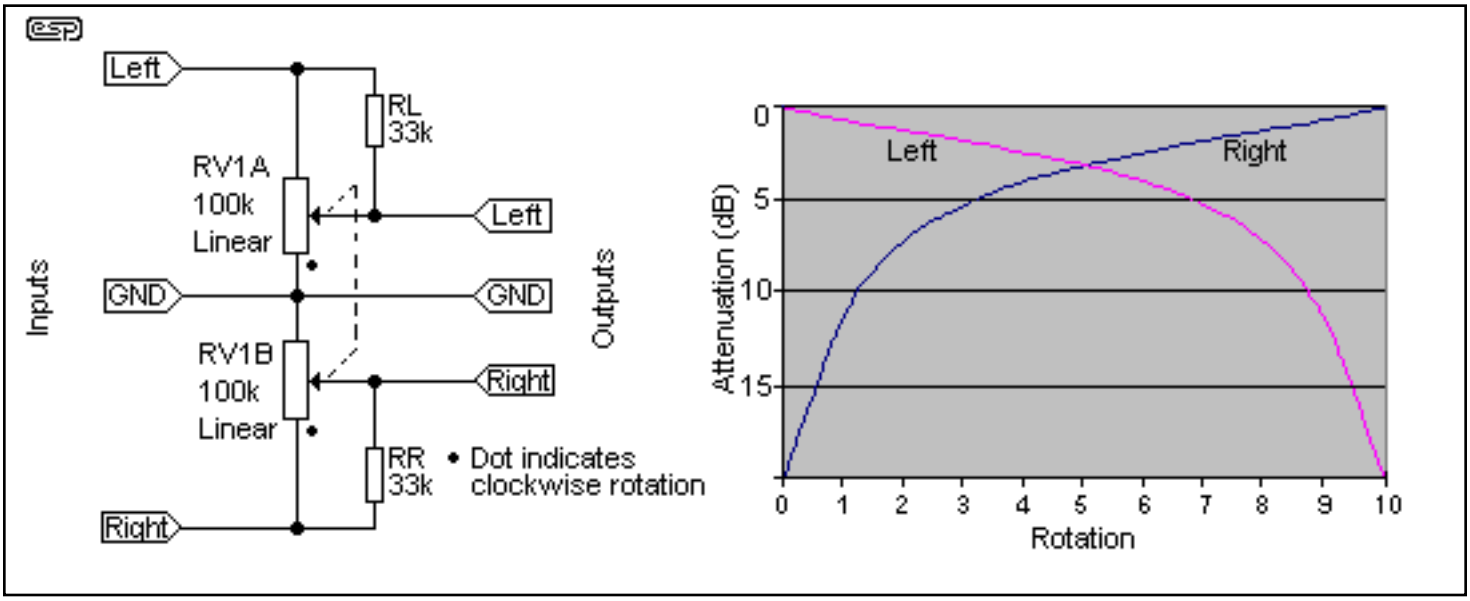


Figure 9- Centre Weighted Balance Control

Needless to say there are many other configurations that can be used, and this is but one. The resistor value (RL and RR) is fairly important - it really should be 35k for a 100k pot, but the error when using 33k is minimal (about 0.16 dB at centre position).

One of the goals of circuit design is to utilise available components. This is not necessary if you make 10,000 of something, since at these quantities special orders will cost little or no more than the normally available components. When you are making one for yourself (or perhaps two - one for a friend for example), specially designed components are not an option due to the setup costs (this could easily be thousands of dollars / euro / pounds). Even in quantities of several hundred, available components are still (usually) cheaper.

The balance control above is an example of a dual log / reverse log pot, created with a standard dual gang pot and a couple of resistors ... and it works better than a commercial offering is likely to - even if you managed to find one.

For more information on this configuration, see Project 01. Note that as shown, the balance control here is not optimised for any significant impedance at the output, so its performance will change if you connect a volume control to the output.

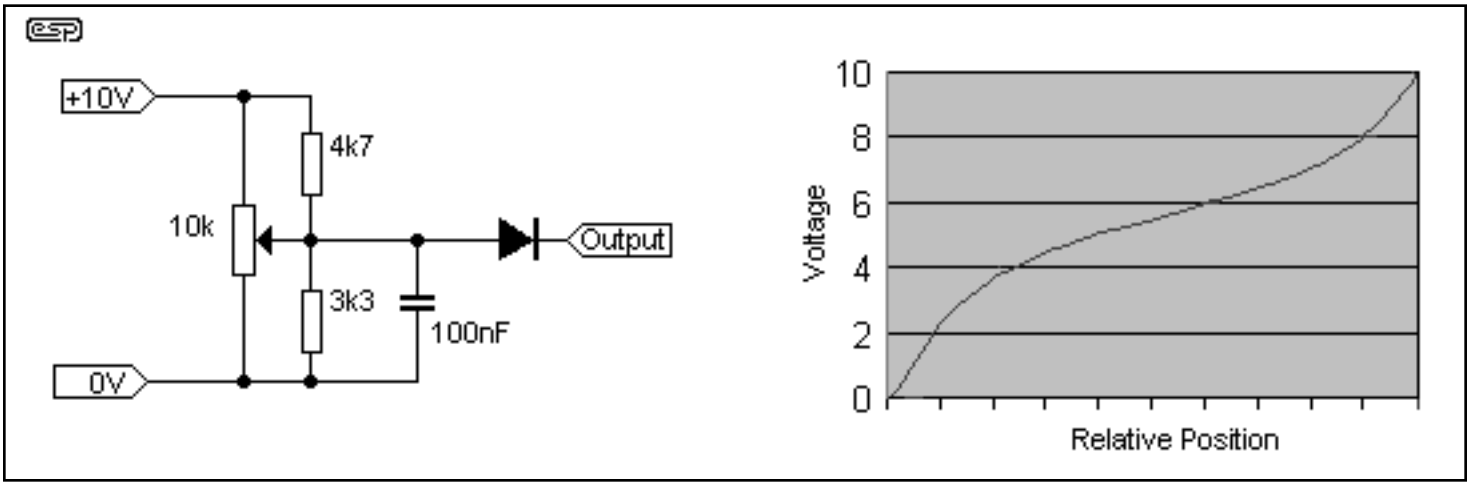


Figure 10 - Creating an "S" Curve for Lighting

Another example of modifying a pot to make it do what you want is shown in the LX-800 Lighting Controller. The faders need an "S" curve, to compensate for the non-linear behaviour of lamps and our eye's sensitivity to light levels. This is also achieved with a couple of resistors across a normal linear pot.

If you don't like the shape (because of the lamps you use, for example), you can simply change the resistor values and modify the curve to suit your exact needs. Since even ordinary log pots are not actually logarithmic anyway, can you imagine getting a pot that would give you an S-Curve? Even worse, if you found that it was not suited to certain lamps, then you would be hard pressed to modify the law to get what you needed. In some cases it would be impossible.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2002/ 2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 22 Jan 2002./ Updated Jan 2003

Elliott Sound Products **Amplifier Power Ratings**

Page Last Updated 25 April 2005

Main Index

Articles Index

This article was originally fairly superficial and frivolous, but has been expanded (a little) to explain the matter better. Amplifier power ratings are usually honest in Hi-Fi equipment, but become very silly when it comes to the 'mass market' systems.

'Exceedingly silly' happens when you look at computer speakers and 'boom boxes' - most of which boast power ratings that (if true) would do a fine job of amplification for a large hall or small stadium. This is quite obviously not the case, as anyone who has used them is aware - well below the point of mild discomfort, it is obvious that distortion is abundant, and the sound (almost literally) falls to pieces. Perhaps on this basis, they should be referred to as power *rantings*

I have a set of computer speakers that are rated at 480W PMPO (yes - four hundred and eighty Watts). I have measured them at less than 5W each before clipping. There is no rhyme or reason that can explain such a difference, except ....

Power Ratings In The New Millennium (and Beyond)

Much has been said - and will no doubt continue to be said - about amplifier power ratings. There has been a disturbing tendency over the last few years to revisit the bad old days where terms such as PMP (Peak Music Power) and PMPO (Peak Music Power Output) have once again raised their ugly heads.

Admittedly, these 'new' power rating are not used by hi-fi manufacturers, other than in

the low-end equipment. These new terms are soundly (no pun intended) based on the science of *marketing*, having nothing to do with actual science or physics. PMPO is mathematically expressed as

$$PMPO = P_{REAL} * k$$

where $\mathbf{P_{REAL}}$ is the real power as calculated by the formula below, and **k** is a constant whose value is approximately equal to one's grandmother's age, divided by the square root of the distance from the office to the advertisement writer's normal place of abode - measured in millimetres, inches, furlongs, statute miles or pounds per square inch (as appropriate) to provide the number you first thought of.

In the unlikely event that the value of **k** cannot be calculated from the above formulae to provide a totally meaningless (but plausible-looking) final result, a randomly selected value of between 20 and 75 should be used.

An alternative (and equally useless) way to measure PMPO is to multiply the supply voltage by the instantaneous short-circuit current from the amplifier - the amplifier does not have to survive this test, and the current only has to exist for around 1us to obtain a satisfactory rating ...

$$PMPO = V_{supply} * I_{peak}$$

where $\mathbf{V_{supply}}$ is the total supply voltage and $\mathbf{I_{peak}}$ is the instantaneous (and possibly destructive) short-circuit current. Given that a 12V plugpack (wall wart) supply may be capable of perhaps 30A for a microsecond or two, this is perfectly acceptable, and gives a PMPO of ...

$$PMPO = 12 * 30 = 360W$$

This is a perfectly acceptable figure, and may be used with gay abandon (in the hope that someone will actually believe it).

Thus (using the first equation) we can now compute the power of an amplifier which manages to impress a voltage (which need not be sinusoidal - an harmonic distortion of up to 400% is considered perfectly acceptable - albeit mathematically impossible) of 8V across a speaker of 8 ohms. Actual (real) power may be calculated by

$P_{real} = V_{RMS} * I_{RMS}$

Since 8V across an 8 ohm load provides 1 Ampere of current, we obtain

$P_{real} = 8 * 1 = 8$ Watts

or, using the power formula

$P_{real} = V^2 / R$ (Voltage squared, divided by impedance in ohms)

thus

$P_{real} = 8^2 / 8 = 64 / 8 = 8$ Watts

PMPO may now readily be calculated, using a median value of 45 which results in a totally satisfactory advertising power (P_a) of

$P_a = 8 * 45 = 360$W PMPO

You will notice that by fiddling with the figures to suit my goal, I have been able to make the PMPO figure the same, using two completely different 'test methods' and 'formulae'. Therefore, the figures *must* be correct, and the result must therefore be genuine .

What ? You think I'm *lying* ? Well spotted gentle reader - the whole process is unadulterated horse-feathers.

It goes without saying that using the complete formula, the final P_a rating could be anything from 500uW (totally unacceptable from a marketing perspective) to several Mega-Watts. Although figures in this latter range have considerable merit, it is probable that even the most gullible of Boom-Box or computer speaker buyers will be a little suspicious, especially when the plug-pack power supply offered (as an option) announces that it provides a mere 1A at 12V (or 12 Watts DC power output - as much as 7 (legitimate) Watts of audio may be obtained from such a supply).

True, this figure will not be comprehensible *per se*, but suspicions may be aroused when a friend's genuine 20W system completely drowns them out with sound. A 160 Watt rated unit's apparent lack of power by comparison is easily explained by the fact that

"This tape/CD/FM radio station was recorded at really low level" or some similar self-delusion.

This is a little more difficult to shrug off nonchalantly if one's Ghetto-Blaster were to be rated at 1.6MW for example. Even those who pay more for their sneakers than others might spend on a tailored suit (I think i may have had one of those ... once), or a real Hi-Fi system, will be forced to wonder why their unit was not supplied with a small nuclear power station to achieve such power.

It is worth noting that it is possible to merely think of a 'good' (i.e. impressive looking) number, call it Watts (PMPO), and use that instead of the potentially tedious mathematical approaches above. This method is just as invalid as the more technical methods described, but is not as much fun.

Warning

There are sites on the net where the author(s) seems to think that there is some logic (however perverse) in the PMPO rating. For example it has been claimed that PMPO is calculated as the maximum instantaneous power an amplifier can deliver, albeit under unrealistic conditions. Such claims are as false and misleading as PMPO - there is *absolutely zero* logic or science involved in making up a PMPO figure, as demonstrated by the following simple exercise.

Let's assume that an amplifier has a supply voltage of 14.8V DC. A little under 15V (unloaded) is not uncommon for so-called 12V plug-pack supplies. For the sake of the exercise, we shall further assume that the manufacturer used a 3,300uF filter cap (this is probably unrealistically high, but will do for the time being). The energy storage of the cap is measured in Joules (Watt/seconds). One Joule is 1W delivered for 1 second. For our 3,300uF cap, charged to 14.8V, we get ...

$$E = \tfrac{1}{2}C * V^2 = \tfrac{1}{2}3{,}300^{E-6} * 14.8^2 = \tfrac{1}{2} = 0.361 \text{ Joule (0.36W/s)}$$

This number is obviously of no use as is, but if we assume that the cap is discharged in 1ms, that gives an instantaneous figure of 360W PMPO, which is much more satisfactory. Goodness me, I seem to have used yet another (almost completely) bogus formula to arrive at the number I first thought of .

Note: The formula for energy in Joules is correct, as is the conversion from seconds to milliseconds. The bogus part is the simple fiddling with numbers to arrive at the answer I wanted.

It pretty much goes without saying that I can think of several other equally meaningless equations that will also give the figure of 360W, but there is absolutely no point in doing so. Just remember that if you see PMPO 'power' listed for an amplifier, the figure is false, and has no meaning at all. Before purchase, try to locate something on the package or in the instruction page that specifies the power in RMS or DIN. Failing that, obtain written assurance from the sales person that the claimed power is real, test it when you get home, then return the product next day. Explain that you were assured that the amplifier was indeed 1kW, but you were unable to obtain more than 5W from it - therefore, the product must be faulty.

Advertising, 1970's Style

In the above mentioned bad old days, there was still a modicum of perverse logic used to calculate 'Power'. Advertisers (after consulting - sorry, interfacing - with someone who could count to more than 10 with their shoes (or trousers) still on), would use the peak value of the RMS voltage (Volts * 1.414), or the more adventurous could even use peak-to-peak (double the peak value).

Using the same value of voltage and impedance as above (namely 8V RMS, 8Ω load), one can calculate the P_{A70} (Advertising Power, as used in the 1970's) to a high degree of uselessness, thus

$$P_{A70}(\text{Peak Music Power}) = (8 * 1.414)^2 / 8\ \text{Ohms} = 16\text{W}$$

Or ...

$$P_{A70}(\text{P-P Music Power}) = 16 * 1.414)^2 / 8\ \text{Ohms} = 64\text{W}$$

Technology has certainly come a long way since then, as is now apparent.

But Seriously, Folks

If, from the above, you have deduced that I am less than favourably impressed by such deceptions, you are correct. Indeed, the term "RMS" power is just as grating to an engineer, since there is no such thing.

'RMS' Power

Power is simply the product of RMS Volts and RMS Amps, and the resulting figure is 'power'. Not 'RMS Power' - or any of the insane derivatives described above - just 'power'.

The term RMS (Root Mean Squared) can only be applied to voltage or current. The RMS value is determined to be the Alternating Current (AC) equivalent of a Direct Current (DC) which creates the same amount of heat in a load.

For a sine wave, this is the peak value, divided by the square root of 2 - i.e. 1.414 (I shall not bore you with the exact reason for this, but it is a scientifically and mathematically accepted fact).

For a perfect square wave, it is the peak value alone, since if the positive and negative peaks were to be rectified (so as to be the same polarity), the result is DC. This condition (or at least close to it) is quite common with guitar amps (the distortion is part of the sound), but should never occur in Hi-Fi, even briefly.

Power should only ever be measured with no clipping. When an amp clips, there is more available power, but higher distortion. It is not uncommon to see amplifier powers rated at 10% distortion. This is quite unacceptable, as this indicates that there is severe clipping of the signal. A good quality amplifier will have less than 0.1% distortion just before clipping, somewhat higher for push-pull valve amps, and a lot higher for single ended triode valves.

When I refer to power in any of my articles, common usage shall prevail, and I (like many others in audio) will reluctantly accept the term RMS Power to mean power. All amplifier power ratings in the project pages (and elsewhere) are 'RMS' unless otherwise stated.

Music Power

The music power of an amp is real, and is generally higher than the continuous power. It is measured by using a tone-burst generator, and is the peak power than an amp can supply for (typically) about 10ms. This is quite reasonable, but not terribly useful when it is examined carefully. Since music is very dynamic, with the peak amplitude exceeding the average by 10 to 20dB (depending on the type of music), an amplifier is never called upon to provide full power all the time (at least if clipping is avoided, which should be all the time).

If the power supply is regulated or has considerable excess power capacity, the continuous and music power ratings will be almost identical. The difference was (at one

time) measured, and was called 'dynamic headroom'. Few amps have a dynamic headroom of better than 1 or 2dB, and the greater the headroom, usually the cheaper the power supply for the rated power.

An amplifier with a much greater music power than its 'RMS' power usually has a transformer and/or filter capacitor that is too small. In most cases, a 90W (RMS) / 100W (music power) amp will not sound louder than a 90W amp with a regulated supply (so RMS and music power are the same). The extra 10W represents a little under 0.5dB, which is barely audible in a comparative listening test.

C# - And on that note ....

Main Index

Articles Index

Note: Although the above is slightly tongue-in-cheek **(SLIGHTLY??)**, it is meant to be taken seriously - this rubbish really happens - just look in the local papers, and on the cartons for computer speakers if you don't believe me!

This epistle is Copyright (c) 1999-2004 Rod Elliott - All Rights Reserved (Yeah, like who else would want it? :-)

The Audio Pages

Elliott Sound Products **How to Wire a Power Supply**

Copyright © 2003 - Rod Elliott (ESP)
Page Created 07 Jun 2003

Articles Index

Main Index

Contents

1.0 Introduction
2.0 How to Wire a Power Supply
    2.1 Schematic Connections
    2.2 Physical Connections
3.0 Additional Components
4.0 Mains Connections
5.0 Fuse Holder
6.0 Mains Switch
7.0 Earth Connection
8.0 Mains Capacitors

1.0 Introduction

This article does not attempt to cover general household or commercial wiring practices - only the internal wiring needed for electrical safety and making your power supply work are covered. For detailed information on wiring practices, you must contact your local supply authority/ company, or obtain a copy of the wiring rules for your country or locality. I am unable to assist with this, as it is highly country specific, and in many countries is also heavily regulated and/or legislated.

Warning: Household electrical current is extremely dangerous, and it may be illegal for you to perform your own wiring, even for equipment that connects via a standard wall outlet. If you are unsure of the procedures, terminology or anything else that may cause a potentially fatal error due to oversight or lack of knowledge, you must seek assistance from a qualified electrician. Remember that if someone is killed or injured as a result of your work, you may be held liable and subject to severe criminal and financial penalties.

2.0 How to Wire a Power Supply

I have been asked many times about PCBs for power supplies for amplifiers. I do not recommend using a printed board for a number of reasons, and these are as follows ...

1. Limited range of capacitor values: When the constraints of a PCB are imposed, the capacitors you choose must be the same physical size as those the board can accommodate. This restriction is so great (IMO) that this precludes the use of a board for almost any DIY power amplifier project.

2. Electrical Characteristics: The normal copper thickness on a printed board is not really sufficient to ensure that there is minimal resistance, so there is a greater likelihood of hum (or buzz) and efficiency may be marginally reduced. In contrast, hard wiring can be as thick as the constructor likes (although it is still necessary to be able to solder to it and the capacitor terminals).

3. Physical layout: A printed board limits the flexibility to mount capacitors in the most convenient place. This is probably one of the most compelling reasons to use hard wiring, since multiple capacitors may be best arranged in a row or a block, depending on the internal construction of the chassis mounted components (e.g. transformer, heatsinks, PCBs, etc.).

4. Other components: A PCB is less than ideal (to put it mildly) for mounting a bridge rectifier, and doubly so if a 35A chassis mount bridge is used. These need to be on a metal panel to obtain heatsinking, and it is very hard to achieve this when a PCB is used.

2.1 Electrical Connections

So, what is the constructor to do? Hard wire the power supply is what. A basic configuration may look like that shown in Figure 1 - this is the schematic for a general purpose supply, suitable for a high end hi-fi power amplifier.

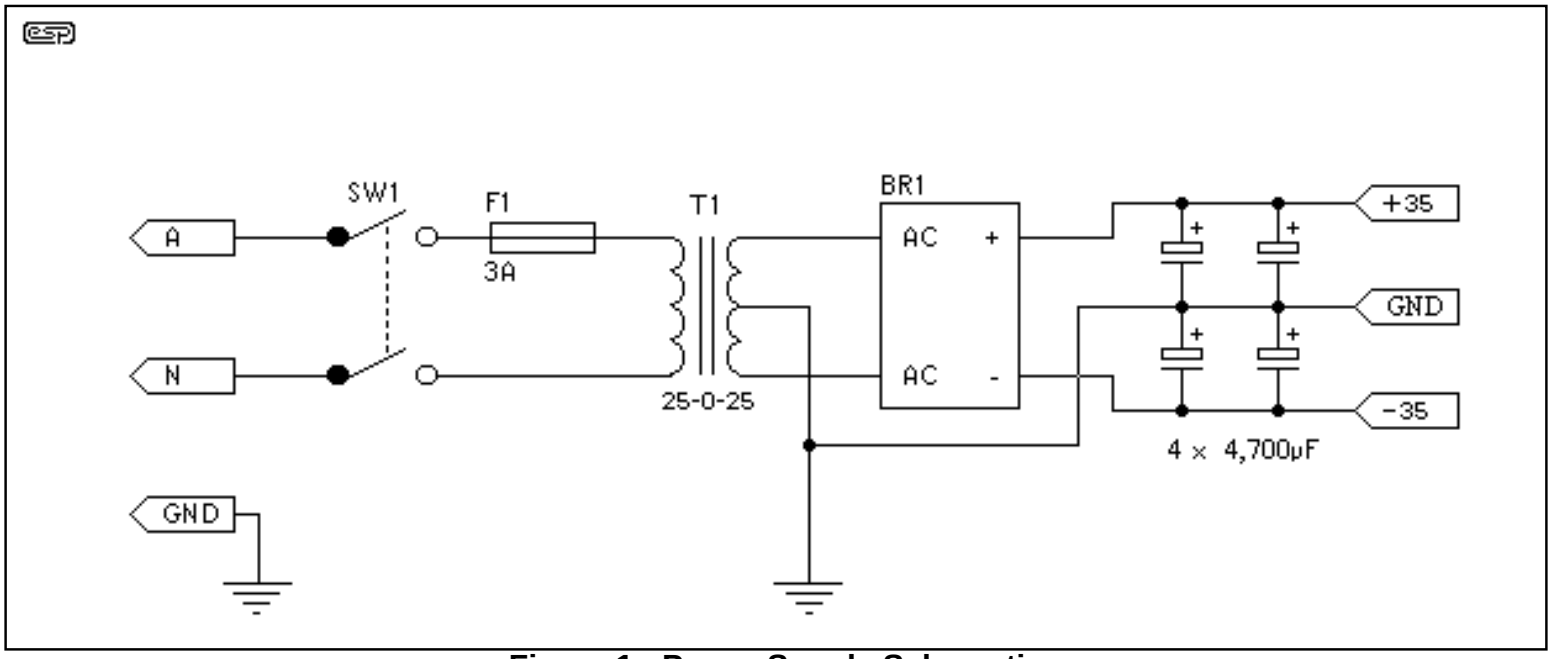
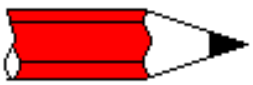


Figure 1 - Power Supply Schematic

This is all well and good, but these are the electrical connections, and have no direct relationship to the physical connections needed. For the purposes of the exercise, we shall assume an IEC mains connector, chassis mounted mains fuse, and a power switch mounted on the front panel.

The diagram in Figure 2 shows one possible physical arrangement of the supply components. If you need to re-arrange the locations, it is a relatively simple matter to move things where you need them, while maintaining the required electrical connections. Some construction articles (and especially kits) will try to enforce a specific layout, but this will not always suit the constructor - especially if s/he happens to have a whole box of 2,200uF capacitors at their disposal (wishful thinking for most :-) where the bill of materials calls for 10,000uF.



It is extremely important that the DC is taken from the capacitors, and *not* the bridge rectifier. If DC is taken from the bridge, it will be noisy, and this can easily get into the amplifier, degrading the signal to noise ratio - particularly under load! The noise may not be heard directly, but will add unwanted signals to the music which may sound "hazy" or "clouded" as a result.

2.2 Physical Connections

The "stylised" drawing below shows how the various components should be connected together. From this, it is possible to extend the basic idea very easily. The diagram assumes that the constructor will use 4 filter caps in a series-parallel arrangement (2 for each supply rail), a 35A chassis mount bridge rectifier, and a toroidal transformer. The colour coded transformer leads are for indentification - they are ***not*** intended to be taken literally. Refer to the manufacturer's specifications to make sure that you get the correct colours! IEC (European)

mains colour coding has been used, and this is now almost a world standard, so should be painless for all. The older standards are also provided below.

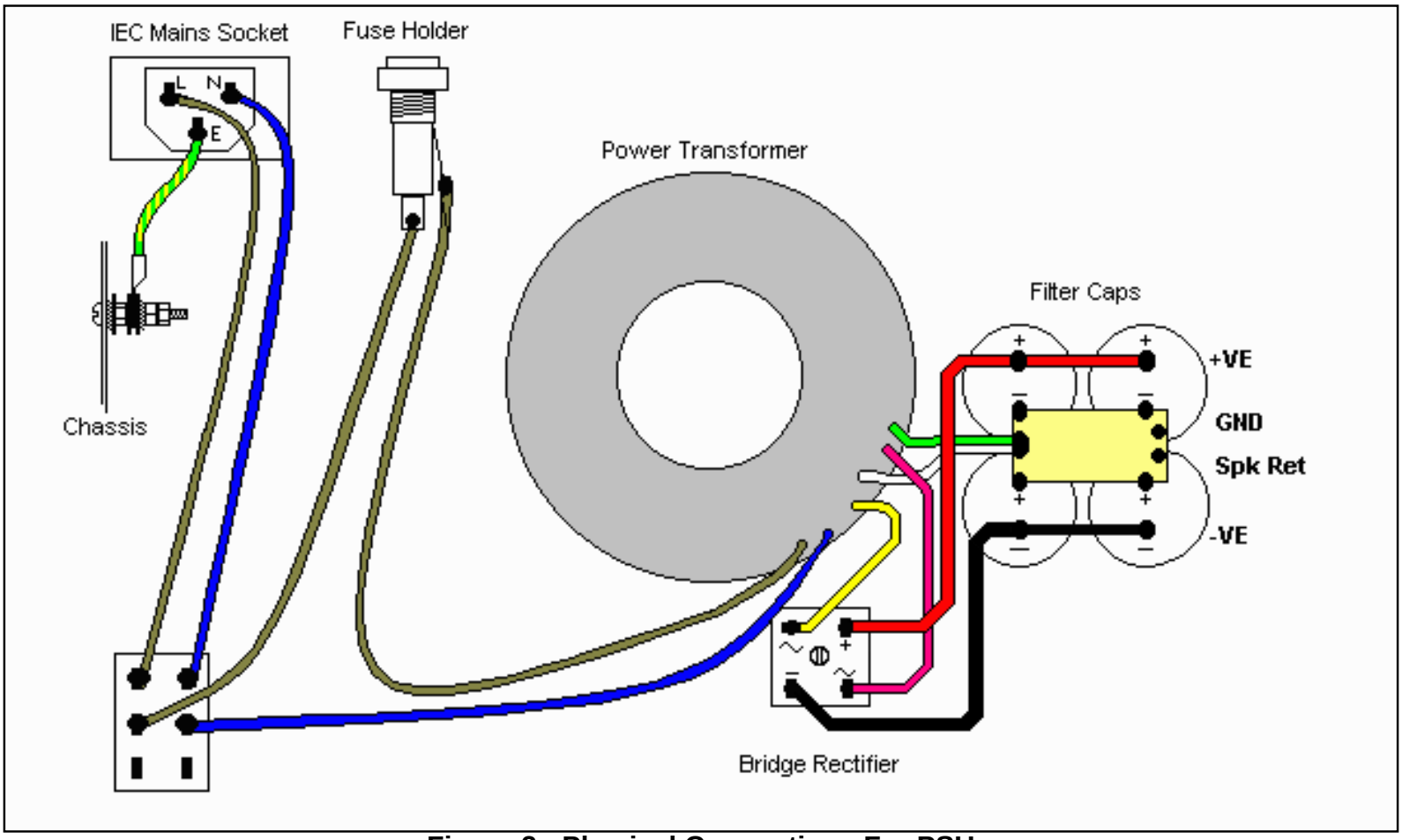


Figure 2 - Physical Connections For PSU

Note that I have not shown the required sleeving over all mains connections. This is essential for electrical safety, and usually just means the the required heatshrink tubing is placed over the wire(s) before attaching and soldering. Make sure that soldering does not heat the tubing, or it will shrink before it is properly located. To this end, make sure that the tubing can be located at least 25mm (1") and preferably more, back from the solder connection.

Rubber boots are available for IEC chassis mount connectors, and large heatshrink can be used to completely encase the fuseholder. Don't forget to feed the wires through the heatshrink or rubber boot before soldering!

As well as proper safeguards against accidental contact with the mains, it is also extremely important to keep mains and low voltage wiring well separated. This means either physical separation, or reinforced insulation between the two sets of wiring. If physical separation is used (and this is the most common and easily achieved), make sure that wherever possible, the minimum distance is 25mm. It should not be possible to squeeze or otherwise coerce the primary (mains) and secondary (low voltage) together under any circumstances. All wiring must be secured using cable ties, and suitable chassis anchors may be needed in some cases to

ensure that all wiring remains properly separated.

With most toroidals, all leads come out of the over-wrapping in (more or less) the same general area. Normally the insulation provided is sufficient to ensure safety, but some additional heatshrink tubing will not go astray if the leads are close together.

3.0 Additional Components

The diagrams above show only the basic parts needed. Other components are used routinely for lower noise, capacitor discharge, etc. A diagram of these extras is not needed, but a brief discussion of them is certainly warranted.

Low Value Capacitors

It is very common to use 100nF or so polyester, mylar or polypropylene capacitors in parallel with the filter capacitors. The use of these ensures that the impedance of the power supply remains low at all frequencies up to several megaHertz. While it will cause no problems to use these components at the power supply itself, most amplifier PCBs will have provision for them on the board. This also ensures that lead inductance between the supply and the amplifier is dealt with. Many power amplifiers also have on-board electrolytic caps - effectively in parallel with the main filter caps.

In addition, capacitors (typically 100nF) are often used across the bridge rectifier, effectively in parallel with each diode. These may help reduce noise, but normally, a properly designed and constructed supply will not require them. However, they do no harm, and may be used if you so desire. Note that if used with a choke input filter, fast diodes must be used - standard speed diodes will overheat. Choke input filters are very uncommon with semiconductor amplifiers, but are seen occasionally with valve amplifiers.

Capacitors may also be used in parallel with the secondary windings, again to reduce noise. Use of caps across the mains is covered below, and great care must be used if you decide to do this. Some common practices are extremely dangerous - especially with 220V or greater mains voltages.

It is very common with valve amplifiers to use a "bleeder" or discharge resistor across the power supply. Although not strictly necessary with low voltage solid state equipment, they don't cause any harm (apart from a small amount of heat). A typical value for supply voltages of +/-30V to +/-60V would be 1k 5W - just make sure they are not mounted too close to the filter capacitors (the heat may reduce the capacitor life).

4.0 Mains Connections

Before discussing the mains, there are several standards of colour coding and nomenclature that need to be covered first. If unsure of any detail, you should seek assistance from a suitably qualified electrical trades person - in some countries it may be illegal to perform any mains wiring unless you are qualified and/or licensed. Make sure that you understand the specific regulations that apply to you - this document is a guideline only, and it is not possible to account for the regulations of all countries.

	Colours			
IEC	**US**	**Old ***	**Lead**	**Also Called**
Brown	Black	Red	Active	Line, Hot
Blue	White	Black	Neutral	Return, Cold, Grounded conductor
Gr/Ye **	Green	Green	Earth	Ground, Safety Earth, Earth Ground, Grounding conductor ***

Table 1 - Wiring Colour Codes

* The "Old" standard was used in various countries (including Australia), and some wiring may still use these colours.

** Gr/Ye - Green with Yellow stripe - this is the standard world wide, although it is not common in the US or Canada at present.

*** There is an important distinction between "Grounding conductor" (safety earth) and "Grounded conductor" (Neutral). These are *US* terms for the conductors and they are ***not*** interchangable, despite the similarity of the names ***!***

The incoming household mains may be connected using a fixed lead, but it is far more convenient to use a connector. The European style IEC connector has world-wide approval, and is recommended. Ready made moulded connector style power leads are available from retail outlets, and are safe and durable. Other lead types may also be available in your area. Be careful that the lead you use is legal in your country - for example, many "specialist" or "high end" leads will be illegal in a great many countries outside the USA - note that this is not one of my attacks on such items, but is a simple fact of electrical safety. Indeed, unless they have UL or CSA approval, you may be at risk in the US and Canada as well, especially if there is an insurance claim in the balance.

If a fixed lead is used, it must be securely clamped at the entry point, and must also be insulated from the chassis with a rubber or plastic grommet. This prevents the lead from damage by metal edges of the entry hole. Cord clamping grommets are available, but be aware that the hole size is critical to the ability of the grommet to clamp the cable securely without damage, and ensure that the lead (replete with grommet) cannot be pulled out.

It is recommended (or required in some areas) that the earth (ground) wire of any fixed mains lead should be longer than the other leads inside the casing to ensure that it is the last to break should the lead clamp fail. This provides some degree of additional safety, but is not infallible.

Use of an approved mains connector is by far the safest and most flexible option.

5.0 Fuse Holder

The mains fuse holder should (must) be a safety type, again depending upon where you live. The safest is an IEC mains connector with integral fuse holder, as it is impossible to access the fuse while the lead is inserted. Where a separate chassis fuse holder is used, it should be connected so that it is not possible to contact the fuse until it is completely clear of internal connections. Most new fuse holders will be designed to meet this requirement, but many older ones will not.

Older style fuse holders allowed physical (finger) contact with a partially withdrawn fuse, which could easily contact an internal conducting part of the holder. The potential for serious injury is quite obvious if power is applied to the unit and the fuse is intact!

6.0 Mains Switch

The ideal mains switch is a double pole switch, to ensure that both active and neutral leads are disconnected when the power is off. This guards against internal components remaining live due to accidental reversal of the mains leads, either because 2 prong mains plugs are used (not recommended), or because of incorrectly wired power outlets or extension leads, for example. These are unfortunately quite common where inexperienced persons have wired the lead, and have not followed the correct colour code.

7.0 Earth Connection

The mains safety earth must be connected to a separate bolt, whose sole purpose is to provide a solid earth connection to the equipment chassis. Where there are separate removable panels, it may also be a requirement where you live that these have a wired connection to the main chassis. This prevents any possibility of the removable panel from becoming "live" should an electrical fault cause the mains to be in contact with the panel - regardless of whether the securing screws are installed or not.

8.0 Mains Capacitors

It is common for many amplifiers to use a capacitor connected between the active (live) lead and neutral. This can provide some useful attenuation of mains borne noise. The capacitor must always be rated for mains AC usage, and preferably for at least 400 VAC - a DC capacitor will fail sooner or later regardless of rated voltage.

In some cases, a capacitor may be used between live and earth (particularly in the US). This is especially common in some guitar amplifiers, and the capacitor is likely to fail - especially at 220 or 240 volts AC, since DC capacitors are commonly used. In some cases, a capacitor may be used from both active and neutral to earth. This is an extremely dangerous practice, and is illegal in many countries. Generally, I do not recommend or condone the use of capacitors from any mains connection to safety or chassis earth. Indeed, under some circumstances these caps can cause residual current devices (RCDs - safety switches) to trip. The use of any capacitor between mains and the chassis places the user at risk of electric shock if the chassis is not connected to safety earth.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 6 June 2003

Copyright © 2001 - Andrew Walmsley

(Edited by Rod Elliott - ESP)

Page Updated 21 Mar 2001

An Amateur’s Guide to Making It Work

Return to Articles Index

Back to Home Page

Contents - Part 1

- Introduction
- It Helps to Have a Purpose
- Avoiding Evaporation Trauma
- Tools, and Their Place in the World
- 'Big' Tools - Nice to have Vs. Have to Have
- The Workshop

Introduction

From the editor (Rod E) - I have inserted some of my own comments, which are identified by ***(rode)*** tacked onto the end. Otherwise the article is almost untouched. Many thanks Andy - A fine piece of work.

A Pursuit Indeed ...

Praise be to electronics. There can be no finer and more honourable pursuit for the man with time on his hands and at least some money to burn.

In theory it is an inexpensive, safe and absorbing hobby with at least the potential for learning what all the pretty little coloured tubes and cans of various shapes and sizes in the back of the television are. Beyond this, you can wow the men and woo the women with your worldly wise talk of linear power supplies, voltage and current amplifiers, pi filters, power transistors and heat sink efficiency. Such talk will eventually guarantee at least one, and possibly more, of the following –

- A seat in the pub when everyone else mysteriously disappears two minutes into the conversation.
- A slap round the side of the head.
- A workshop full of ageing and dilapidated electrical gear, most of which presents a really serious public health hazard, that you've promised to repair free of charge for your family and mates as soon as you find the time.

The gathering of broken electrical gear is a particular one to watch out for. It has been suggested that the recent unexplained disappearance of a number of electronics enthusiasts may have been caused by dimensional instability in their workshops. The theory is that the accumulation of such vast amounts of semi-deceased gear can force into existence a temporal doorway into a world with lead fumes for an atmosphere. This theory has yet to be proved and may be a load of old cobblers. However, it always pays to be wary.

---

1. It Helps to Have a Purpose

Adopting a (slightly) more serious note, the remainder of this article will address some of the issues that you will need to be aware of if you are coming into the field of electronics as an absolute beginner. As with any activity there are some dangers, but the risk of suffering any form of harm can be reduced to practically zero by adopting a few simple working practices and taking a careful and methodical approach to the work in hand.

On the upside, the rewards to be had when you have learned enough to consider yourself a competent amateur are many and varied. Its impossible to make an exhaustive list as everyone gets something different out of their hobby. However, there are some general benefits and these include ...

- A deeper understanding and appreciation of the technology that is a constant in modern life.

- The ability to diagnose and repair simple faults in your own equipment which would otherwise have to be professionally repaired or replaced completely. As an example, an ageing hi-fi amp of mine died recently whilst in use. Five minutes with the multimeter

helped me to confirm that the bridge rectifier in the power supply unit had failed. In this case the repair cost me nothing as I had an equivalent bridge in my parts drawer. Even if I'd had to buy one, it would have cost me no more than two or three pounds sterling. Not bad if you consider that binning the entire amp and replacing it with a new one of a similar quality would have set me back about a hundred times the cost of the new rectifier.

- The satisfaction of correctly constructing kits and designing and building your own circuitry.

- The scope to acquire faulty electrical items at little or no cost and restore the unit to working order in your own time. I'm quite a fan of old audio/video equipment and you'd be amazed if I were to tell you about all the lovely pieces of gear that I have seen destined for the skip when the fault was nothing more serious than a tired transistor or even a blown fuse. A good example of this is my own record deck – a 1974 Bang and Olufsen Beogram 4000 which was going to be binned after three "professional" repairers had failed to bring it back to life. I picked the unit up for ten pounds sterling, spent half an hour replacing the mains supply fuses in the case, and grinned broadly when I plugged it in and watched it come to life. Its been used every day for three years now and has worked reliably on all occasions. What's more, it gives every indication of being quite happy about the prospect of running perfectly for another 26 years, so long as it gets the occasional bit of TLC. As an added bonus, when I want back and talked to the guy in the shop where Id bought it, he was so pleased to hear the record deck working that he gave me amplifier that went with it for nothing.

In the next section, we abandon our traditionally light-hearted approach and discuss the very serious issue of electrical safety. If you read nothing else at all in your life then this read this.

---

2. Avoiding Evaporation Trauma

In the last section, I aimed to convince you of the fun that you can have being an amateur sparky. If you find that you don't fancy it at all then I suggest that you quit now before you start buying loads of relatively expensive gear which will be of no use whatsoever to you.

Electronics is usually a lot of fun, and I try to reflect this in the light-hearted nature of these articles. However, for this part of the series I'm afraid that I have to get very, very, very serious about something.

Yep, you've guessed it – ***electricity***.

In very basic terms, electricity is the flow of electrons along a conductive material such as copper. In order to use electricity in anger, this flow of electrons must be impeded by some sort of resistive or capacitive load. This can be an electrical circuit of varying complexity, a coil such as found in an electric motor, or a filament such as found in a normal lightbulb. I accept that this is verging on a gross over-simplification of the truth, but it will more than adequately serve our purpose at present.

One of the characteristics of this flow of electrons is that it will always follow the path of least resistance to the point of lowest potential. Since the human body is around 98% water, and given that water suitable saturated with mineral salts (that's us!) is an excellent conductor of electricity, the potential for you to involuntarily and unexpectedly become the conductor of a great deal of electricity is considerable. For the amateur electronics enthusiast, this risk is increased a hundredfold since your chosen hobby will inevitably bring you into potentially intimate contact with electricity on a regular basis.

It seems to be an article of faith amongst many that the lower the electrical voltage, the less potential danger exists when working close to it. This is complete nonsense when you consider that the static electricity shock that you can get by touching the body of a car on a hot, dry day can be in the order of 30 – 50,000 volts. Whilst such shocks can be irritating, it is extremely unlikely that you will be inconvenienced beyond this due to the very, very low electrical currents involved, and the instantaneous nature of the discharge from the car body to your own. However, when working with direct current (DC) or alternating current (AC) sources at much lower voltages, the result of physical contact with a live wire can be almost instantaneous death.

Indeed, a current of 50mA (barely enough to make a low wattage lamp even glow) is sufficient to send your heart into a state called "ventricular fibrillation", where the heart muscles are all working out of synchronisation with each other. Little or no blood is pumped, and you will die within about 3 minutes unless help is immediately at hand.

Sometimes (but less often), your heart will simply stop. If this happens, it is possible that with external heart massage that it might re-start, and occasionally it might even re-start by itself - rare, but it can happen.

However, worry not. As I have said, this risk can be reduced considerably with the application of a few simple working practices, a careful and methodical approach to the job, and a large helping of simple common sense. The list below is intended to give some pointers to what the correct working standards should be. What it is not is the de facto standard for electrical safety. It is expected that you will use the recommendations below in conjunction with the absolutely basic principles such as not overloading plug points, not mixing electricity with water, and not leaving live bare wires dangling within reach of anybody.

If you are still in the dark after reading the list then I strongly suggest that you do not even consider proceeding with electronics as a hobby until you have located and attended an approved course on all aspects of electrical safety, and you are more than satisfied that your understanding of the subject is correct and thorough. Many educational institutions such as colleges and universities run such courses during the evening. They are generally quite cheap to enrol on, and usually run for one or two evenings a week for a period of five to six weeks.

- Treat DC and AC electricity at **ALL** voltages with the utmost respect and caution. This includes all household and battery supplies, no matter how small the battery may be. Statistics suggest that the survival rate for people with gunshot wounds is far higher than for those who have suffered serious electrocution.

- Unless you are **ABSOLUTELY** sure of what you are doing, **NEVER** work on a live chassis under any circumstances. Even experienced engineers are loathe to work on live equipment unless it is absolutely necessary. Prior to working on any form of equipment, ensure that it is isolated from the mains by physically disconnecting the plug from the mains socket. If the mains socket is switched then also ensure that you have switched the socket off. If you must work in close proximity to any form of electrical socket then stick insulating tape across the front of the socket to prevent electrocution due to tools or fingers coming into inadvertent contact. In addition, take careful note of what I say in all sections here before you even *think* about breaking out the screwdrivers.

- **Never** under any circumstances be tempted to "jerry rig" your latest creation to the mains using bare wires and matchsticks (or similar) shoved into a plug socket. You may laugh but I've seen such suicidal stupidity perpetrated on many occasions. **This is an incredibly stupid thing to do !** When connecting any device to the mains, use a good quality plug from a reputable manufacturer, and ensure that the plug is correctly assembled and that all connections and covers are secure and tight prior to connection. For plugs with integral fuses, such as those supplied in the U.K., ensure that the fuse is of the correct rating for the device.

- Of course, there are rare situations where it is necessary to work on a live chassis – setting the bias current on a power amplifier is one example. Prior to commencing work of this type, it is **IMPERATIVE** to satisfy yourself that **ALL** electrical connections are correctly insulated in order to prevent accidental contact. This is particularly important in equipment that you have built yourself, though I have seen wiring in quite expensive commercial products which is nothing short of reprehensible. When making insulated connections I have three sizes of heatshrink sleeving. After ensuring that the connection is well made and has no sharp edges which may puncture the sleeving, I alternately apply and shrink each size of sleeving, finishing with the largest. In this way,

I am satisfied that there can be no accidental contact.

- In addition to the other points raised above, **NEVER** wear a watch, ring, necklace or any other form of conductive jewellery whilst working on electrical devices, whether they are live or not.

- If, like me, you have switched mains sockets mounted on your workbench, **NEVER** assume that the appliance that you are working on is completely isolated from the mains when you switch the socket off. To be safe, always completely remove the mains plug from the socket and lay it well away from any live contacts in a place on the bench where it cannot be knocked. In addition, give serious consideration to having a qualified electrician install a residual current device (RCD*) on the mains supply to your work room. These devices work on the principle that if more current is flowing on one supply pole than the other, then some must be going where it is not intended to. Under these circumstances, the device will cut out the supply of electricity extremely quickly. I have witnessed a situation where someone came into accidental contact with a mains cable immediately after inadvertently slicing it in two with a pair of hedge trimmers. Luckily, the cable was connected to the distribution board via an RCD which cut out immediately, and the unlucky gardener suffered no ill effects whatsoever. Had the circuit been protected with a more traditional cartridge fuse, the situation may have been far different. Even if a fuse does blow in a situation where someone is being electrocuted (and this is not at all guaranteed), this can take up to three seconds, which is more than long enough for the unfortunate person to be killed outright.

- Be aware of the fact that any equipment which utilises power reservoirs such as capacitors in its design may well be live for many hours or days after it has last been powered up. A excellent example here is the power supply for any form of amplifier. Nearly all power supplies for these devices utilise capacitance smoothing after the rectifier, and in some cases the capacitors used can store a great deal of power for a considerable length of time after the device has been switched off and disconnected from the mains supply. When working with equipment of this type, if you are in any doubt at all then it is far safer to assume that there is power stored in the capacitors, and proceed accordingly.

- In addition to the points raised above, be aware of the fact that **ALL** Cathode Ray Tubes in televisions, computer monitors and oscilloscopes can store lethal levels of charge at many, many thousands of volts for weeks after they have been switched off and disconnected from the mains. The design, construction, repair and servicing of any device incorporating CRTs is a very specialised and dangerous task unless you know exactly what you are doing. Indeed, I know of many extremely competent electronics

enthusiasts who will not even consider touching devices incorporating these components. To be safe, I suggest that you don't either.

- Microwave ovens have probably killed more technicians than any other electronic device. The capacitor can store a huge electrical charge, and this charge can remain poised to pounce on any unsuspecting technician for weeks. These animals are seriously dangerous, and must be treated with the utmost respect - or avoided altogether. ***(rode)***

* RCDs are also known as earth (ground) leakage circuit breakers ***(rode)***

This may seem to be an extensive list, but I am loathe to make an apology even if you feel that I may have overstressed the point. Whilst I do find that the majority of people behave sensibly and apply common sense when working with electricity, I have seen some inexcusable examples of sheer stupidity in my time, and these have generally been as a result of people not thinking about what they are doing. Mistakes with electricity are at the best costly, and at the worst fatal.

In the next section, we make a thankful return to slightly more light-hearted matters and discuss how to assemble a decent toolkit without having to remortgage the kids.

---

3. Tools, and Their Place in the World

Its always said that a bad workman always blames his tools. A corollary of this is a good workman can only be as good as the tools he is using. In this section, we'll look at the toolkit you will need to assemble in order to get started as an electronics hobbyist.

Firstly, a word on the buying of tools in general.

I'm sure everyone who is reading this has strolled around the odd Sunday market in the course of their lives and encountered the mythical 50,000-piece toolkit of dubious geographical origin that cost little more than a decent set of screwdrivers, and which seems to meet all your needs in one fell swoop without breaking the bank. Whilst these kits may seem to be excellent value on the face of it, and the tools that they contain may look to be little different to their far more expensive counterparts stocked by well-established retailers, they are generally not up to the reasonably hard daily use to which you will put them in the pursuit of your hobby.

In the case of tools such as hammers, spanners, screwdrivers and drill bits, the very cheap ones are not only a waste of money as they won't last two minutes, they are downright dangerous. Even when undertaking very light duties in the workshop, you'd be surprised at the

stresses and strains on a tool as simple as the humble spanner or screwdriver. In order to perform correctly under these conditions, good quality tools are well designed for the job in hand, and the metal from which they are made is correctly tempered and heat treated so that the tool will give many years of trouble free service before a replacement is required. Cheaper tools are generally poorer quality copies of the better designs, and there is no guarantee that the metal has been correctly treated at all. At the very least, such a tool may slip whilst in use and damage the workpiece. At the worst the tool may bend, snap or even shatter whilst in use, causing personal injury.

As a general rule of thumb, when buying tools you should go to a well established retailer and only buy tools made by reputable manufacturers. It is far better to spend some time saving up to buy the best tools you can afford, rather than to compromise on cheaper ones which will not last as long, and which may not be as satisfying to use. This is particularly the case with tools that will be put to heavy use such as hammers, spanners, pliers, screwdrivers, cutters and strippers.

When it comes to power tools, what I have said about buying quality items is especially pertinent. It is important that tools such as power drills, jigsaws, circular saws, routers and planers are of good quality, and this extends not only to the tool itself, but to any accessories or blades which are fitted to it. Be especially wary of ultra-cheap power tools that seem to offer the world for little money. They will not last long, they most certainly won't be up to the job in hand, and you may end up completely spoiling what you are working on for the want of something decent.

Before we start with our wish list, it's worthwhile mentioning that there are very, very few injuries suffered when a well designed tool of good quality is used correctly for its intended purpose. The well known banged heads, scuffed elbows, skinned knuckles and puncture wounds that give us the walking wounded of the DIY wars are only caused when the right tool is used for the wrong job or a tool is forced beyond its designed limits. Using the correct tool generally means that you'll get the job finished quicker, you won't injure yourself, and you won't break or spoil anything else whilst working.

- **Screwdrivers -** The better quality screwdrivers have tempered shafts and hardened tips so that they won't slip on the screw head and damage the screw, the workpiece, or most importantly, you. Try to aim to have at least four sizes of flat head screwdriver, and four sizes of posidrive screwdriver, along with a decent set of flat headed and posidrive jeweller's screwdrivers for dismantling or assembling smaller components. For ease of use, those with the softer contoured handles are far better than those with the traditionally shaped plastic or wooden handles, and they generally provide far better insulation – this is a must when working with electrical gear. You'll be surprised at just how much use your screwdrivers will be put to, and how much it will hurt if one slips, so buy the best that you can possibly afford.

In use, it is important to remember that a screwdriver is only for the fixing and removal of screws. It is not a chisel, a hammer, a crowbar, a counterweight, a hook, or any of the other things that the many improvised uses I have seen may suggest. Years of experience has taught me that screwdrivers are sentient beings in their own right. If you misuse them they WILL bite back. Some day I'll get round to publishing my extensive collection of scars to reinforce this point – there's no teacher like pain :-)

- **Spanners -** You won't need anything in the monkey wrench league for electronics work. A good quality set of open ended and ring spanners which go from 3mm to 13mm in 1mm increments will be ideal for your purpose. If you're going to be restoring old equipment then it may be an idea to get hold of similar-sized set of imperial spanners as well, though these needn't be anywhere near the top of your list otherwise. As usual, make sure that what you're buying is the best you can afford. The drop forged chrome vanadium spanners from well-known manufacturers such as Draper (in the U.K.) or Snap On (worldwide) are made to exacting standards out of high quality materials. Whilst the initial outlay from buying a set may have you sitting in a cool room for a while to recover, they really will provide you with a lifetime of unfailing service. In the U.K., cheap and nasty spanners are known as ‘knuckle f*****s’ - nobody wants hands like an ageing prize fighter when you're trying to fiddle that 2mm nut on in a space it's taken you the best part of two months to get into.

- **Pliers and Cutters -** Again, there's no need for a pair of water pump pliers with 40 inch knurled vice grips. Try to aim for a good quality set of pliers which consists of two or three sizes of needle nosed pliers, a couple of sizes of flat nosed pliers, and a couple of different configurations of wire cutter. As usual, look for good quality well made examples with comfortable moulded grips that provide high electrical resistance – this is very important. Cheap and nasty pliers are particularly prone to shattering when being used with enthusiasm. When they do give way, bits of razor sharp metal fly everywhere at incredible speeds, so you're not even safe when standing behind someone who's using a pair of bargain basement specials.

- **Knives -** Frowned down upon by some enthusiasts, but invaluable in my view. I have a set of scalpels for fine work, a pair of Swiss Army knives for general use, and a couple of large lock knives for stripping the insulation from heavy duty cables that won't fit in my wire strippers. Do not even think about using something other than a good quality lock knife when you're going to be applying some pressure to the workpiece in order to cut through it – pen knife blades can simply close against your fingers whilst in use. No further description necessary I fancy.

The ubiquitous Stanley™ knife is a good investment - avoid the many cheap copies, as they often have a less than perfect locking mechanism. Searching the workshop floor for the missing bit (prior to rushing off for microsurgery) is not the ideal way to spend one's Sunday afternoon. ***(rode)***

- **Hammers -** There's nothing to be ashamed of – even the most refined of us occasionally need to resort to the rough stuff in order to finish the job. To be fair, it's rare that you'll ever need a hammer when working with electronics, the two just don't go together. However, having a small toffee hammer tucked away in the corner of your toolkit can be invaluable, for example when working on racking or equipment mounts. To be honest, it's so long since I bought a hammer that I have no idea how to go about selecting the correct one. They used to be sold by the weight of the head and I have two in my toolkit stamped at six ounces and one pound respectively (roughly 170 and 220 grams). If possible, try and get hold of a ball peen hammer with a head weight of about 150 grams and a good quality wooden handle. For the use that I suspect it will get, going broke buying it is not essential, but remember to steer away from the really cheap and nasty stuff.

  On the subject of hitting things, a set of centre punches is essential if you intend drilling holes (which is hard to avoid most of the time). A centre punch allows you to carefully mark a small indentation on the workpiece before you start drilling, and it will stop the drill bit from wandering all over the place until it finally decides to make a hole in the wrong place. "Automatic" centre punches are also available - these do the same thing, but you only need to press down on the tool - a spring provides the tension, and a pressure sensitive release "lets go" at the preset pressure and makes a nice little indent for you. ***(rode)***

- **Reading Lamp -** What ? Nothing for nipping, gripping, bending or knocking ? Nope, just a plain old reading lamp. Nowadays reading lamps come in all shapes and sizes ranging from the standard old anglepoise to space-aged low voltage halogen designs with the transformers hidden inside an art-deco base. I would avoid the really expensive models as I suspect you're paying a lot of money for design and styling which will never be appreciated in the workshop. A simple flexible necked reading lamp with a nice heavy base and a 40-60 watt bulb fitted to it will be more than adequate for your needs.

  While you are at it (especially for fine work, and more so if you are getting old, like me :-) a good magnifying lamp or headband magnifier is worth its weight in ruined circuit boards. ***(rode)***

- **Drill and Drill Bits -** Its a good idea to have a large mains-powered drill for heavier work such as drilling metal casework, and a smaller battery powered drill for lighter duty tasks

such as working with plastic, or drilling very small holes using bits which would be prone to snapping if used in a power drill. In the case of both the battery drill and the mains powered unit, if you can buy those that have electro-mechanical devices which sense the pressure on the drill trigger and vary the motor speed accordingly then all the better. This is an especially useful feature on the battery-powered drill and means that you can work with very delicate items at extremely low speed to minimise the risk of damage to the workpiece. As usual, buy the best unit that you can afford from a reputable manufacturer.

This is also true for the drill bits. It is often more economical to buy these in a set and you should aim for good quality items made from high speed steel. If you can buy a set of drill bits which also includes one or two hand or power reamer bits then this is ideal. Whilst all drill bits will eventually go blunt and need replacing, the better quality ones do give much longer service than the cheaper ones. When drilling any material, considerable friction is generated between the drill bit and the workpiece. It is important that you don't allow the bit to become too hot, as this can sometimes cause it to soften and become ineffective.

During drilling, do not apply too much pressure to the drill as you will almost certainly snap the bit. Over time, you will develop a ‘feel’ for when the bit has nearly penetrated the bottom of the hole, and you will know when to ease pressure on the drill in order to avoid the jaws of the chuck coming into contact with the workpiece and damaging the surface.

To extend the usefulness of your mains power drill, you will also want to invest in a good quality drill stand and press. Its probably a good idea to buy one made by the manufacturer of your drill as you can then be sure that everything will fit together correctly. If the drill press comes with an optional vice that attaches to the bed of unit and allows workpieces to be held rigidly whilst drilling then all the better.

- **Taps and Dies -** Nope, not the ones on your bath, mate. Once a suitably sized hole has been drilled in a piece of metal, a tap is used to cut threads into the edges of the hole to admit and hold a threaded fixing. A die is the exact opposite of a tap in that it is screwed onto a shaft of metal (usually referred to as a ‘blank’) to cut threads and make a bolt. There is a dizzying number of standards in the world for specifying and cutting threads of different pitches in different sized holes or blanks, and whole volumes have been written on the subject. I don't propose to go into these at all here (though the insomniac may find an excellent cure for his affliction within their pages). In response to this multitude of standards, many tool suppliers sell a kit of taps and dies covering the thread sizes and pitches that will be most commonly encountered when working with mainstream equipment.

Taps come in a number of types but we're only really interested in tapered taps which are normally used for cutting threads in panels, and so-called ‘blind’ taps which are used for cutting threads in holes which do not pass through the whole depth of the material. A good example of the latter would be holes bored in one side of a heatsink for transistor mounting bolts. When buying your kit, ensure that it contains all the necessary sizes of tapered and blind taps. Using taps and dies is nowhere near as easy as it first looks.

A detailed discussion is really beyond the scope of this document, and it would be far better to find someone who is skilled in their use and ask them to show you. Once you are satisfied that you have mastered the basics, practise on some scrap metal until you are sure of what you are doing. During the course of this learning process, don't be at all discouraged if you turn out some dreadful examples. As I have hinted, there is a definite knack to using these tools and time is the best teacher.

- **Wire Strippers -** These will separate the professional amateurs from the amateur professionals every time. Unless you're working with thick armoured cable, or very heavy duty solid core cable, you're showing yourself up if you try and remove the insulation from the wire with your thumb and the blade of a knife. Not only is it guaranteed that you'll eventually slice a chunk of yourself off, but it requires an incredibly light touch and almost a lifetime of practise to feel when the blade has passed through the insulation and is about to damage the conductors. In multi-cored cable such as mains cable, you can cut through the outer insulator and damage the inner cable sheaths in one smooth movement. Assuming that you even notice this potentially dangerous mistake, the only thing to do is cut the whole lot off and start again. I'm sure that if all the cable that this splapdash approach wastes each year were to be put end to end then we'd be well on the way to the moon in no time.

  If you're doing it right, you're going to be using your wire strippers a hell of a lot in electronics. If you don't splash out on any other tool in your kit, you'll be thanking yourself for years if you show a little extravagance when buying this one. The very best wire strippers are nicely balanced with padded handles, well machined spring loaded mechanisms and replaceable blades. The standard for wire sizes is called (not surprisingly) the Standard Wire Gauge (SWG). All good wire strippers are marked in some way to indicate how the tool should be set in order to strip the insulation from wire of a certain SWG. If the tool is set to an SWG which is too high for the wire you are working with, it will just damage and stretch the insulation without removing it. If set too low, then both the insulation and some or all of the conductors (assuming stranded wire) will be removed. If you are working with solid core cable, the conductor may seem to be undamaged until you try to solder it to the fitting, at which point it will snap. In all cases,

the damaged section of wire must be cut from the length, the tool correctly set, and the operation repeated. If you have a serious amount of disposable income to blow on a pair of wire strippers (and I suggest that if you have then you should), you can buy some with an "intelligent" spring loaded mechanism that "senses" when the blades have cut through the insulation, and stops them before they reach the conductors. The blades then move backwards in the jaw to break the newly cut insulation away, all in one smooth movement. Its not necessary to set the SWG for the wire that you're using, and they'll strip anything first time, every time. Years ago, I went without beer for a whole month to buy a pair of those – it was (damned) difficult at the time but I appreciate my strength of will every time I use them.

- **Soldering Iron and Desolder Tool -** You'll be using your soldering iron almost as much as your wire strippers so it's worthwhile getting a good one. Go to a shop where the assistant will let you pick up a number of similar irons and chose the one that feels comfortable and well-balanced. The heating element should be ceramic cored and the iron should be designed in such a way as you can change the tips as and when necessary. If you can select a model with low tip leakage current then all the better. As well as the iron itself, you'll need a good heavy stand, ideally with an integral sponge. Its also a good idea to buy a number of different tips of varying sizes for different types of work. The more advanced (and expensive) temperature controlled soldering irons and soldering stations are overkill for the beginner, though you may want to think about one if you're doing a lot of work, or if you become very serious about your hobby.

  Desolder tools come in a dizzying array of shapes and sizes. For the hobbyist, the spring-loaded plunger design is more than adequate. Be aware that these tools do need to be dismantled and cleaned, practically after each session, and the tips and ‘O’ rings do need to be replaced periodically. If you can buy a supply of ‘O’ rings and tips along with the desolder tool itself then all the better.

  With regard to the solder itself, buy a decent quality 60/40 solder, and **NEVER EVER** use plumber's flux when working with electronics. Its not necessary, makes a hell of a mess, and is highly corrosive.

- **Files -** Buy a good quality set of needle files for close work, and a set of larger files for general metal working. The now familiar caveat of buying good quality is still relevant even for files. Cheap and nasty files have uncomfortable handles which will do nothing but give you blisters and snap when you're using them. The teeth are poorly machined and they'll just make a mess out of all your hard work. A piece of metal worked properly with a good quality file can sometimes not be distinguished from one which has been machined. A couple of flat (preferably "bastard" cut), round and square files in different sizes will allow you to make odd shaped holes for connectors and switches.

- **Hacksaw(s) -** A good quality full size hacksaw and a "baby" version - but not the plastic bodied types - they are useless. Cutting pot and switch shafts, reducing screws to a sensible length and a multitude of other essential cutting tasks will be a lot harder if you have to chew through them with your teeth. ***(rode)***

- **Steel Rule -** Absolutely invaluable this. Buy a good one with both metric and imperial divisions. A couple of cheap plastic rules will also come in handy. For marking out panels before drilling, an accurate 90 degree square is a good investment.

- **Vice -** Loose women ? Gambling ? Drinking to excess ? If you're already good at any of these then you don't need me to tell you how to do them better, and if you've already got such a wild life then when will you find time for electronics ? :-)

  What I do mean is the humble old bench vice. You don't need a huge engineer's vice for normal electronics work, a small woodworker's vice will do just nicely. If you can get one that comes equipped with a quick release mechanism so that you can move it to whichever corner of the workbench takes your fancy then this is excellent. One which will allow you to fashion and fit jaw covers made of wood or plastic is also a good idea, as the standard cast or wrought iron jaws will almost certainly damage a lot of the components that you will be working with, no matter how careful you are.

  On the subject of holding things in place, I also find that a collection of artery forceps and a number of different sizes of ‘G’ cramp are often just what you're looking when there just aren't enough fingers or hands to go round. (Also see miscellaneous, below.)

  Finally, for just a few of your hard-earned dollars (pounds, shekels ….) you can get hold a tool with a heavy base, infinitely flexible arms about six to ten inches in length ending in a pair of crocodile clips, and with an integral magnifying glass attached. What's more, they don't complain and drop the wire when it starts getting hot. I don't know what they're called elsewhere, but they're commonly known as ‘helping hands’ in the U.K., and they're the best thing since sliced bread when it comes to connecting cables to small plugs or sockets. You can arrange the cable and the socket just how you want and then apply the soldering iron without fear of anything flying off into the distant corners of the workbench. For example, if you chose to do a lot of work with older hi-fi equipment as I do, you'll often encounter the old five and seven pin DIN plugs and sockets. Don't even try to solder one of these without a pair of helping hands.

- **Pop Rivet Gun -** Whilst one of these isn't vital for a beginner, they are useful. When actually installing the rivet in the workpiece, you're asking a lot of the gun so buy a decent one, and don't forget to maintain your stock of rivets and washers. The times that I've gone to the rivet drawer to discover that I've used up the size that I need are without number, and this always seems to happen on a Sunday afternoon when everywhere is shut.

  I can't resist a further word on the Sunday afternoon phenomenon at this point. Whilst I have no proof of this (what would you expect ?), I've a feeling that the forces of physics and the natural vindictiveness of small inanimate objects are in a particular state of harmony after about 2.00 PM on a Sunday afternoon. If it's a wet and cold Sunday afternoon then this harmony is a particularly good one. If it's snowing and you have a desperate need for the thing that you are working on the next day then the harmony is practically angelic. Expensive and difficult to obtain components that you could attach a spanner to and happily swing from (on any other day of the week) will readily snap, bend, run away after being dropped and magically metamorphose into something the wrong size after 2.00 PM on a Sunday afternoon, and there's not a bloody thing that you can do about it as everywhere is shut. To add insult to injury, you just **know** that you haven't got a spare – you don't have to even both looking, you just **know**. I've talked to a lot of hobbyists and mechanics who have built an entire belief system of uncanny internal consistency around this phenomenon, to the point where they will get very drunk on a Saturday night for the **sole** purpose of averting the temptation to do **anything** on a Sunday afternoon – it's nothing to do with riotous quaffing or having a good time, it's all about avoiding the Sunday afternoon sinking feeling - honest :-)

  The above is slightly different in other countries where a somewhat more liberal approach to regulated shopping hours is common - the effect will simply perform an appropriate time shift to ensure maximum annoyance and/or frustration. ***(rode)***

- **Miscellaneous *(rode)* -** Gripping "things" - spring clamps, clothes pegs, G-cramps and a tiny (as in *really* small) bench vice will be used once in a blue moon, or every day. It is almost guaranteed that if you don't have a good selection of these essentials, their immediate requirement to allow you to complete the job will increase tenfold. Also include masking tape and rubber bands (the former holds anything in place temporarily, and the latter are great for keeping a pair of pliers closed on the workpiece :-)

  Consider raiding your local medical supply outlet for a pair (or two) of artery forceps. These used to be considered “unusual”, but they are now available from many electronics retailers who have finally realised the usefulness of these little devices. They may be used as tiny pliers, clamped onto sensitive component leads to act as a heatsink, or used to retrieve that (blessed) small screw from the nether regions of a

chassis where it has taken root. These are probably my most used small “gripping tool” of all.

A large plastic kitchen cutting board can also be very useful for cutting paper labels (to be stuck to front panels and covered with clear plastic), plus other uses which will dawn on you when you see all the knife marks on your workbench. Even slicing open plastic packets of resistors or capacitors is made easier and safer if you have a proper surface to do it on. Oh yes, add a couple of pairs of scissors to the list while you are at it.

An array of pencils, a steel scriber, and a few permanent and non-permanent felt tip pens will always be used, as well as a notepad. I know I'm stating the obvious, but you'd be surprised how often these basic essentials are nowhere to be seen.

Andy has made a good point with all his suggestions about good quality tools, but sometimes the 50,000 piece set for $29.99 actually makes good sense. Some of the tools will break, but in the meantime, you have a huge array of things you can use until you find out what you really need - you then replace the broken cheapie with a good quality equivalent, knowing in advance that you will use it (you must, otherwise you wouldn't have broken it, right?). Don't trust any of the cheapie sets for anything other than "light duties" - and sometimes some of the other bits of the kit can be modified into something you really need, even for a single use job - a blowtorch, angle grinder and hammer can modify almost any tool, and if you mess it up, you haven't blown any serious money. This is a technique I have used many times, and as a result have a very wide range of "interesting" tools (the original purpose of which is now lost in time in many cases :-)

Earthed wrist strap and earthed mat for working on static sensitive devices (e.g. MOSFETs and CMOS logic).

Insulating safety gloves for working on valve amps and mains supplies. - Not to everyone's taste, but useful if you are sufficiently paranoid (not a bad trait when electricity is involved).

Wirewool and fine emery paper for cleaning component leads etc.

A first aid kit, especially if the workshop is away from the house.

A self-contained smoke detector in case accidents with the soldering iron etc.

A suitable brick wall (head-hitting for the use of) for when things don't go quite right :-)

(With some additional input from Geoff Moss, my unpaid, but very appreciated editor)

---

4. ‘Big’ Tools - Nice to Have Vs. Have to Have

We've now exhausted the list of smaller tools that you'll need to make a start in electronics. As you progress, you'll undoubtedly gather unto yourself a wide and varied selection of useful tools above and beyond what I've mentioned in this section, and you'll be amazed at just what you can buy. For example, I recently saw a tool that you can fit to the end of your power drill that will make a four inch deep perfectly square cutout of any size in any material short of solid steel ! No more belting the hell out of a cold chisel just to fit a socket in the wall I fancy !

Before we move onto setting up a workshop, it's worthwhile briefly mentioning a number of tools which may not have an immediate application to electronics, but which are often worth their weight in gold when carrying out related work. Please remember to always wear the correct protective equipment when using power tools. At the very least, always have a pair of safety specs handy as a matter of course.

- **Drill Press -** A small cheap drill press will be more accurate than an expensive hand held drill, no matter how good you are with the latter. For drilling heatsinks and even PCBs, they have no equal. Most are now made in China and are a tad dubious, but if set up with care will be more than adequate for most jobs where small or perfectly vertical holes are essential. This is especially true if you are drilling holes that are to be tapped. ***(rode)***

- **High-Speed Rotary Engraver -** These are made by various manufacturers, and are invaluable for making small modifications to cases, PCBs, and just about anything else that is not the shape you want it to be. For a small tool, they are expensive, but add a few sanding drums and disks to the basic set, and you have a tool that has a multiplicity of uses for fine work. ***(rode)***

- **Angle Grinder -** Go anywhere, do anything, sharpen anything and cut anything. Watch out for sparks and dust and ALWAYS wear good safety specs. I've had the odd grinding disc explode on me and it's NOT funny !

- **Circular Saw -** Adds infinite macho appeal but wear safety specs and please keep your fingers out of the way

- **Electric Plane -** I can't imagine anything better when wood refuses to fit !

- **Jig Saw -** Just remember that they make different saw bits for wood and metal. When cutting aluminium, always ensure that the workpiece is well lubricated with light machine

oil

- **Heat Gun -** Great for mess-free paint stripping and there's nothing better for heat shrink sleeving. Do try to remember that these are not to be used as hair dryers :-)

- **Power Sander -** For those times when you've got better things to do than spend forever rubbing things down by hand. Orbital sanders are excellent for fine finishing. Belt sanders are particularly effective (and evil).

- **Band Saw -** Excellent for case work, and a far better alternative to a jig saw when making long cuts in sheet metal

- **Router** - Can be used to cut perfect (including recessed) holes for speaker mounting, and odd shaped cutouts for terminal panels and the like. Rounding bits allow you to make all the sharp edges go away on a speaker box, which is great for reducing refraction and protecting small people from injury when they crash headlong into your cabinets. Other router bits can make fancy trims or just simple slots for shelving or reinforced cabinet assembly. ***(rode)***

- **Sheet Metal Punches -** Whilst not cheap, a set of these useful little things will guarantee that you'll never bend or break anything again whilst trying to force a 12mm HSS drill bit through it at suicidal speed.

- **Welder -** Of all electric arc welders, nothing beats a MIG (Metal Inert Gas) set to my way of thinking. They're versatile, easy to use, cheap to run, and you can get some excellent results with one after a bit of practise. If you're really going for it then think about acquiring a small oxygen and acetylene welding set for metal cutting, brazing and really fine welding work. When working with electric arc welders, ALWAYS use an approved welding mask (there are NO exceptions to this), and remember that a recently welded workpiece is guaranteed to make you sizzle if you touch it, even when the bright red glow has gone away!

  With oxyacetylene equipment, bear in mind that both gasses are HIGHLY explosive if incorrectly (or even correctly) fiddled with. You won't just take out yourself and your workshop, you'll very likely also take out the entire street (I'm not kidding here).

- **Lathe -** Using one of these properly takes a lot of mastering, but once you've got there,

you can make practically anything you need. Many hobbyists and home mechanics believe that the lathe has almost totemic powers against the Sunday afternoon syndrome referred to above.

- **Air Compressor -** A tool of little use by itself, unless you just want an unusual source of noise and heat. With the right attachments, you can spray paint, blow swarf and dust out of the hardest to reach places, and operate a whole array of air tools. Not specifically something for the average home hobbyist, but one of these really has some posing value.

---

4. The Workshop

Okay, you're skint as a flint (i.e. completely broke!) but replete with tools, and you're itching to get going. All that you need now is the workshop.

In this section, I'll provide some guidelines on how to establish a workshop, how to build a workbench, and how to store components and tools so that what you need is always to hand without too much searching and turning over of rocks. I'll also make a brief mention of how to maintain your tools once they are installed and racked up in the workshop, and how to approach your work correctly once the workshop is in place and being used for its intended purpose.

- **The Right Place at the Right Time -** The first challenge is to select the place in your house, garage, outbuilding, shed, shack or cave that will become your workshop. I know that your choice will be limited by available buildings, space not dedicated to other uses, and the overall size of your house and outbuildings, so consider the following solely as guidelines.

    - Try to establish your workshop in a space no smaller than 20 feet by 40 feet (around 6 metres by 12 metres give or take).

      Naturally, your personal situation (or s/he who must be obeyed) will dictate the feasibility of this - subtle force may be required in extreme cases (i.e. every time). ***(rode)***

    - Ideally, the space should be one that can be permanently dedicated as a workshop. There's nothing worse than having to clear all your tools and gear away in the middle of a half finished job. In addition, the best way to avoid becoming cheesed off with the job in hand is to know that you can walk away, close the door on it, and return whenever you wish. Always remember that this

will be your hobby, and that it should always be fun and a form of relaxation and recreation, not a chore. In some cases I've finished projects in a weekend when I've been in the right mood. In other cases it's taken me six months to a year to complete a piece of work, and I still have projects ongoing now that I started years ago. Having a dedicated workshop allows me to do pursue my hobby in this way.

- The ideal floor covering for this space will be smooth concrete with a coating of floor paint, or floorboards covered with hardboard and linoleum, depending on the type of building. Carpet will inevitably get stained, etched, burnt and dirty and dropped components have a knack of disappearing into it forever.

- Ensure that the space has an adequate source of ambient lighting, ideally both natural and artificial, which you can then supplement with your desk lamp for close work.

- Ensure that there is a safe and reliable mains supply to the workshop. I strongly suggest that where practical you take a fused spur directly from the distribution board into the workshop board, and feed each workshop socket from a separate RCD. There are different types of RCD, some will tolerate a little more leakage or surge current and take a little longer to trip, others are extremely sensitive and will trip at the first sign of any form of current imbalance on the active and neutral poles. Given the use to which the workshop will be put, always go for the second type of RCD. Okay, this may seem a little over-cautious, and will almost certainly not be cheap, but we're talking about a room in which electricity will be played with on an almost continuous basis.
  In comparison with the cost of a human life, a bit of mains wiring and couple of RCDs is an infinitesimally small price to pay. My own workshop is wired in exactly this way. I have of course had the odd mishap over the years, and the majority were due to unexpected component failure, and one or two have been due to working too late into the night and getting tired, or just being bloody stupid. In all cases my RCDs have protected me from any harm by tripping instantly, and I've lived to tell the tale.

  Where you must have a temporary work space, portable "in-line" RCDs are available, and one of these should be used as a matter of course. Also make sure that you use the test facility that is provided on these units regularly, to make sure that it is working properly - a false sense of security can get you killed. Do try not to become a statistic! ***(rode)***

  On a final note, as is the case with all electrical installations, if you're not **absolutely certain** of what you're doing then for God's sake get the work carried

out by a competent and correctly qualified electrician. You can have all the safety devices in the world, but if the wiring itself is a health hazard then you may as well replace your RCDs and fuses with cut off lengths of six inch nails for the good that they'll do you. Even if like me you're a relatively competent amateur, have a correctly qualified electrician at least look over all your wiring before you throw the switch for the first time. If nothing else, at least you've then got peace of mind.

- If the workshop will be in an outbuilding such as a shed or detached garage, bear in mind that there will eventually be a fair amount of expensive equipment stored in it, and secure the premises accordingly. Your local police crime prevention unit can generally advise you about how to go about securing your workshop, but common sense and a little low-grade paranoia are just as good. It goes without saying that if there is a window into the workshop then it should be covered by a net curtain, and valuable items should not be left on display.

- If you live in the U.K., or a country with a similar climate (hard to imagine but there you go), you will need to think about keeping the workshop dry and warm. If the workshop is in your house then this isn't much of a problem. If you are working in an outbuilding then ensure that it is free of leaks and draughts, and that condensation is kept to a minimum all year round. Electricity and damp conditions are not particularly good bedfellows, and keeping all your tools and components in a damp and cold environment will do nothing at all to prolong their life. As a secondary concern to the well-being of your tools, it's a good idea if you can also be warm and comfortable whilst working, and suitable heating should be arranged to ensure this.

- **The Workbench -** The place where it will all happen. I've spent many, many happy hours in front of my workbench and I attribute a lot of these to the thought that I put in when building it. A good way to start is to take your chosen workshop chair, sit down in it and decide the height at which you will be comfortable working. Ensure that your chosen height will allow you to keep your back and neck relatively straight, and to work with your forearms at an angle of about ninety degrees relative to your upper arms. Assuming the rough room dimensions referred to in section 1, and the presence of a window in your chosen space, try to run the workbench down the full length of one of the long walls as close to the window as possible, and make it between a meter and a half and two meters in width to give you plenty of space for your tools, your test equipment and the project currently in hand. Where practical, support the workbench on sturdy brackets anchored to a load-bearing wall with rawlplugs, or to the wall frame rather than the cladding if the workshop is to be in a timber framed building.

  Supplement the brackets with a number of legs along the front of the bench which should be securely anchored to the floor. The top of the bench can be covered with thin

aluminium sheet or heat-resistant Formica, or you can do the whole job in one fell swoop by getting hold of a suitably sized lump of kitchen worktop. This is an excellent option if you don't want to mess around with Formica and evo-stick, or if aluminium sheeting wouldn't exactly blend in with the room's existing décor. Tools which will be permanently attached to the bench, such as your band saw or drill press should be fitted within easy reach, but away from the section that is to be your permanent working area. Its a good idea to attach about eight mains sockets with double pole switches and integral neon tell-tales to the top of the bench, and to feed these from the RCD board that you can attach to the wall beside your bench.

Since you are now completely broke (as we established earlier :-), good workbench tops are expensive! An alternative that I have used (am still using, actually) is the "hollow core" interior door. These are usually inexpensive, and are surprisingly strong because of the cardboard matrix inside. Minimal bracing is needed (or none at all), and they will happily support 1kW power amplifiers without a complaint. Surface treatment is optional, since most are pre-painted with an undercoat and they look quite good as well as being the cheapest worktop you will ever get. Just don't drop heavy things on them, or the hardboard surface *will* break. Not normally a problem - mine has moved premises several times and has been in constant use for well over 15 years with no protective covering at all ! They are not suitable for heavy bench mounted tools such as drill presses, grinders or a vice, so don't even think about it. ***(rode)***

Storing tools is a fairly simple job, and you can make an excellent rack with a decent sized sheet of MDF and a pot of assorted self-tapping screws. There's the added bonus here that you can lay everything out just as you want it without having to compromise as you would with a shop-bought tool rack. Once you've decided what will go where, use a pencil to trace the outline of the tool onto the MDF backboard. It may all seem obvious when the tools are in the rack, but when they're all out on the bench and it comes time to put them away, all you'll see is an irregular matrix of random-sized self-tapping screws. If you think this is the voice of experience speaking here then you'd be absolutely right :-)

- **Storing Components -** I've tried the lot – margarine cartons, empty syrup tins, home made boxes, carefully sub-divided desk drawers, cardboard boxes. You name it. In the end I had to bite the bullet and buy a big metal rack containing many, many, many small clear plastic drawers. I eventually expanded this installation to include a couple of rails and about ten parts bins for the bigger items such as motors, trannies and large capacitors. The whole lot cost me about forty quid so it wasn't cheap, but now everything is correctly labelled and I can find what I want when I want it. If you can find a cheaper and more efficient way of storing components then by all means go ahead and do it – this one works for me.

- **Looking After Your Tools and Yourself -** This is nowhere near as difficult as it may sound. Ensure that the things such as knives and wire cutters which should be sharp stay sharp, and replace items such as drill bits, files and screwdrivers as and when they become blunt. Always use the right tool for the right job and remember that if you are forcing a tool whilst grunting and sweating profusely with effort then you're using the wrong tool, or the right tool in the wrong way. Stop what you are doing immediately, take stock of the situation, and decide on an alternative approach before either you or the workpiece are damaged.

  The moving parts of tools such as pliers and wire strippers do need to be lubricated from time to time. Use only a small amount of light machine oil and lubricate only the linkages, not the business end of the tool.

  A good mental attitude in the workshop can make the difference between a successfully completed project of which you are rightfully proud, and an expensive piece of junk which you will have to bin. If you watch a competent amateur or a trained profession at work, you'll see a composed individual working calmly and methodically towards completion of the task in hand. His work area will be well laid out and tidy, and all tools will be carefully selected and correctly used. Only commence work on your projects when you want to. Never rush to finish a project and don't drive yourself unnecessarily. Remember that this is a hobby – it should make you happy and enable you to relax and derive reward from what you do. Before starting work, make sure that external distractions are at a minimum, that the lighting is correct, and that you are warm and comfortable.

  Never work beyond your capabilities. If you don't know what you are doing then stop immediately and use a suitable source of information to learn what you need to know. You have to be very lucky indeed for all your guesses to be correct ones when working with electronics. More often than not the end result of an uninformed guess is sparks, blown fuses, smoke, a nasty smell and a steep retrospective learning curve.

  Though the temptation when you're on a roll is to keep working until you finish, don't become over tired as you will make mistakes and you may end up ruining an expensive piece of equipment in the process. Its far better to do a bit and then have a break than to go for an all night bender in the hope that everything will go well and the coffee won't unexpectedly dry up.

  As a final note, if you like a couple of beers as I do, don't attempt to do any work after imbibing. I once wired up a guitar amp PSU whilst a little intoxicated, and I thought the standard of my work was excellent. It was fortunate that I decided not to throw the switch as a quick inspection the next morning showed badly routed wires, missing or

badly positioned insulation sleeving, and a pair of incorrectly wired bridges – this is just why so many countries now have strict drink driving laws. After a couple of beers you think you're doing fine when what you're actually doing is making a right pig's ear out of the whole thing.

And finally, why not go back and re-read section 2 on electrical safety ?

In the next instalment, we'll look at correct soldering technique, and the right way to assemble circuits in general, I'll then complete the series (for now) with a short article on casing up your new masterpiece safely and attractively.

Return to Articles Index

Back to Home Page

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Andrew Walmsley and Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Andrew Walmsley) and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Andrew Walmsley and Rod Elliott.

Page created and copyright (c) 18 Mar 2001

Elliott Sound Products **Beginners' Guide to Transformers**

Copyright © 2001 - Rod Elliott (ESP)
Page Published & Updated 24 Mar 2001

Transformers - The Basics
Section 1

Articles Index

Main Index

Contents - Section 1

- Preface
- Introduction
- 1. Magnetism and Inductors
- 2. Magnetic Core Terminology
- 3. How a Transformer Works
- 4. Interesting Things About Transformers
- 5. Examples of Transformer Uses
  - 5.1 Push-Pull Valve Output Stage
  - 5.2 Single Ended Triode (SET) Output
  - 5.3 Line Level Applications
- 6. Safety
- 7. Noise
- Section 2
- References

Preface

One thing that obviously confuses many people is the idea of flux density within the transformer core. While this is covered in more detail in Section 2, it is important that this section's information

is remembered at every stage of your reading through this article. For any power transformer, the maximum flux density in the core is obtained when the transformer is idle. I will reiterate this, as it is very important ...

For any power transformer, the maximum flux density is obtained when the transformer is idle.

The idea is counter-intuitive, it even verges on not making sense. Be that as it may, it's a fact, and missing it will ruin your understanding of transformers. At idle, the transformer back-EMF almost exactly cancels out the applied voltage. The small current that flows maintains the flux density at the maximum allowed value, and represents iron loss (see Section 2). As current is drawn from the secondary, the flux falls slightly, and allows more primary current to flow to provide the output current.

It is not important that you understand the reasons for this right from the beginning, but it *is* important that you remember that *for any power transformer, the maximum flux density is obtained when the transformer is idle*. Please don't forget this

---

Introduction

As you look through this article, you may be excused for exclaiming "This is for beginners? - the man's mad. Mad, I tell you!" This is probably fair comment, but transformers are not simple, and there is no simple way to provide all the information you need to understand them properly. There are sections here that probably go a little bit deeper than I originally intended, but were just too interesting to leave out.

There are parts of this article you may want to skip over, but I suggest that you do read all of it if you can. A full understanding to the extent where you can design your own transformer is not the aim, but the majority of the information is at the very least interesting, and will further your general electronics knowledge.

For those who wish to delve deeper, Section 2 does just that. It is recommended reading, even for beginners, as there is a great deal to be learned about transformers, despite their apparent simplicity.

The principles that allow us to make use of electro-magnetism were only discovered in 1824, when Danish physicist Hans Oersted found that a current flowing through a wire would deflect a compass needle. A few years after this, it was found that a moving magnetic field induced a current into a wire. From this seemingly basic concept, the field of electromagnetism has grown to the point that society as we know it would not exist without the many machines that make use of these discoveries.

Transformers are essential for all modern electronics equipment, and there are very few devices

that do not use them. Each transformer type has a specific use, and it is uncommon that a transformer made for one application can be used for another (quite different) purpose.

Before embarking on a description of the different types, the basic theory must be understood. All transformers use the same basic principle, and only the finer points ever change. A transformer works on the principle of magnetic coupling to transfer the energy from one side (winding) to the other.

Transformers are bi-directional, and will work regardless of where the input is connected. They may not work as well as they otherwise might, but basic functionality is unchanged. An ideal transformer imposes no load on the supply (feeding the primary) unless there is a load across the secondary - real life components have losses, so this is not strictly true, but the assumption can be used as a basis of understanding.

Power transformers are rated in Volt-Amps (VA). Using Watts is of no use, since a load that is completely reactive dissipates no power, but there are still Volts and Amps. It is the product of "real" voltage and current that is important - a wattmeter may indicate that there is little or no real power in the load, but the transformer is still supplying a voltage and a current, and will get hot due to internal losses regardless of the power.

Transformer cores have a quoted permeability, which is a measure of how well they "conduct" a magnetic field. Magnetism will keep to the path of least resistance, and will remain in a high permeability core with little leakage. The lower the permeability, the greater is the flux leakage from the core (this is of course a gross simplification, but serves well enough to provide an initial explanation of the term).

A transformer may be made with various materials as the core (the magnetic path). These include

- Air - provides the least coupling, but is ideal for high frequencies (especially RF). Permeability is 1.
- Iron - A misnomer, since all "iron" cored transformers are steel, with various additives to improve the magnetic properties. Permeability is typically about 500 and upwards.
- Powdered Iron - Steel magnetic particles formed into a core and held together with a bonding agent, and fired at high temperature to create a ceramic-like material with very good properties at medium to high frequencies (over 1 MHz). Especially suited to applications where there is a significant DC component in the winding or for very high power. Permeability is typically 40-90.
- Ferrite - A magnetic ceramic, usually using exotic magnetic materials to obtain extremely high permeability and excellent high frequency performance (from 50kHz to over 1MHz). An astonishing range of different formulations is available for different applications. Permeability is from about 500 up to 9,000 or more.

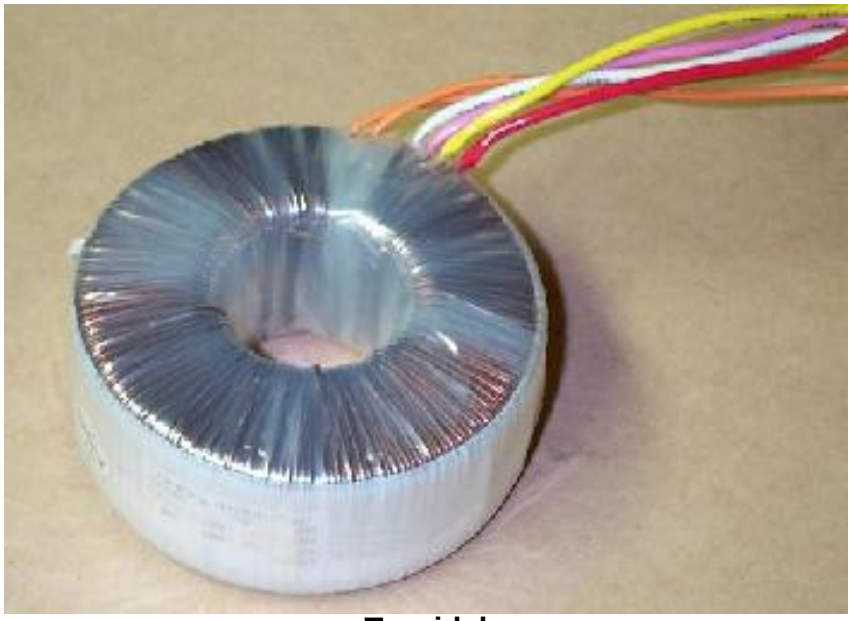
Technically, powdered iron and ferrites are both classified as soft (see below) ferrites, but they have very different characteristics, even within the same "family". They are generally unsuitable for low frequency operation, except at low levels. Ferrites are often used as signal transformers (such

as isolation transformers for telecommunications or other small signal applications), where the high permeability makes them an ideal choice for small size and high inductance.

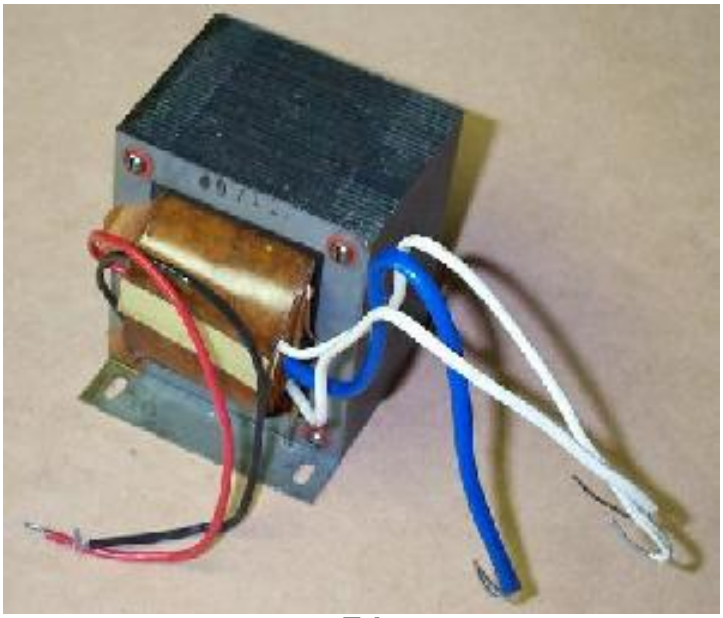
Core materials are generally classified as "soft" - this has nothing to do with their physical properties (they are all hard to very hard), but is a reference to their ability to retain magnetism (remanence). Hard magnetic materials are used for magnets, and they have a very high remanence, which is to say they retain a very large proportion of the original magnetic field that was induced into them during manufacture.

All switchmode power supplies use ferrite transformers, since conventional laminations cannot be made thin enough to prevent huge losses in the core.

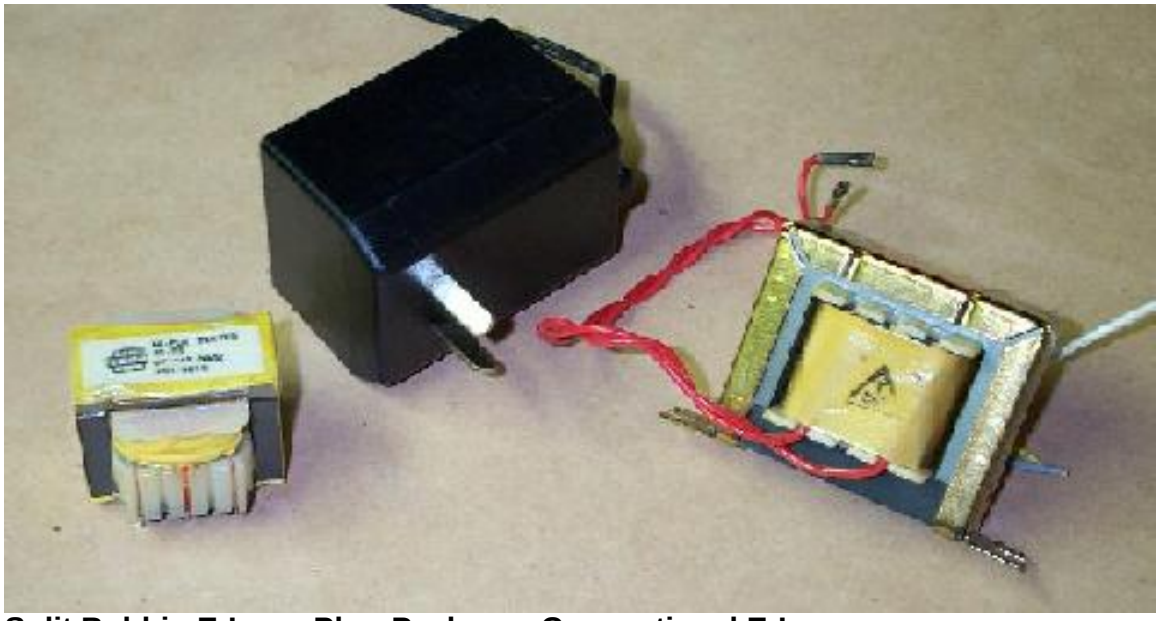
Many limitations exist in any core material. For low frequency power applications, grain-oriented silicon steel (about 4% silicon) is by far the most common, as it has a very high flux density before saturation. Almost all other materials are inferior in this respect, one of the main reasons this material is still so common.



Toroidal



E-I



Split Bobbin E-I Plug-Pack Conventional E-I

A small sample of some transformers is shown above (not to scale). The toroidal and E-I transformers are the same power rating, and a small selection of little transformers and a plug-pack (wall transformer, wall-wart, etc) are shown as well.

---

1. Magnetism and Inductors

The transformer is essentially just two (or more) inductors, sharing a common magnetic path. Any two inductors placed reasonably close to each other will work as a transformer, and the more closely they are coupled magnetically, the more efficient they become.

When a changing magnetic field is in the vicinity of a coil of wire (an inductor), a voltage is induced into the coil which is in sympathy with the applied magnetic field. A static magnetic field has no effect, and generates no output. Many of the same principles apply to generators, alternators, electric motors and loudspeakers, although this would be a very long article indeed if I were to cover all the magnetic field devices that exist.

When an electric current is passed through a coil of wire, a magnetic field is created - this works with AC or DC, but with DC, the magnetic field is obviously static. For this reason, transformers cannot be used directly with DC, for although a magnetic field exists, it must be changing to induce a voltage into the other coil.

Try this experiment. Take a coil of wire (a loudspeaker crossover coil will do nicely for this), and a magnet. Connect a multimeter - preferably analogue) to the coil, and set the range to the most sensitive current range on the meter. As you move the magnet towards or away from the coil, you will see a current, shown by the deflection of the meter pointer. As the magnet is swung one way, the current will be positive, the other way - negative. The higher the coil's inductance and the stronger the magnet (and/ or the closer it is to the coil), the greater will be the induced current.

Move the magnet slowly, and the current will be less than if it is moved quickly. Leave it still, and there is no current at all, regardless of how close the magnet may be. This is the principle of magnetic induction, and it applies to all coils (indeed to all pieces of wire, although the coil makes the effect much greater).

If you now take a handful of nails and place them through the centre of the coil, you will see that the current is increased many times - the magnetic field is now more concentrated because the steel nails make a better magnetic path than air.

The ability of a substance to carry a magnetic field is called permeability, and different materials have differing permeabilities. Some are optimised in specific ways for a particular requirement - for example the cores used for a switching transformer are very different from those used for normal 50/60Hz mains transformers.

The permeability of transformer cores varies widely, depending on the material and any treatment that may be used. The permeability of air is 1, and most traditional cores have a much higher (i.e. > 1) permeability. A couple of notable exceptions are aluminium and brass, which are sometimes used to reduce the inductance of air cored coils in radio frequency (RF) work. This is much less common than a ferrite "slug" core, which increases the inductance and is used to tune some RF transformers.

As well as permeability, magnetic cores (with the exception of air) have a maximum magnetic flux they can handle without saturation. In this context, saturation means the same as in most others - when a towel is saturated, it can hold no more water, and when a magnetic core is saturated, it can carry no more magnetic flux. At this point, the magnetic field is no longer changing, so current is not induced into the winding.

You will be unable to saturate your nails with the magnet, as there is a very large air gap between the two pole pieces. This means that the core will always be able to support the magnetic flux, but the efficiency is also very much lower because the magnetic circuit is open. Nearly all the transformers you will see have a completely closed magnetic circuit, to ensure that as much of the magnetism induced into the core as possible will pass through the winding(s).

There are some cases where a tiny air gap will be left deliberately, and this is done routinely when a transformer or coil must sustain a significant DC component as well as the AC. This is covered briefly below, but there is more on this subject in the second section of the article.

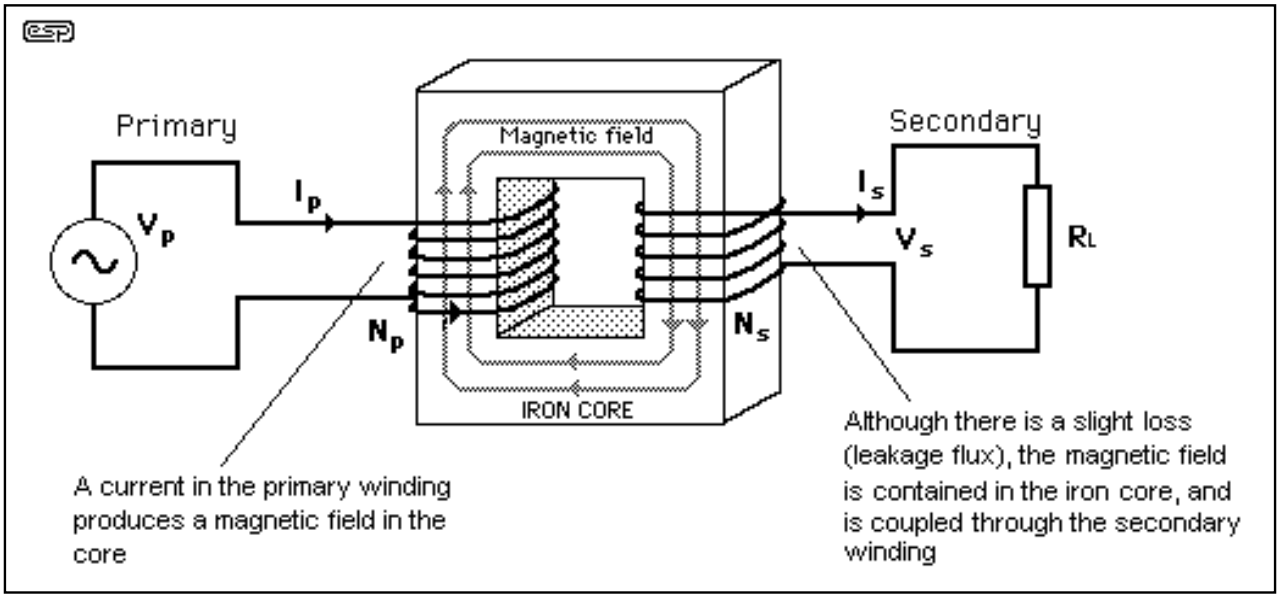


Figure 1.1 - Essential Workings of a Transformer

Figure 1.1 shows the basics of all transformers. A coil (the primary) is connected to an AC voltage source - typically the mains for power transformers. The flux induced into the core is coupled through to the secondary, a voltage is induced into the winding, and a current is produced through the load.

The diagram also shows the various parts of a transformer. This is a simple transformer, with two windings. The primary (denoted as such during the design) will induce a magnetic field into the core in sympathy with the current produced by the applied AC voltage. The magnetic field is concentrated by the core, and nearly all of it will pass through the windings of the secondary as well, where a voltage is induced. The core in this case is typical of the construction of a "C-Core" transformer, where the primary and secondary are separated. More common is the "traditional" EI (ee-eye) type, which although somewhat out of favour these days is still used extensively. This is shown below.

The magnitude of the voltage in the secondary is determined by a very simple formula, which determines the "turns ratio" (N) of the component - this is traditionally calculated by dividing the secondary turns by the primary turns ...

1.1.1 **N = Ts / Tp**

Tp is simply the number of turns of wire that make up the primary winding, and Ts is the number of turns of the secondary. A transformer with 500 turns on the primary and 50 turns on the secondary has a turns ratio of 1:10 (i.e. 1/10 or 0.1)

1.1.2 **Vs = Vp * N**

Mostly, you will never know the number of turns, but of course we can simply reverse the formula so that the turns ratio can be deduced from the primary and secondary voltages ...

1.1.3 **N = Vs / Vp**

If a voltage of 240V (AC, naturally) is applied to the primary, we would expect 24V on the secondary, and this is indeed what will be measured. The transformer has an additional useful function - not only is the voltage "transformed", but so is the current.

1.1.4 **Is = Ip / N**

If a current of 1A were drawn by the primary in the above example, then logically a current of 10A would be available at the secondary - the voltage is reduced, but current is increased. This would be the case if the transformer were 100% efficient, but even this - the most efficient "machine" we have - will sadly never be perfect. With large transformers used for the national supply grid, the efficiency of the transformers will generally exceed 95%, and some will be as high as 98% (or even more).

Smaller transformers will always have a lower efficiency, but the units commonly used in power amplifiers can have efficiencies of up to 90% for larger sizes. The reasons for the lost power will become clear (I hope) as we progress. For the time being, we shall consider the transformer to be "ideal" (i.e. having no losses) for simplicity.

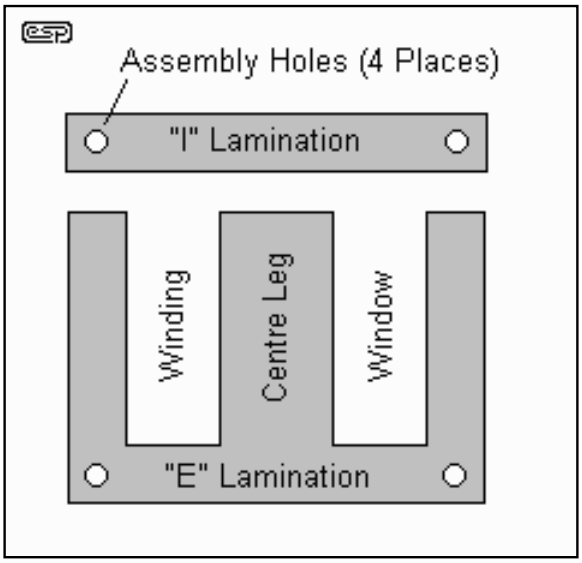


Figure 1.2 - E-I Laminations

The conventional E-I lamination set is still extensively used, and a few pertinent points are worth mentioning. The centre leg is always double the width of the outer legs to maintain the cross-sectional area. Likewise, the "I" lamination and the "back" of the E are the same width as (or sometimes slightly larger than) the outer legs. The winding window is where the copper windings live, and in a well designed transformer will be almost completely full. This maximises the amount of copper and reduces resistive losses because the windings are as thick as they possibly can be.

---

2. Magnetic Core Terminology

This list is far from complete, but will be sufficient to either get you started or scare you away. I have included the symbols and units of only three of the entries below, since most are of no real interest.

Coercivity -is the field strength which must be applied to reduce (or *coerce*) the remanent flux to zero. Materials with high coercivity (e.g. those used for permanent magnets) are called *hard*. Materials with low coercivity (those used for transformers) are called *soft*.

Effective Area - of a core is the cross sectional area of the centre limb for E-I laminations, or the total area for a toroid. Usually this corresponds to the physical dimensions of the core but because flux may not be distributed evenly the manufacturer may specify a value which reflects this.

Effective length - of a core is the distance which the magnetic flux travels in making a complete circuit. Usually this corresponds closely to the average of the physical dimensions of the core, but because flux has a tendency to concentrate on the inside corners of the path the manufacturer may specify a value for the effective length.

Flux Density - (symbol; B, unit; Teslas (T)) is simply the total flux divided by the effective area of the magnetic circuit through which it flows.

Flux linkage - in an ideal inductor the flux generated by one turn would be contained within all the other turns. Real coils come close to this ideal when the other dimensions of the coil are small compared with its diameter, or if a suitable core guides the flux through the windings.

Magnetomotive Force - MMF can be thought of as the magnetic equivalent of electromotive force. It is the product of the current flowing in a coil and the number of turns that make up the coil.

Magnetic Field Strength - (symbol: H, unit; ampere metres ($A\ m^{-1}$)) when current flows in a conductor, it is always accompanied by a magnetic field. The strength, or intensity, of this field is proportional to the amount of current and inversely proportional to the distance from the conductor (hence the -1 superscript).

Magnetic Flux - (symbol: Φ; unit: Webers (Wb)) we refer to magnetism in terms of lines of force or flux, which is a measure of the total amount of magnetism.

Permeability - (symbol; μ, units: henrys per metre (Hm^{-1}) is defined as the ratio of flux density to field strength, and is determined by the type of material within the magnetic field - i.e. the core material itself. Most references to permeability are actually to "relative permeability", as the permeability of nearly all materials changes depending upon field strength (and in most cases with temperature as well).

Remanence - (or remnance) is the flux density which remains in a magnetic material when the externally applied field is removed. Transformers require the lowest possible remanence, while permanent magnets need a high value of remanence.

I mention these here for the sake of completeness, but their real importance is not discussed further in Section 1. Section 2 of this article will revisit the terms, and their importance is somewhat enhanced in context.

---

3. How a Transformer Works

At no load, an ideal transformer draws virtually no current from the mains, since it is simply a large inductance. The whole principle of operation is based on induced magnetic flux, which not only creates a voltage (and current) in the secondary, but the primary as well! It is this characteristic that allows any inductor to function as expected, and the voltage generated in the primary is called a "back EMF" (electromotive force). The magnitude of this voltage is such that it almost equals (and is effectively in the same phase as) the applied EMF.

No simple calculation can be made to determine the internally generated voltage. As described in Part 1 of this series, for a sinusoidal waveform, the current through an inductor lags the voltage by 90 degrees. Since the induced current is lagging by 90 degrees, so too is the induced voltage (back EMF). For the sake of simplicity, imagine an inductor or transformer (no load) with an applied voltage of 100V. For the effective back EMF to equal the applied AC voltage (as it must), the actual magnitude of the induced voltage (back EMF) is 141V. If this is all to confusing, don't worry about it. Unless you intend to devote your career to transformer design, the information is actually of little use to you, since you are restrained by the "real world" characteristics of the components you buy - the internals are of little consequence. Even if you do devote your life to the design of transformers, this info is still merely a curiosity for the most part, since there is little you can do about it.

When you apply a load to the output (secondary) winding, a current is drawn by the load, and this is reflected through the transformer to the primary. As a result, the primary must now draw more current from the mains. Somewhat intriguingly perhaps, the more current that is drawn from the secondary, the original 90 degree phase shift becomes less and less as the transformer approaches full power. The power factor of an unloaded transformer is very low, meaning that although there are volts and amps, there is relatively little power. The power factor improves as loading increases, and at full load will be close to unity (the ideal).

Now, another interesting fact about transformers can now be examined.

We will use the same example as above. A 240V primary draws 1A, and the 24V secondary supplies 10A to the load. Using Ohm's law, the load resistance (impedance) is therefore 24/10 = 2.4 Ohms. The primary impedance must be 240/1 = 240 Ohms. This is a ratio of 100:1, yet the turns ratio is only 10:1 - what is going on?

The impedance ratio of a transformer is equal to the square of the turns ratio ...

3.1.1 **$Z = N^2$**

Transformers are usually designed based on the power required, and this determines the core size for a given core material. From this, the required "turns per volt" figure can be determined, based on the maximum flux density that the core material can support. Again, this varies widely with core materials.

A rule of thumb can be applied, that states that the core area for "standard" (if indeed there is such a thing) steel laminations (in square centimetres) is equal to the square root of the power. Thus a 625VA transformer would need a core of (at least) 25 sq cm, assuming that the permeability of the core were about 500, which is fairly typical of standard transformer laminations. This also assumes that the core material will not saturate with the flux density required to obtain this power.

The next step is to calculate the number of turns per volt for the primary winding. This varies with frequency, but for a 50Hz transformer, the turns per volt is (approximately) 45 divided by the core area (in square centimetres). Fewer turns are needed for a 60Hz transformer, and the turns per volt will be about 38 / core area. Higher performance core materials may permit higher flux densities, so fewer turns per volt might be possible, thus increasing the overall efficiency and regulation. These calculations must be made with care, or the transformer will overheat at no load.

For a 625VA transformer, it follows that you will need about 432 turns for a 240V primary, although in practice it may be less than this. The grain-oriented silicon steels used in better quality transformers will often tolerate much higher total flux per unit area, and fewer turns will be needed.

You can determine the turns per volt of any transformer (for reasons that will become clearer as we progress) by adding exactly 10 turns of thin "bell wire" or similar insulated wire to an existing transformer, wound over the existing windings. When powered from the correct nominal supply voltage, measure the voltage on the extra winding you created, and divide by 10 to obtain the turns per volt rating for that transformer.

Now, what earthly use is this to you? Well, you might be surprised at what you can do with this knowledge. Assume for a moment that you have a transformer for a fair sized power amplifier. The secondary voltage is 35-0-35V which is much too high to power the preamp circuit or even its power supply - but you will be able to do that with a single 16V winding. Another transformer would normally be used, but you can also add the extra winding yourself. This is almost too easy with toroidal transformers, but with others it may not be possible at all. If the transformer uses (say) 2 turns per volt, a mere 32 extra turns of bell wire (or similar) will provide 16V at the typical 100mA or so you will need. Add a 10% margin, and you still have only 36 turns to add, and this can be done in a few minutes. Make sure that the extra winding is securely taped down with a good quality tape (Kapton is highly recommended if you can get it). Do ***not*** use ordinary electricians' tape - it is not designed for the temperature that transformers may operate at under consistent load.

NOTE: Ensure that there is no possibility whatsoever of the added winding shorting between turns - this will cause the smoke to escape from the insulation in a spectacular fashion, and you may ruin the transformer itself.

---

4. Interesting Things About Transformers

As discussed above, the impedance ratio is the square of the turns ratio, but this is only one of many interesting things about transformers (well, *I* think they are interesting, anyway :-)

For example, one would think that increasing the number of turns would increase the flux density, since there are more turns contributing to the magnetic field. In fact, the opposite is true, and for the same input voltage, an increase in the number of turns will decrease the flux density and vice versa. This is counter-intuitive until you realise that an increase in the number of turns increases the inductance, and therefore reduces the current through each coil.

I have already mentioned that the power factor (and phase shift) varies according to load, and this (although mildly interesting) is not of any real consequence to most of us.

A very interesting phenomenon exists when we draw current from the secondary. Since the primary current increases to supply the load, we would expect that the magnetic flux in the core would also increase (more amps, same number of turns, more flux). In fact, the flux density decreases! In a perfect transformer with no copper loss, the flux would remain the same - the extra current supplies the secondary only. In a real transformer, as the current is increased, the losses increase proportionally, and there is slightly less flux at full power than at no load.

---

5. Examples of Transformer Uses

This is only a brief discussion of the many uses of transformers. I have avoided switchmode supplies in this section, and will only present the most common linear applications. Power supply applications are covered more fully in Section 2, and also in the article on Linear Power Supply Design.

It would be impossible to cover all aspects of transformers and their uses, since they are so diverse and are used in so many different things. Computer network interface cards, modems, through to power amplifiers and microwave ovens, car and marine ignition systems, Tesla coils and moving coil phono preamps are a very small sampling of the diversity of the humble transformer (well, maybe it is not so humble after all :-)

5.1 - Push-Pull Valve Output Stage

Apart from the obvious uses in power supplies, transformers are used in other areas as well. Valve power amplifiers nearly all use a transformer for the output stage, and this converts the high impedance of the anodes to the loudspeaker impedance, as well as providing the voltage feed to the output valves. No biasing or other support components have been shown here - for more information on this, have a look at How Amplifiers Work. <

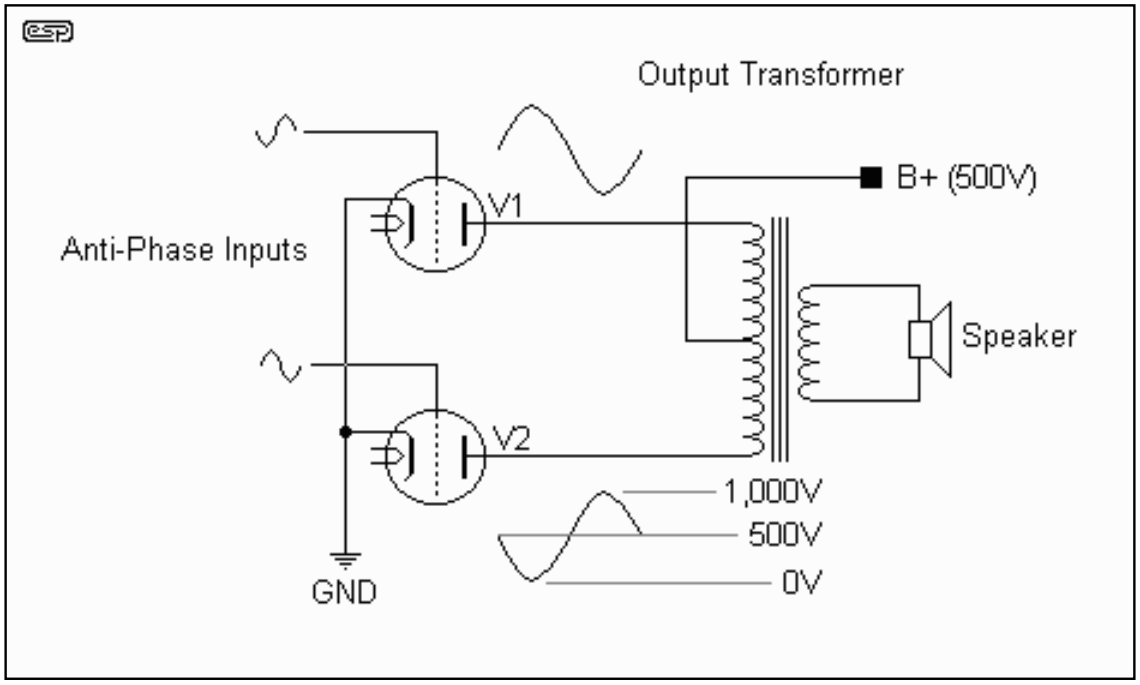


Figure 5.1 - Push-Pull Valve Output Stage

Figure 5.1 shows how this works. The primary winding acts in a manner that may surprise you at first, but it is quite in accordance with all the theory. The supply voltage shown is 500V, and we will assume that the valve can turn on hard enough to reduce this to zero alternately at each end of the winding. This is never the case, because valves do not have a low enough internal impedance, but it makes the explanation simpler :-)

Neither valve will draw appreciable current with no signal, and the amount drawn does not magnetise the core. The reason is simple - an equal amount of current is drawn through each section of the primary winding, but effectively in opposite directions. The magnetic field created by one half of the winding is cancelled by that from the second half, leaving a nett steady state magnetic field of zero.

When valve V1 turns on completely, the voltage at its end of the winding is reduced to zero, and the voltage at the anode of V2 is 1,000 volts. This must be the case, or the transformer theory is in tatters. The primary is operating as an "auto-transformer". Likewise, when V1 turns off and V2 turns on, the situation is reversed. You may well ask why 2 valves are needed at all? The voltage from one valve is quite capable of swinging the voltage from one extreme to the other, it would seem.

This is not the case. Since the valve can only turn on, it will only be able to supply current for 1/2 of the waveform. A Class-A push-pull design will normally have each valve carrying 1/2 of the maximum peak current required. In the case of a push-pull design, there is no core saturation because of the DC current (which cancels out as before), so although two valves are needed, the transformer will be smaller and will have very much better performance. Single-ended Class-A amps require a very large core with an air-gap to prevent saturation. This reduces the performance of the transformer dramatically, and increases distortion and gives a poorer low frequency response because of the lower inductance. High frequencies can also be adversely affected,

because the air-gap causes some of the magnetic flux to "leak" out of the core. This is the cause of leakage inductance (covered in more detail in Section 2).

It is worth noting that the effective peak to peak swing across the entire transformer primary is 2,000V. When V1 is turned on completely, it has zero volts (for our example only) at the plate, and V2 turns on it has a plate voltage of 1,000V. V2 has exactly the same voltage peaks, but they are 180 degrees out of phase. The total voltage across the transformer is therefore the sum of the two voltages. From an AC perspective, the B+ supply line can be considered the same as zero volts (remember it will be bypassed with a large capacitance).

The RMS voltage is easily calculated from the standard formula ...

5.1.1 **$V_p = V_{p-p} / 2$**

To obtain the peak value from peak to peak, then ...

5.1.2 **$V_{rms} = V_p / \sqrt{2}$**

To find the RMS value.

In this case, the peak to peak voltage is 2,000V, so peak is 1000V. The RMS value is 707V.

5.2 Single Ended Triode (SET) Output

Figure 5.2 shows the basic arrangement of a SET amplifier output stage. The full DC current must flow through the transformer primary, and as discussed above, an air-gap must be introduced into the core to prevent saturation. Because an air gap reduces the efficacy of the magnetic path, the core needs to be considerably larger than would otherwise be the case.

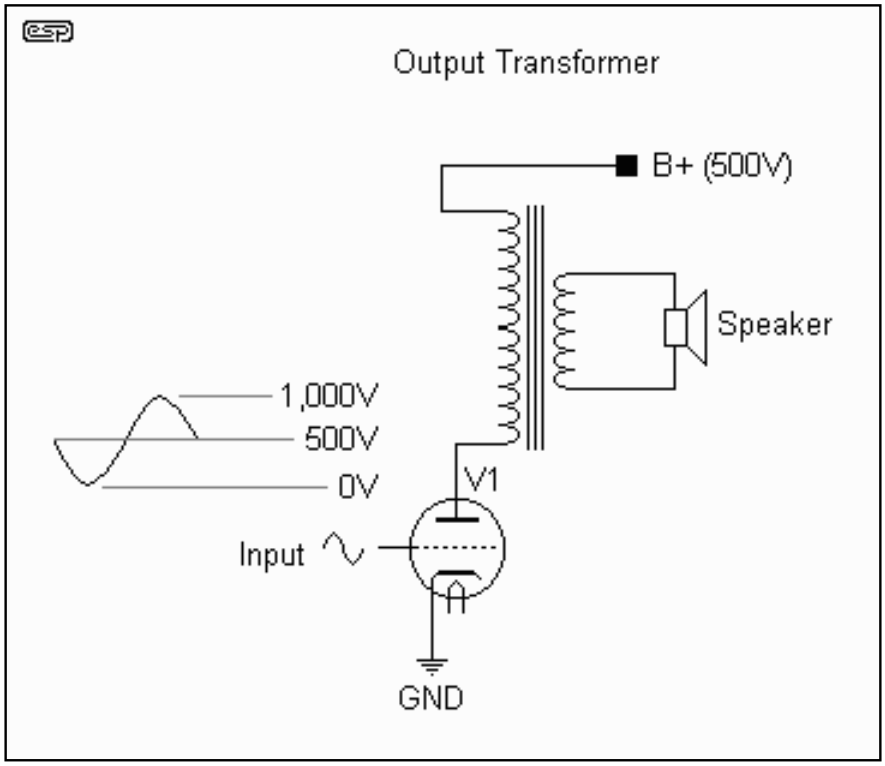


Figure 5.2 - Single Ended Triode Output Stage

The core operates with only one polarity of flux, which varies with the signal. One might think that this alone would reduce distortion, since the flux never crosses the zero point, but this is not the case. It is still necessary for the flux to change, and the characteristics of magnetic materials indicate that the resistance to change (rather than the absolute polarity of the magnetic field) is the dominant factor. The valve (and transformer primary) must now carry a current equal to the peak AC current demanded by the load - subject to the transformation ratio, of course. Maximum negative swing (valve turned on) will double this current, and it will be reduced to nearly zero as the valve turns off (positive swing). As the current is reduced below the average standing (quiescent) current, the voltage across the transformer increases in the opposite polarity - hence the fact that the plate voltage exceeds the supply voltage.

For the same power output, the valve in a single ended circuit must be considerably larger than that required for a push-pull circuit, because of the higher dissipation needed for the extra current. There are also many other issues with this arrangement, and they will be covered in more detail in Section 2.

Not the least of these is that the (probable) advantage of the additional voltage swing when using a centre tapped transformer is now gone, so the maximum RMS voltage that can be developed is 353V - a significant drop in primary AC voltage.

5.3 Line Level Applications

Transformers are also used for "line-level" low power applications, typically balanced microphone inputs and line output stages. A transformer is unsurpassed for real-world balanced circuits, as the input or output is truly floating, and requires no connection to earth. This means that common mode signals (i.e. any signal that is common to both signal leads) are almost completely rejected.

Figure 5.3 shows a transformer balanced input, converting to unbalanced. The signal is amplified, and sent to the output transformer for distribution as a balanced signal again. The "amplifier" will typically be a mixing console, and will take mic or line level signals as the input (having run from the stage to the mixing area), and the final mixed output is sent back to the stage for the main (Front of House) public address amplifiers and speakers. There may be in excess of 100 metres of cable from the microphone to the mixer and back to the main amps, and barely any noise will be picked up in the process.

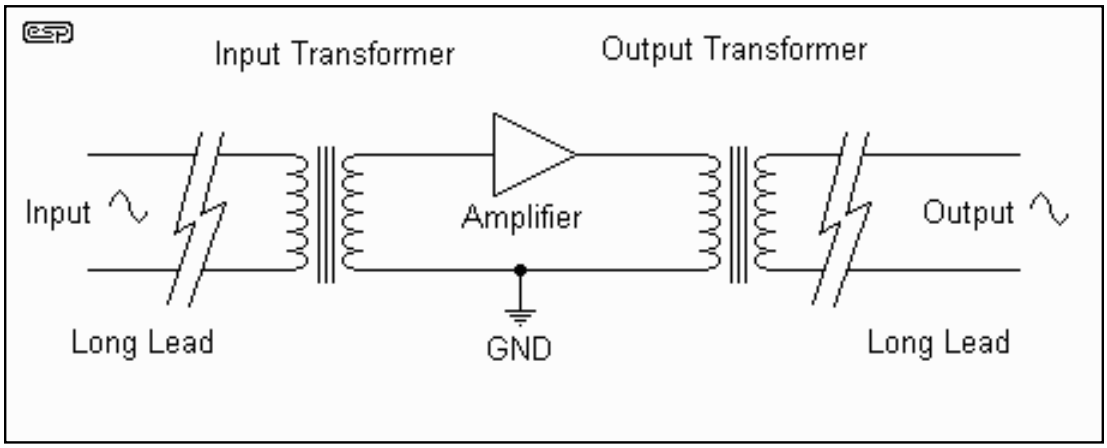


Figure 5.3 - Balanced Microphone and Line Outputs

The telephone system used to be completely dependent on transformers to feed the signal from the exchange (or Central Office in the US) to the customer premises equipment (CPE). The phone switch used in offices, (PABX - Private Automatic Branch Exchange, or PBX for the US) equipment still uses transformers for nearly all incoming circuits whether analogue or digital. The principle is exactly the same as for the audio application shown above, except that for telephone circuits there is usually a DC voltage present to power the CPE (in the case of a home telephone) and to provide some basic signalling. All modern PABX circuits use ferrite cored transformers, with DC isolation circuitry to ensure that no DC flows in the transformer windings, as this degrades the performance in the same way as with the output transformer for a SET power amplifier.

Audio applications for transformers in balanced circuits came from the telecommunications industry where the concepts were first thought of. A telephone line may be 4km or more in length, and is not shielded, so a method of eliminating hum and noise was essential.

---

6. Safety

Safety is a major consideration for any power transformer (and in the case of telecommunications, the isolating transformers), and electrical contact between primary and secondary must not be allowed under ***any*** realistic fault condition. All countries have safety standards that apply to transformers where electrical isolation is important, and if in any doubt about the safety of a transformer for a particular purpose, make sure that you verify that the transformer complies with the relevant standard(s). It is well beyond the scope of this article to cover all the possibilities of standards and compliance issues, so I shall leave that to you.

Many power transformers are fitted with an internal "once only" thermal fuse that will become open circuit in the event that a preset temperature is exceeded. This temperature is chosen to be the maximum safe temperature of the windings before the insulation melts or breaks down, so in the event of a fault, the thermal fuse will open before the insulation is damaged and the component becomes potentially dangerous. It also helps to prevent the risk of fire (and no, this is not intended to be humorous - a friend of mine had his house burned to the ground because of a faulty power transformer in a video recorder - as determined by the fire investigators. True story!). See Figure 6.1 (below) as an example of how bad things can get if the transformer is not protected.

Once the thermal fuse opens, the transformer must be discarded, as it is usually not possible to gain access to the fuse for replacement. This is not as silly as it may sound, since the thermal effects on the insulation cannot be predicted, and the transformer may be unsafe if it were still able to be used.

There are transformers that are designed to be "intrinsically safe", and these usually have the windings on separate sections of the core, not in physical contact with each other. If the core is connected to the electrical safety earth (which is usually a requirement), no method of failure (including a complete meltdown) in the primary will allow mains voltage to appear at the secondary. Side by side windings are the next safest, and although primary and secondary are on the same bobbin, the material used is selected to withstand high temperatures and will maintain separation of the windings. Toroidal cores and other concentrically wound transformers are the least safe, since the only separation between primary and secondary is a rather thin layer of insulation. This is one of the reasons that thermal fuses are often used with toroids.

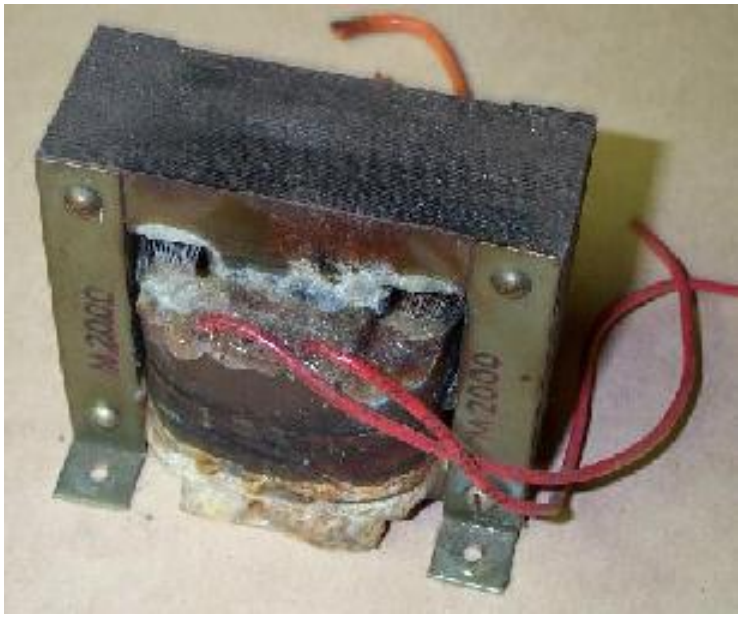


Figure 6.1 - Transformer Meltdown

Figure 6.1 shows a transformer I removed from a repair job. It is a complete meltdown, and the remains of the plastic bobbin can be seen quite clearly. In any circuit, it is extremely important to protect the user from coming into contact with the mains should this happen. In this case, the bobbin had melted away from the windings, dribbled on the base of the equipment, and generally

made a big mess. Despite all this, there was no electrical connection between primary and secondary or the laminations. This was a well made transformer (it failed due to gross continuous overload, not any failure in the transformer itself).

Proper safety earthing is the only real way to ensure that a transformer that fails catastrophically (such as that shown) does not cause the chassis to become live - not all transformers are created equal when safety is concerned. Correct fusing will ensure that the fuse blows - hopefully *before* the electrical safety is compromised. A thermal fuse would have prevented the situation from becoming as bad as shown, but the transformer would have been just as dead.

---

7. Noise

Transformers make noise. This is not only the electrical noise that is created by the nasty current waveform through the windings, diodes and into the filter capacitors, but actual audible noise. One source is winding vibration, due to the wire moving because of the magnetic field and the current flowing through the conductors. This is to be avoided at all costs, since constant vibration will eventually wear away the insulation, the windings will short circuit, and the transformer is ruined. Fortunately, this is rather unusual, but it can (and does) happen on occasion.

Most of the noise is from the laminations or other core material, which contract when subjected to an intense magnetic field. This is called magnetostriction, and happens to a greater or lesser degree with all magnetic materials. A stethoscope will verify the source of the noise, and there is little or nothing that will stop it. A resilient mounting will stop most of the noise from being acoustically amplified by the chassis, and generally the noise will be worse at no load. In some cases, a transformer may have been designed for 60Hz, but is used at 50Hz. In this case, the flux density will probably exceed the maximum allowable for the core, and the transformer will get much hotter than it should, and will almost certainly be a lot noisier as well. Toroidal transformers will generally be much quieter than EI laminated (i.e. conventional) types.

Most (all?) transformers designed specifically for 60Hz will eventually fail with 50Hz mains, due to overheating. The reverse is not true, and 50Hz transformers can be operated quite safely on 60Hz.

Another problem with E-I laminations is that they may not have been fastened together well enough, and this allows the outer laminations in particular to vibrate. Better quality conventional transformers will commonly be impregnated with varnish (sometimes under vacuum) and baked in a moderate oven until tender .... oops, I mean until the varnish is completely dry. This binds the laminations and windings together, preventing noise, and also making the transformer more resistant to damage by water or other contaminants, and/ or under conditions of high humidity (such as in the tropics).

Section 2

Click here to view the second part of the article. As I am sure you have noticed, transformers are

not so simple after all.

References

These references are common to both sections of the article, although most only apply to Section 2. Countless different Web pages were researched during the compilation of these articles, and although some were interesting, the majority were of minimal use. Of those who I actually remember (a daunting task in itself, considering the sheer amount of searching I had to do), I must "thank" the following Web pages (in alphabetical order) ...

- Amidon
- ATDL School (US Army)
- Jensen Transformers
- Mitchell Electronics Corporation
- Tomi Engdahl - (ePanorama.net)

In addition, I have used various other references, but notably (in order of usefulness) ...

- Radiotron Designer's Handbook - F Langford-Smith (4th Edition)
- Magnetism: quantities, units and relationships (E-mail: R.Clarke@surrey.ac.uk)
- Basic Electronics - Grob, Third Edition (McGraw-Hill)
- Motorola Rectifier Applications Handbook (3rd Edition)

Download Silvio Klaic's neat little transformer calculator from his website

---

The following (slightly edited) definitions are from Units of Measurement

Units of Measurement site copyright by Russ Rowlett and University of North Carolina at Chapel Hill.
Definitions used by permission.

Tesla (T) - flux density (or field intensity) for magnetic fields (also called the magnetic induction). The intensity of a magnetic field can be measured by placing a current-carrying conductor in the field. The magnetic field exerts a force on the conductor which depends on the amount of the current and the length of the conductor. One Tesla is defined as the field intensity generating one newton of force per ampere of current per meter of conductor. Equivalently, one Tesla represents a magnetic flux density of one Weber per square meter of area. A field of one Tesla is quite strong: the strongest fields available in laboratories are about 20 Teslas, and the Earth's magnetic flux density at its surface, is about 50 microteslas (μT). One Tesla equals 10,000 gauss. The Tesla, defined in 1958, is named after Nikola Tesla (1856-1943), whose work in electromagnetic induction led to the first practical generators and motors using alternating current (much to the annoyance of Edison, who claimed DC was "safer").

Weber (Wb) - magnetic flux. "Flux" is the rate (per unit of time) in which something crosses a surface perpendicular to the flow. In the case of a magnetic field, then the magnetic flux across a perpendicular surface is the product of the magnetic flux density, in Teslas, and the surface area, in square metres. If a varying magnetic field passes perpendicularly through a circular loop of conducting material (one turn), the variation in the field

induces a electric potential in the loop. If the flux is changing at a uniform rate of one Weber per second, the induced potential is one volt. This means that numerically the flux in webers is equal to the potential, in volts, that would be created by collapsing the field uniformly to zero in one second. One Weber is the flux induced in this way by a current varying at the uniform rate of one ampere per second. The unit honours the German physicist Wilhelm Eduard Weber (1804-1891), one of the early researchers of magnetism.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 17 March 2001./ updated 25 Jun 2005

Elliott Sound Products **Beginners' Guide to Transformers - Part 2**

Copyright © 2001 - Rod Elliott (ESP)
Page Last Updated 03 Apr 2001

Transformers - The Details
Section 2

Articles Index

Main Index

Contents - Section 2

- Section 1
- Introduction
- 8. Windings in Series and Parallel
  - 8.1 Series Connections
  - 8.2 Parallel Connections
- 9. Valve Output Transformer Example Calculation
- 10. Compromises
- 11. Losses
  - 11.1 Iron Losses
  - 11.2 Copper Losses
  - 11.3 Regulation
  - 11.4 Other Losses
- 12. Sample Measurements
- 13. Core Styles
  - 13.1 Air Gaps
- 14 Core Materials
- 15 Distortion
- References
- Units
- Magnetic Terminology

Introduction

For those brave souls who have ploughed their way through the first section - I commend you! As you have discovered, transformers are not simple after all, but they are probably far more versatile than you ever imagined. They are, however, real world devices, and as such are prey to the failings of all real components - they are imperfect.

This section will concentrate a little more on the losses and calculations involved in transformer design, as well as explain in more detail where different core styles are to be preferred over others. Again, it is impossible to cover all the possibilities, but the information here will get you well on your way to a full understanding of the subject.

The first topic may seem obvious, but based on the e-mails I get, this is not the case. Transformers can have multiple windings, and these can be on the primary or secondary. Windings can be interconnected to do exciting and different things, but from a safety perspective it is ***imperative*** that primary and secondary windings are kept segregated.

There are several references to "shorted turns" within this article. If any two turns of a winding short to each other, the current flow is limited only by the DC resistance of the shorted section of the winding. The current flow is enormous, and with even one shorted turn, the transformer is no longer serviceable and must be discarded or rewound. No shield or other conductive material may be wrapped around a core and joined, as this creates a shorted turn capable of possibly hundreds of amperes. The exception to this is the magnetic shield sometimes used with E-I laminated transformers, but this is wrapped around the entire transformer, and is not considered as a "turn" as it is not in the winding window with the primary and secondary.

It is also worth noting that a transformer behaves quite differently depending upon whether it is driven from a voltage source (i.e. very low impedance, such as a transistor amp or the mains) or a current source or intermediate impedance. This will be covered in a little more detail further on in this article.

---

8. Windings in Series and Parallel

Many transformers are supplied with two (or more) secondaries. In many cases, the data sheet will indicate that the windings may be connected in parallel or series. For example, a toroidal transformer may be rated at 2 x 25V at 5A (250VA). With the windings in parallel, the available current is 10A, but only for a single voltage of 25V AC. Connect the windings in series, and you get 50V at 5A, or by referencing the centre tap to earth, the familiar 25-0-25 designation.

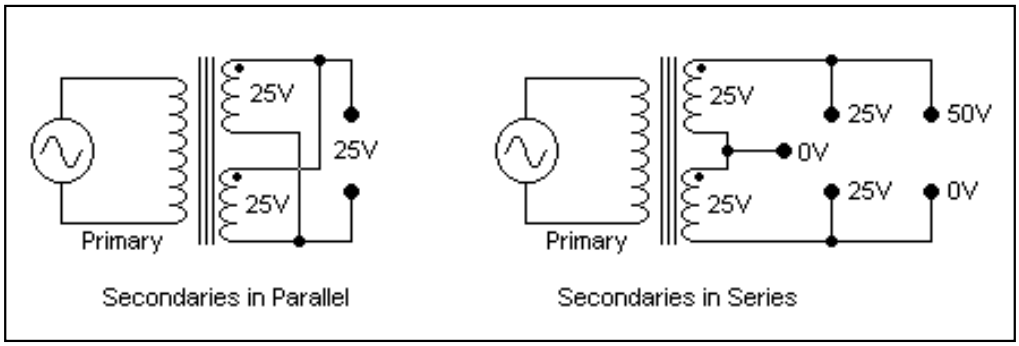


Figure 8.1 - Windings in Series and Parallel

There are some rules that apply to winding interconnections - if you break them, you may break your transformer as well. Note the dots on the windings - this is the traditional way to identify the start of a winding, so that the phase may be determined.

Antiphase wiring will not harm a transformer when wired in series (although the zero volts output for equal windings is somewhat limited in usefulness). Parallel antiphase connection will destroy the transformer unless the fuse blows - which it will do mightily. Always use a fuse when testing, as a simple mistake can be rather costly without some form of protection for the transformer and house wiring!

8.1 Series Connections

Windings may be connected in series regardless of voltage. The maximum current available is the rating specified for the ***lowest*** current winding. Windings may be connected so as to increase or decrease the final voltage. For example, dual 25V windings may be connected so as to produce 50V or zero volts - although the latter is not generally useful :-)

When windings are connected in phase the voltages add together, and if connected out of phase, they subtract. A 50V, 1 amp winding and a 10V 5 amp winding may therefore be connected to provide any of the following ...

- 10V @ 5A - The 10V winding by itself
- 50V @ 1A - The 50V winding by itself
- 60V @ 1A - Both 50V and 10V windings, connected in series and in phase
- 40V @ 1A - Both 50V and 10V windings, connected in series and out of phase

The above example was used purely for the sake of example (such a transformer would not be useful for most of us), but the principle applies for all voltages and currents. Series connections are sometimes used in the primaries as well, mainly for equipment destined for the world market. There are several common mains supply voltages, and primary windings are connected in various combinations of series and parallel to accommodate all the variants.

8.2 Parallel Connections

Parallel connection of transformer windings is permitted in one case only - the windings must have exactly the same voltage output, and must be connected in phase. Different current capacities are not a problem, but it is rare to find a transformer with two windings of the same voltage but different current ratings.

Even a 1V difference between winding voltages will cause ***big*** problems. A typical winding resistance for a 5A winding might be 0.25 ohm. Should two such windings be connected in parallel, having a voltage difference of 1V, there will be a circulating current limited only by the resistances of the windings. For our example, the total winding resistance is 0.5 ohm, so a circulating current of 2A will flow between the windings, and this is completely wasted power. The transformer will get unexpectedly hot, and the maximum current available is reduced by the value of the circulating current.

Should the windings be connected out of phase, the circulating current will be possibly 100A or more, until the transformer melts or the fuse blows. The latter is generally to be preferred.

The transformer manufacturer's specifications will indicate if parallel operation is permitted. If you are unsure, measure the voltages carefully, and avoid parallel connection if the voltages differ by more than a couple of hundred millivolts. There will always be a difference, and only the manufacturer's winding tolerances can predict what it will be. With toroidal transformers, the windings are often bifilar, meaning that the two windings are wound onto the transformer core simultaneously. The tolerance of such windings is normally very good, and should cause no problems.

---

9. Valve Output Transformer Example Calculation

In Section 1, I described a very basic push-pull valve output stage. Now it is time to examine this a little more closely. We shall use the same voltages as were obtained in the basic description of Section 1 - an RMS voltage of 707V. It must be said that the following is not intended to be an accurate representation of valves, as the losses in real life are somewhat higher than indicated here. This is for example only. We shall also take the (typical) losses as 10%, and adjust the secondary impedance accordingly.

A valve (tube) amplifier is required to drive an 8 ohm loudspeaker. The primary impedance (called the Plate-Plate impedance for a push-pull amplifier) is 6,000 Ohms, and the supply voltage is 600V. Allowing for losses of 100V across each valve, the maximum voltage swing on the plates (anodes) of the valves is 1kV p-p (or effectively 2kV peak to peak on the transformer primary). What is the output power?

Secondary impedance will be 7.2 ohms, based on the 10% loss ...

Zs = 8 / 1.1 = 7.2 ohms

The impedance ratio is calculated first ...

Z = 6,000 / 7.2 = 833

The turns ratio may now be determined

N = √833 = 28.8 (29:1)

The voltage ratio is the same as the turns ratio, so the peak to peak voltage to the speaker is

Vs (p-p) = Vp / N = 2,000 / 29 = 69V

To convert this to RMS ...

Vp = 1/2 Vp-p = 34.5V
RMS = peak * 0.707 = 24V
Power is therefore 24^2 / 8 = 72W

Notice that at each calculation, the figures were rounded to the closest (or next lowest) whole number.

This was for convenience, but the way I did it also gives a conservative rating that is more likely to be met in practice.

Ouch! Sorry, that *was* a bit nasty for this time of day :-)

A bit nasty or not, it is a reasonable representation of the reality of an output transformer design, but naturally real (as opposed to my "invented" figures) will be substituted. Typically the losses across the output valves will often be far greater than indicated here. but that depends on the valves used (and the topology - triodes behave very differently from pentodes or tetrodes).

Just to complete this section and to put the above into perspective, I have included a few figures (taken from the 1972 Miniwatt Technical Data manual) for the EL34/ 6CA7 power pentode - quite possibly my all-time favourite output valve.

Class	Mode *	Plate Volts	Plate Current	Screen Volts	Screen Current	Grid Bias	Load Impedance	Power Output	Comments
Class-A	S-E	250	100	265	15	-13V	2,000	11W	Plate supply = 265V, THD** 10%
Class-AB	P-P	375	2 x 75 ## 2 x 95	365	2 x 11.5 2 x 22.5	-19V	3,400 (p-p) #	35W	Cathode bias resistor 130 ohms, common screen resistor, 470 ohms, THD 5%
Class-B	P-P	775	2 x 25 2 x 91	400	2 x 3.0 2 x 19	-39V	11,000 (p-p)	100W	Plate supply, 800V, THD 5%
Class-A (Triode)	S-E	375	70	-	-	-25V	3,000	6W	Cathode bias resistor 370 ohms, Screen tied to plate, 400V plate supply, THD 8%
Class-AB (Triode)	P-P	400	2 x 65 2 x 71	-	-	-28V	5,000 (p-p)	16W	Screen tied to plate, Cathode bias resistor 220 ohms, THD 3%

Table 9.1 - Abbreviated Data For EL34 Power Pentode

*	S-E: Single Ended, P-P: Push-Pull
**	THD - Total Harmonic Distortion (this is for the valves only, and does not include transformer distortion)
#	p-p: Plate to Plate impedance
##	First figure is no load, second figure is full power

As can be seen quite readily, the distortion of the S-E configurations is much worse than the push-pull versions. Not only that, but (to maintain relevance :-) the transformers are larger and harder to design, and even then will be worse than their push-pull counterparts. In the maximum efficiency configuration, power output is 100W, and distortion is still lower than for either of the single ended configurations. The losses across the output valve in this mode are about 58V, but are considerably higher for any of the cathode biased versions - as one might expect.

This will be revisited in another article on the design of valve amplifiers.

---

10. Compromises

It is very important that the core does not saturate (see below), since there will be no variation of flux, no back EMF, and excessive current will be drawn - especially at no load. The final design of any

transformer is a huge compromise, and there is a fine line between a transformer that will give acceptable regulation and one that gets too hot to touch at no load.

Somewhat surprisingly, the flux density in the core actually decreases with increased load current drawn from the secondary. Even though the primary is drawing more current, this is transferred to the secondary and thence the load - it does not cause the flux density to increase. The flux density decreases largely due to primary resistance, which causes the effective primary voltage to decrease. Any voltage lost to resistance (remember Ohm's law?) is voltage that is "lost" to the transformer, and serves no function in the transformation process. It does cause the transformer to get hot (or hotter) than at no load.

Also, the normal variation of mains voltage must be allowed for. A transformer running at the very limit of saturation at nominal supply voltage will overheat if the mains is at the upper (normal) limit. A transformer that is designed to run at the limit will have superior regulation compared to a more conservative design, but this is of little consequence if it fails in normal use.

For audio transformers, there are even more compromises.

---

11. Losses

As discussed earlier, a transformer is a real component, and therefore has losses. These are divided into two primary types, but there are other "hidden" losses as well. All losses reduce efficiency, and affect frequency response. The low frequency limit is determined by the primary inductance, and this is proportional to the area (and consequent mass) of the transformer core. High frequency losses are caused by eddy currents in the core (see below), and by leakage inductance and winding capacitances.

None of these can be eliminated, but by careful selection of core material, winding style and operational limits, they can be reduced to the point where the transformer is capable of doing the job required of it.

11.1 Iron (Core) Losses

Core losses are partly the result of the magnetising current, which must keep forcing the magnetic field in the core to reverse in sympathy with the applied signal. Because the direction of flux is constantly changing, the transformer core is subject to a phenomenon called hysteresis, shown in Figure 11.1

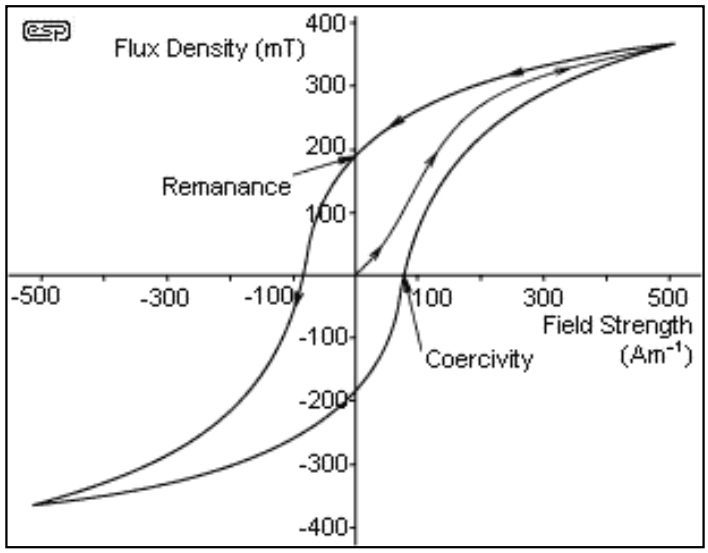


Figure 11.1 - The Hysteresis Loop

When the magnetomotive force is reversed in a magnetic material, the residual magnetism (remanence - also known as remnance) in the core tries to remain in its previous state until the applied flux is too great (coercivity). It will then reverse, and the same situation will occur twice for each cycle of applied AC. The power required to force the flux to change direction is the hysteresis loss, which although usually small, is still significant. I am not about to go into great detail on this, but a Web search will no doubt reveal more information than you will ever need.

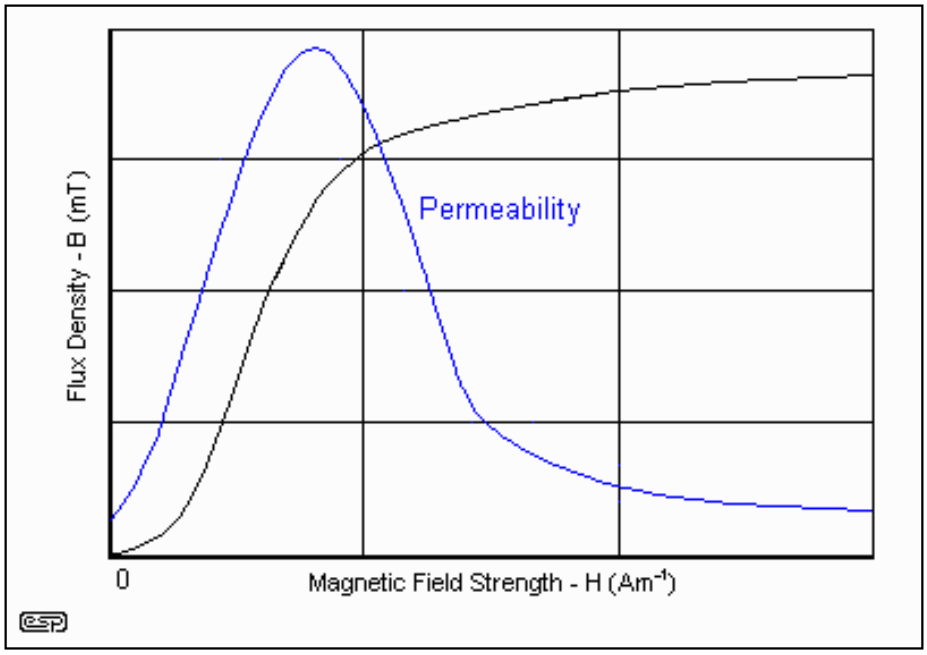


Figure 11.2 - B-H Curve

As can be seen from the two magnetic field drawings, the flux density (B) is dependent upon the applied magnetic field strength (H). For the example shown, the "knee" of the curve coincides with the point where permeability starts to fall. Above this, a progressively larger change in the magnetic field is required to increase the flux density. This is saturation, and most transformers will be designed to operate at or below the knee. Above the knee is dangerous, as a small increase in applied voltage will not produce the required increase in back EMF, and the primary current will increase disproportionately to the rise in voltage. In other words, the transformer will be too sensitive to applied voltage, and will

possibly self destruct if the mains voltage were even slightly higher than normal.

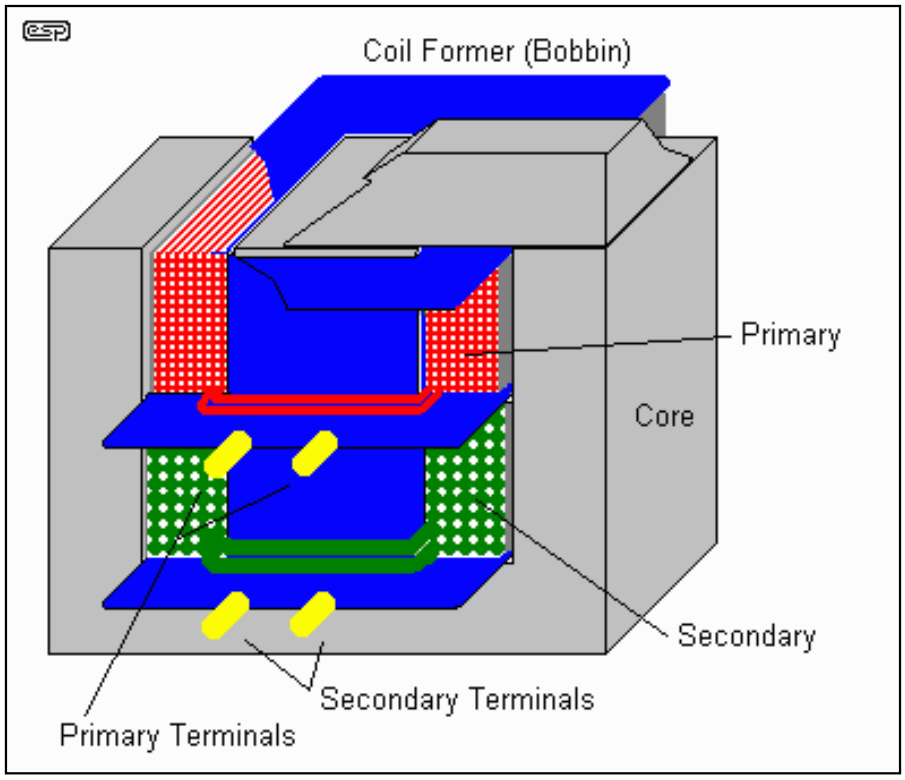


Figure 11.3 - Cutaway View of a Transformer

The transformer shown is a "split bobbin" type, having separate sections on the former for the primary and secondary windings. This reduces the capacitance between windings, and also provides a safety barrier between the primary and secondary. For some applications, this is the only winding method that meets safety standards. It is also very simple to add an electrostatic shield between the windings - a flat plate of thin metal is cut so that it can be slipped over the bobbin, and the ends are insulated so that it does not create a shorted turn. This is connected to earth, and prevents noise from being capacitively coupled between windings. The shield would logically be placed on the secondary side of the bobbin divider for safety.

In addition, there are so-called "eddy current" losses. These are small circulating currents within the magnetic core, as shown (exaggerated) in Figure 11.4, and these cause the core material itself to get hot. Each of these eddy current loops acts as a tiny shorted turn to the transformer, and to reduce the effect, the core is laminated - i.e. made from thin sheets of steel, insulated from each other. The thinner the laminations, the smaller are the eddy current losses, but they will never be eliminated. Eddy current losses increase with frequency, requiring different techniques for high frequency operation, and are the major contributor to the iron losses in any transformer.

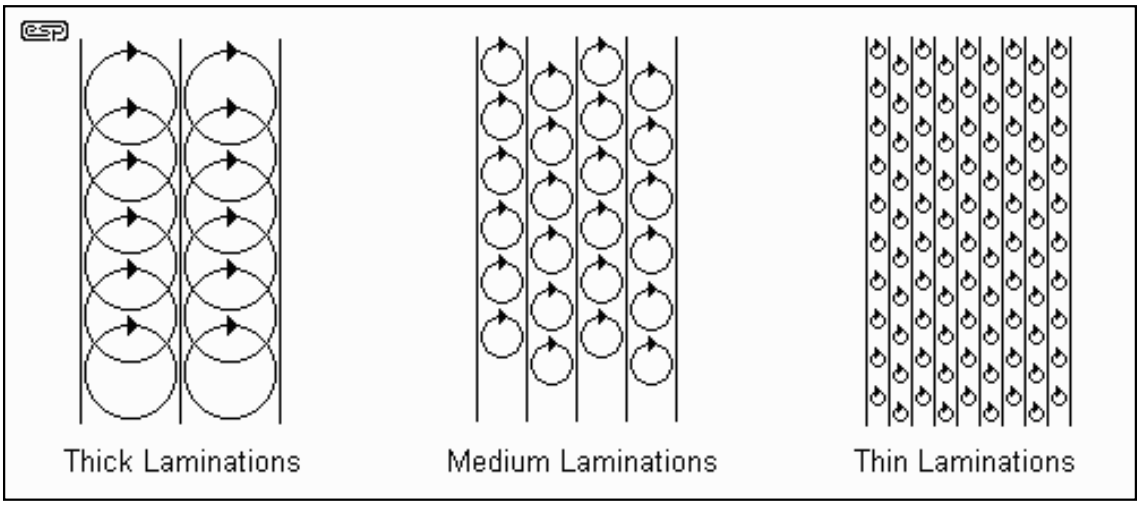


Figure 11.4 - Eddy Currents in Laminations

The eddy currents are shown for three lamination thicknesses. Although not shown (for the sake of clarity), the current loops are constantly overlapping, and are effectively infinite in number. The thick laminations allow the loops to be larger, and therefore the lamination section is cut by more magnetic "lines" of force, so the currents (and losses) are larger. For high frequencies (above 10kHz), it is generally not possible to make laminations thin enough to prevent the losses from becoming excessive, and ferrite materials are preferred. These effectively have a huge number of incredibly small magnetic particles, all insulated from each other, and eddy current loops are very small indeed. Even so, ferrite materials are normally specified up to a few hundred kilo-Hertz for power applications, before the losses become too great again.

Iron losses of both types are the primary source of losses in any transformer that is operating at no load or only light loading. At no load, the core flux density is at its maximum value for any given applied voltage / frequency combination. Power transformers are usually designed to operate below the knee of the saturation curve (this is essential with toroidal types), with sufficient safety margin to ensure that the core can never saturate.

Saturation involves a dramatic loss of permeability (and therefore inductance), and causes the primary current to rise disproportionally to an increase of voltage. Significant waveform distortion occurs once the core starts to saturate.

As a load is drawn from the secondary, the primary must supply more current, and this means that the resistance of the primary winding becomes significant. Any voltage 'lost' to winding resistance is effectively no longer part of the applied voltage, so core flux is reduced.

11.2 Copper Losses

Following on from the previous point, the voltage lost to winding resistance is copper loss, and all such losses must be dissipated as heat. Consider a transformer at idle, with 240V on the primary. The primary resistance may be in the order of 5 ohms, and the idle current perhaps 20mA. The loss is determined by the normal power formula, and in this case is ...

$$\mathbf{P = I^2 * R} = 0.02^2 * 5 = 2\text{mW}$$

V = R * I = 5 * 0.02 = 100mV

For all intents and purposes, the full 240V is applied to the primary. When the transformer is loaded, this changes. Let's assume 1A primary current and look at the figures again ...

P = I² * R = 1.00²* 5 = 5W
V = R * I = 5 * 1.00 = 5V

Now, the effective primary voltage is only 235V, because 5V is 'lost' due to winding resistance. Naturally, if the voltage is lower, the flux density must also be lower.

Minimising copper loss in both primary and secondary is essential, but there are limits to what can be achieved. These are imposed by the available space for the winding, and just how much copper the manufacturer can get into that space. Allowance must still be made for insulation and manufacturing tolerances.

You will see that in Figure 11.3 the windings are shown stacked directly on top of each other. Surely a more efficient winding can be made by making use of the "valleys", minimising the winding height and allowing heavier windings. Ah, if only life were that simple! The windings are traditionally made from left to right, then right to left, so the turns in each layer are at a slight angle relative to the layer below or above. It is therefore not possible to interleave the windings properly, and if you were to design a transformer based on the erroneous assumption that interleaving would work, the winding would not fit into the window.

For the normal layered construction (i.e. primary closest to the core, and secondary over the top), we also have to allow for insulation between primary and secondary, and in some cases additional insulation is used between layers of larger transformers because of the large voltage difference between the outer limits of each winding. These are another set of compromises that must be made, all of which mean that the windings must be thinner than we might like, and thus the losses are increased.

Because any length of wire has resistance, there will always be winding resistance. The greater the resistance for a given current, the more power is dissipated as heat - this is a complete loss. At no load, there is virtually no loss, since the currents are low, but as secondary current increases, so too do the copper losses.

Copper loss is the primary source of loss at any appreciable power from a transformer. Conventional rectifiers as used in semiconductor amplifier power supplies cause the resistance to be more significant than would otherwise be the case. See Linear Power Supply Design for more details on these losses.

Ultimately, copper losses limit the power available from a transformer. Since all copper loss results in heat, this becomes a limiting factor, so once you reach the point where the temperature rise cannot be limited to a safe value, the size of the core must be increased. This allows the manufacturer to use fewer turns per Volt, and the larger core has more space for the windings. The wire size can therefore be increased, so copper losses are brought back to the point where overheating is no longer a problem. This process continues from the smallest transformers to the largest - each size is determined by the VA rating and allowable temperature rise.

11.3 Regulation

Copper loss is responsible for a transformer's regulation - the ratio of voltage at no load versus full load. Regulation is almost always specified into a resistive load, which considering the way nearly everyone uses transformers, is virtually useless. It is rare that any transformer is operated into a purely resistive load - the vast majority will be used with a rectifier and filter capacitors, and the manufacturer's figure is worthless. Actually, it is worse than worthless, as it misleads the uninitiated to expect more voltage than they will obtain under load, and causes people grief as they try to work out why their amplifier (for example) gives less power than expected.

Naturally, there are some to whom any measurement is sacrilege, so none of this applies to them

The output voltage is (nearly) always specified at full load into a resistance. So a 50V, 5A transformer will give an output of 50V at a sinewave output current of 5A. If the regulation of this transformer were 4%, what is the no-load voltage?

The answer is 52V. Regulation is determined quite simply from the formula ...

$$\mathbf{Reg\% = (V_N - V_L) / V_L * 100 / 1}$$

Where $\mathbf{V_N}$ is no-load volts, and $\mathbf{V_L}$ is loaded volts.

As determined earlier, this assumes a sinusoidal output current, and this just does not happen with a rectifier / filter load. It may be found that this same transformer has an apparent regulation of 8 to 10% when supplying such a load. See Linear Power Supply Design for more information on this topic (there is little point in doing the article twice :-)

The regulation with rectifier loads is a complex topic, but you will need to know the ramifications before you start construction of your latest masterpiece, rather than find out later that all your work has resulted in much lower output power than you expected. Not that you can change it for any given transformer, but at least you will know what to expect.

11.4 Other Losses

Since the transformer is not an ideal device, it has unwanted properties apart from the losses described so far. The other losses are relatively insignificant for a power transformer, but become difficult to manage for transformers intended for wide bandwidth, such as microphone transformers and valve output transformers.

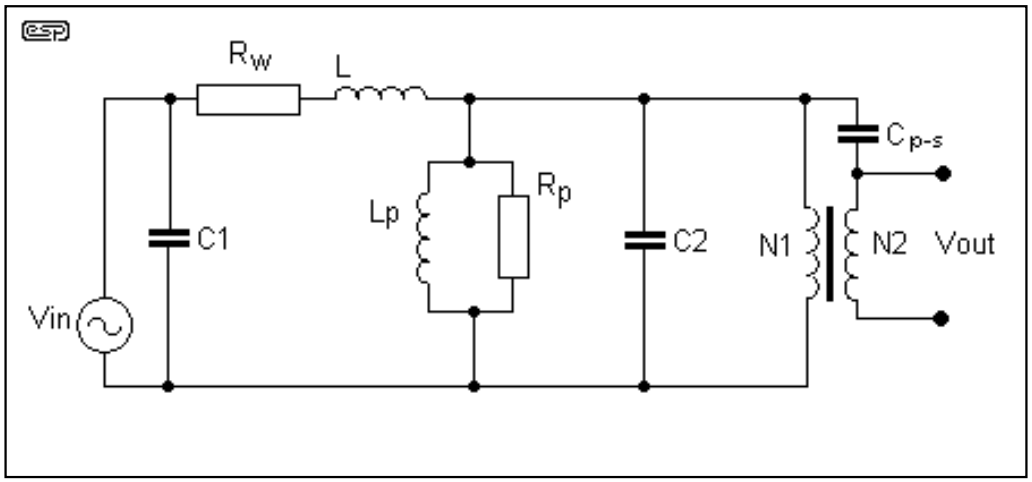


Figure 11.5 - Transformer Simplified Equivalent Circuit

The equivalent circuit shown in Figure 11.5 is greatly simplified, but serves to illustrate the points. Since the windings are usually layered, there must be capacitance (C1 and C2) between each layer and indeed, each turn. This causes phase shifts at high frequencies, and at some frequency, the transformer will be "self resonant". This is not a problem with power transformers, but does cause grief when a wide bandwidth audio transformer is needed.

In addition, there is some amount of the magnetic field that fails to remain in the core itself. This creates a "leakage" inductance (L) that is effectively in series with the transformer. Although small, it tends to affect the high frequencies in particular, and is especially troublesome for audio output transformers. This is typically measured with an inductance meter, with the output winding short circuited. Any inductance that appears is the direct result of leakage flux.

Lp is the primary inductance, and as you can see, there is a resistor in parallel (Rp). This represents the actual impedance (at no load) presented to the input voltage source, and simulates the iron losses. The series resistance (Rw) is simply the winding resistance, and is representative of the copper losses.

C_{p-s} is the inter-winding capacitance, and for power transformers can be a major contributor to noise at the output. This is especially irksome when the transformer is supplying a hi-fi system, and mains borne noise gets through and makes horrid clicks, electronic "farts", electric motor whine, and various other undesirable noises in the music. Toroidal transformers are very much worse than conventional (E-I) transformers in this respect, because of the large area of each winding. An electrostatic shield will all but eliminate such noises, but these are expensive and uncommon with toroids (pity).

The capacitance between the primary and secondary can become troublesome with this technique, and although possible, an electrostatic shield (actually, a number of electrostatic shields are needed) adds considerably to the cost, but creates a minimal overall benefit. This winding method is not used (or needed) with low frequency power transformers, and would lead to greatly reduced electrical safety because of the difficulty of insulating each section from the next. This latter problem also exists with an output transformer, but is easier to control because one side of the secondary is earthed.

Another problem exists when the capacitance between primary and secondary is high - electrical noise on the primary is coupled through the capacitance to the secondary. This can lead to mains noise getting through the entire power supply and into the amplifier in extreme cases. To combat this problem, an

electrostatic shield is sometimes used, and this is connected to earth. Note that the shield cannot be joined in a complete circle around the winding, as this would create a shorted turn that would draw a tremendous current and burn out the transformer.

There is a technique that is used for valve output transformers, shown in Figure 11.6 - you will not find this method used in power transformers, as it is completely unnecessary.

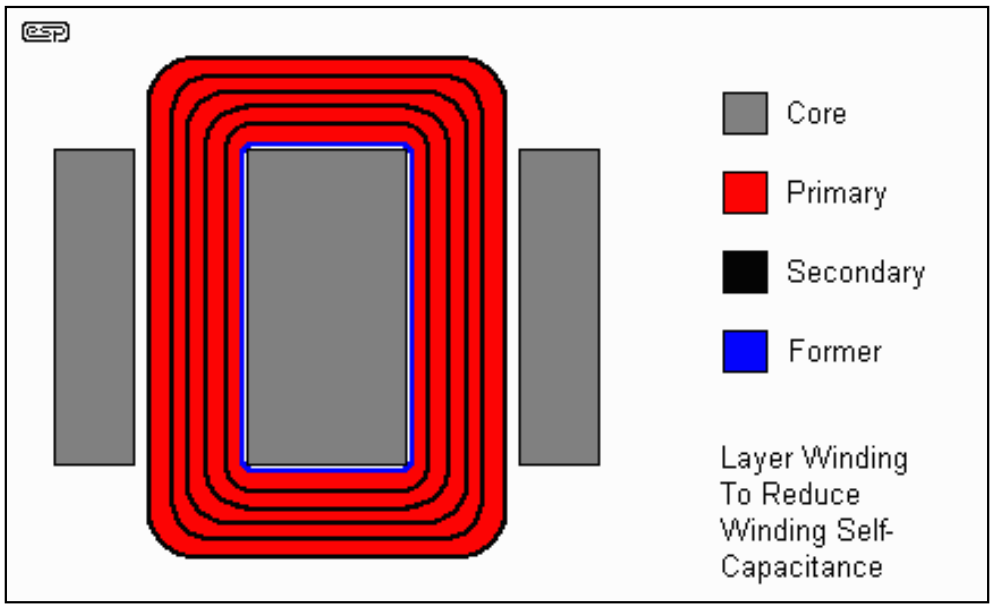


Figure 11.6 - Interleaved Winding for Extended HF Response

The trick to winding transformers to minimise the winding leakage inductance and self capacitance is called "interleaving", but this results in much greater inter-winding capacitance (between primary and secondary). The most common way this is done is to use a multi-segmented winding, as shown in the sectional drawing of Figure 11.6. This type of winding is (or was) quite common for high quality valve output transformers, and the extension of frequency on the top end of the audio spectrum is very noticeable.

---

12. Some Measurements

I measured the characteristics of a small selection of transformers to give some comparative data. I excluded regulation from this, as it is difficult to make a suitable load at short notice, and loads tend to get rather hot even with short usage. Most manufacturers will provide this information in their specifications, but be warned that this refers to a resistive load, and regulation will be *much* worse when supplying a conventional rectifier and filter capacitor (see above, and the Power Supply Design article for more details). It is also worth noting that an inductance meter is often of little use with large iron cored transformers, unless it operates with a sinusoidal waveform at (or near) the design frequency of the transformer. The inductances shown are calculated, since the measured values with my meter were a long way off.

Type	Rating	Inductance	Resistance	Turns/Volt	Magnetising	Core Loss	Reactance	Mass (kg)
Toroidal	500VA	34.7 H	2R4	2	22mA	5.28W	10.91k ohms	5.0
Toroidal	200VA	63 H	5R1	3	12mA	2.88W	20k ohms	2.7

E-I	200VA	4.36 H	6R6	2	175mA	42W	1.37k ohms	3.2

Table 7.1 - Measured Characteristics of Some Transformers

The toroidals are clear winners in terms of core loss in particular, but it must be said that the E-I transformer tested is not really representative of the majority. This is one of a few left that I had specially made, and they were deliberately designed to push the saturation limits of the core. These transformers run quite hot at no load, but give much better regulation than a more conservative design - the vast majority of such transformers. They were actually designed to run just above the "knee" of the B-H curve for the laminations used, and although somewhat risky, none has failed (to my knowledge) since they were made about 20 years ago. I use a pair of them in my hi-fi system, which has been in daily use for 10 years now.

It is also worth noting that the mass is lower than for a more "traditional" transformer design - a conventional design of the same power rating would be expected to weigh in at about 5kg.

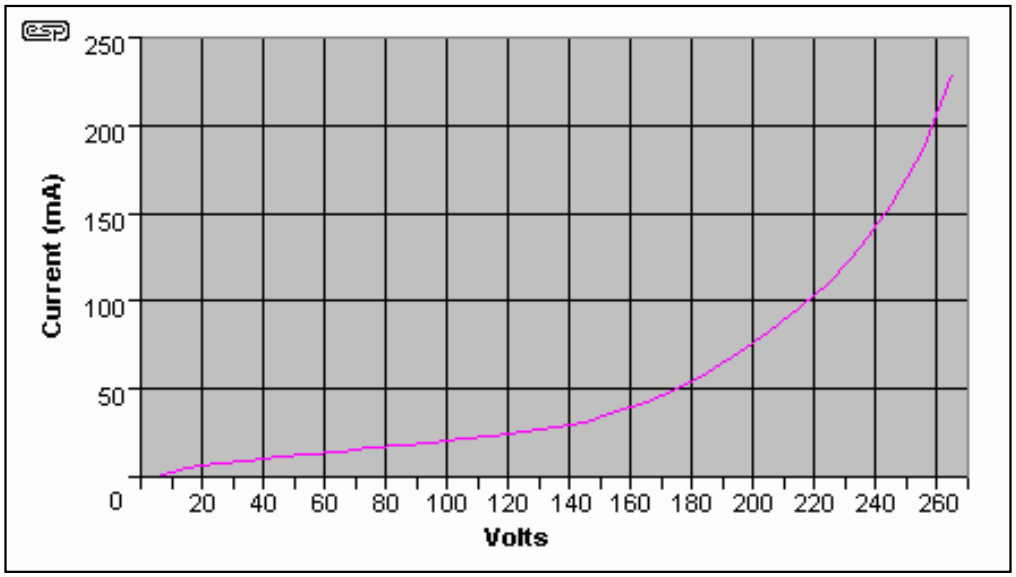


Figure 12.1 - Current vs. Voltage for the E-I Transformer

To take my measurements to the logical limit, I measured the magnetising current of my sample E-I transformer. Look closely at the graph in Figure 12.1, and you will see a typical BH curve (as shown in Figure 11.2 but with the axes reversed). As you can see, at 240V input, the transformer is operating at the knee of the curve, and is well on the way to saturation. There was no point doing this for the toroidals, as they are operated well below saturation level and I would be unable to (conveniently) measure them.

Toroids usually have a more pronounced knee, and a correspondingly steeper rise in current once the saturation limit has been reached. This is primarily because of the fully enclosed magnetic path, which has no air gaps at all (E-I laminated transformers have a small but significant gap where the laminations are joined. This is unavoidable in any practical transformer, but has little affect on performance in real life.

---

13. Core Styles

There is a huge array of different core shapes, and each has its own advantages and disadvantages. The two most common for commercial and DIY audio equipment are the E-I "shell" core and the toroidal core, but there are many others.

Ferrites in particular are moulded, and therefore have many specialised shapes to suit various applications, as well as the more traditional shapes shown below.

Toroidal cores are made from a continuous strip grain oriented silicon steel, and are bonded to prevent vibration and maximise the "packing density". It is important that there are no gaps between the individual layers, which will lower the performance of the core. The sharp corners are rounded off, and they are usually coated with a suitable insulating material to prevent the primary (which is always wound on first) from contacting the core itself.

I don't propose to even attempt them all, but one iron core that warrants special mention is the "C" core. These were once very popular, but have lost favour since suitable winding machines became available for toroids. They are still a very good core design, and are especially suited where an intrinsically safe transformer is required (i.e. where the primary and secondary windings are physically separated), and this technique also ensures that the inter-winding capacitance is minimal. C-cores are made by rolling a continuous strip into the desired shape, and after bonding, it is cut in half. To ensure the best possible magnetic coupling (i.e. no air gap), the cut ends are machined and polished as a pair - it is very important to ensure that the two are properly mated or unacceptable losses will occur. The core halves are commonly held together with steel banding, similar to that used for large transport boxes.

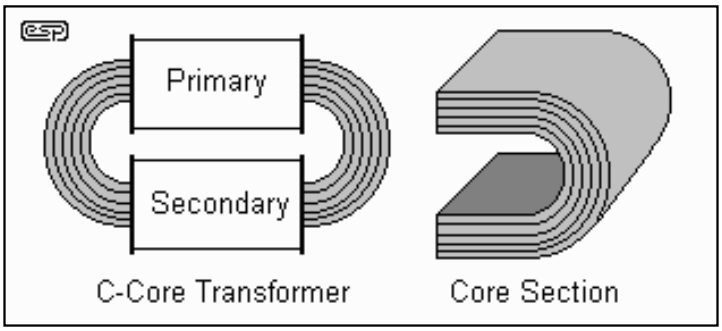


Figure 13.1 - C-Core Transformer

The main disadvantage of the single c-core arrangement shown above is that its leakage inductance is rather high. Although both windings could be placed onto a single bobbin with a pair of cores, it is more common to use four 'C' sections as shown below. This provides more iron (twice as much) and allows fewer turns for a given voltage. Naturally, the double c-core as shown below is not intrinsically safe, because both windings are wound together in the same way as for an equivalent E-I transformer.

C-cores are not as efficient as toroidal cores, but are easier to wind with conventional coil winding machines. The overall efficiency lies between the E-I core and the toroidal.

Figure 13.1A - Double C-Core Transformer

A sample of ferrite cores is shown in Figure 13.2 - this is but a small indication of the selections available, and most styles are also available in many different grades to suit specific applications.

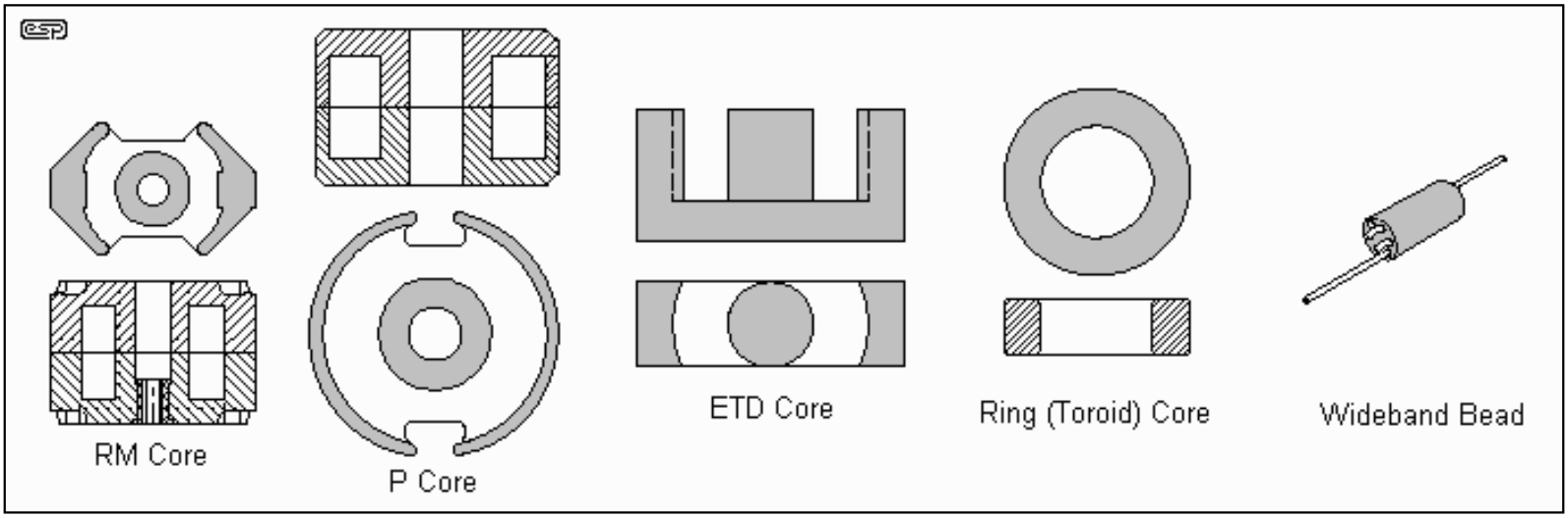


Figure 13.2 - Some Ferrite Core Styles

The diagram in Figure 13.3 shows the correct way to stack an E-I transformer. Sometimes manufacturers will use 2 or 3 laminations in the same direction, then the same in the other. This cuts costs, but the transformer performance will never be as good. Alternate laminations minimise the air gap created between the E and I sections due to imperfect mating of the two. It is essential that the laminations are packed as tightly as possible so that the effects of the air gaps are minimal.

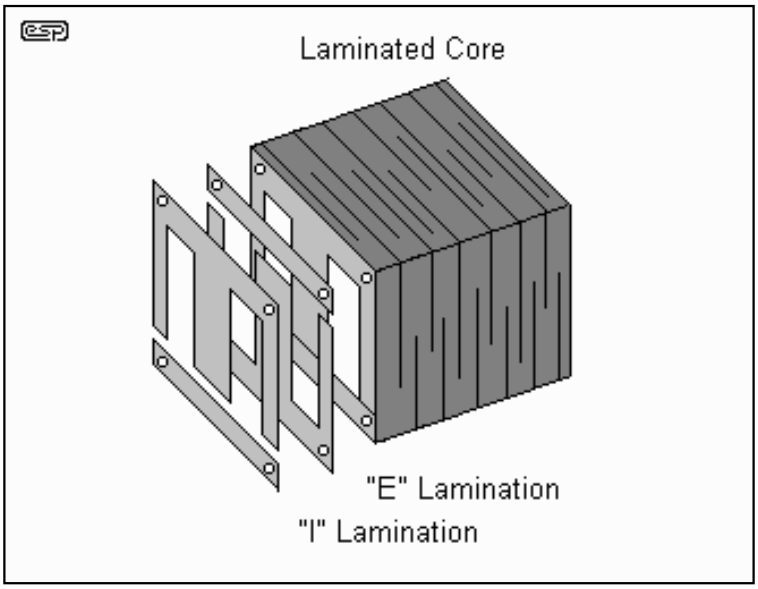


Figure 13.3 - E-I Lamination Stacking

When a transformer using E-I laminations is bolted together, it is important that the bolts are insulated from the core. If not, this would allow large eddy currents to circulate through the end laminations and the bolts, reducing performance dramatically. For safety, the core should always be bonded to mains earth unless the transformer is rated as "double insulated".

"Yes, but what good is that? The laminations are insulated from each other anyway." The inter-lamination insulation is sufficient to prevent eddy currents, but cannot withstand the mains voltage, so in case of electrical breakdown, the core may become "live" if not earthed.

In order to reduce the radiated flux from an E-I transformer core, you will sometimes see a copper or

brass band* wrapped around the winding and the *outside* of the core, as shown in Figure 13.4. This acts as a shorted turn *to the leakage flux only*, and greatly reduces magnetic interference to adjacent equipment. Such measures are not needed with toroidal transformers, as leakage flux is very much lower, and the core is completely enclosed by the windings.

(* While I am sure that many people would love to see their local brass band wrapped around a transformer, this is not what I had in mind. It does create an interesting mental picture though :-)

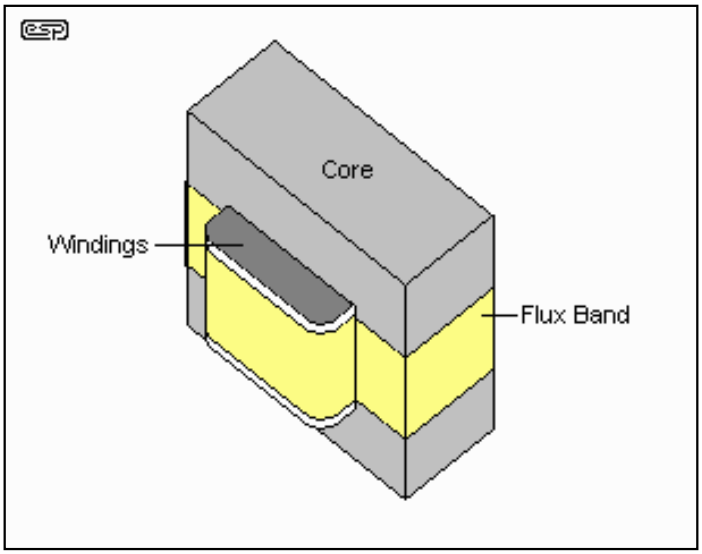


Figure 13.4 - Flux Banded Transformer

Just in case you were wondering, the dimensions of E-I laminations are worked out so that the laminations can be created with no material waste (other than the holes). The relative dimensions are shown below, and are just a ratio of the real dimensions, which will naturally be in millimetres or inches.

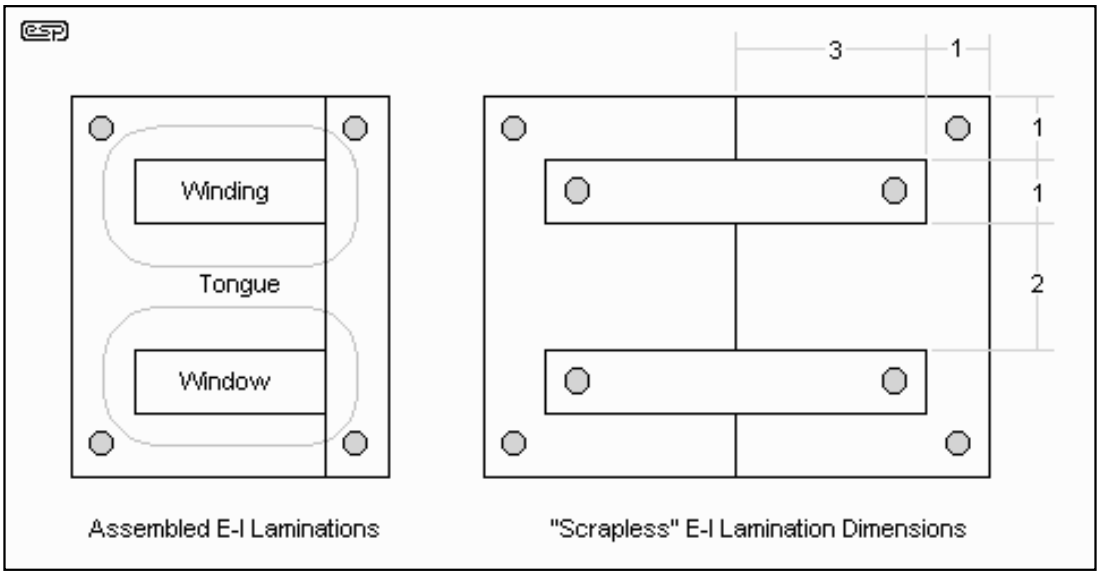


Figure 13.5 - Assembled Laminations and Punching Dimensions

The magnetic path length is the average for the dual path shown in the assembled lamination drawing, and is generally assumed to be 12 (units). This may be thought a little pessimistic, but is the commonly accepted figure. The winding window size is restricted by the punching dimensions, and it is critical that the maximum usage is made of the limited area available. Should the winding wire be too thin, there will

be plenty of room, but copper losses will be excessive. Make the winding wire too thick, and the completed winding will not fit into the available space. Additional space must be allowed for the winding bobbin, and for inter-winding insulation and the final insulation layer.

13.1 Air Gaps

DC flows in the windings for any transformer that is used for "flyback" switching supplies or SET power amplifiers, to name but two. The effect is that the DC creates a magnetomotive force that is unidirectional, and this reduces the maximum AC signal that can be carried before saturation in one direction. Indeed, the DC component may cause saturation by itself, so the transformer would be rendered useless as a means of passing the AC signal without *severe* degradation. Even the use of a half wave rectifier will introduce an effective DC component into the windings, and these should be avoided at any significant power level (i.e. more than a few milliamps).

To combat this, transformers that are subject to DC in the windings use an air gap in the core, so it is no longer a complete magnetic circuit, but is broken by the gap. This lowers the inductance, and means that a larger core must be used because of the reduced permeability of the core material due to the gap. An air gap also increases leakage inductance because of the flux "fringing" around the gap, and resistive (copper) losses are increased as well, because more turns will be needed.

It is beyond the scope of this article to cover this in great detail, but it does impose some severe restrictions on the design of transformers where DC is present. This is (IMO) one of the biggest disadvantages of the SET amplifier so popular with audiophiles, as it almost invariably leads to unacceptable compromises and equally unacceptable distortion (both harmonic and frequency).

In some designs, it is possible to eliminate the DC component by using a tertiary winding that carries ... DC. If the additional winding can be made to induce a flux that is equal and opposite that of the bias current, then the quiescent flux in the transformer can be reduced to zero (where it belongs). The disadvantage with this is that it requires an extra winding, and ***that*** takes up valuable winding space on the core. It is also a difficult technique to get right, and is not often seen these days. It was a popular technique in telecommunications equipment at one time, and meant that smaller transformers could be used for the same (or better) performance.

E-I transformers all have a minuscule "air gap" because of the way the laminations are assembled. With care, this can be almost be considered negligible, but it cannot be eliminated. C-cores will have their cut ends machined to minimise the effect, but again, it cannot be eliminated entirely. The toroidal core has no air gap at all, and is therefore more efficient (magnetically speaking) - they are utterly intolerant of DC in the windings.

---

14. Materials

There is an enormous range of core materials, even within the same basic class, so I will mention only a few of the most common. All materials have some basic requirements if they are to be used with AC (for transformers, rather than solenoids or relays, which can operate with DC). The core cannot be solid and electrically conductive, or excessive eddy current will flow, heating the core and causing very high

losses. Therefore, all cores use either thin metal laminations, each electrically insulated from the next, or powdered magnetic material in an insulating filler. The list below is far from exhaustive - there are a great many variations of alloys, and I have mentioned only a few of those that are in common use.

Silicon Steel (General Information)
Typically, soft (i.e. low remanence) magnetic steel will contain about 4% to 4.5% silicon, which lowers the remanence of the steel and reduces hysteresis losses. Normal mild steel, carbon steel or pure iron has quite a high remanence, and this is easily demonstrated by stroking a nail (or screwdriver) with a magnet. The nail will become magnetised, and will retain enough magnetism to enable it to pick up other nails. The addition of silicon reduces this effect, and it is very difficult to magnetise a transformer lamination strongly enough so it can pick things up.

This is not to say that the remanence is zero - far from it. When a transformer is turned off, there will often be residual magnetism in the core, and when next powered on, it is common for the transformer to make noise - both toroids and E-I transformers can sometimes make a "boing" noise when power is applied. That this phenomenon is intermittent is a combination of several factors ...

- What was the polarity and magnitude of the mains at switch off
- What is the polarity and magnitude of the mains at switch on
- To what extent has the core de-magnetised itself between events

The longer a transformer is left unpowered, the lower the remanent flux, and the less likelihood there is of an excessively high inrush current. If the mains is applied when at it's peak value, inrush current is at it's lowest. Conversely, if the mains is applied at the zero crossing point, inrush current will be maximum - this is exactly the reverse of what you would logically expect. The inrush current lasts for several cycles, and is made much worse with a rectifier and filter capacitor on the output. The capacitor is a short circuit when discharged, and large capacitors will take longer to charge.

Silicon steel and other metal (as opposed to ferrite) materials are normally annealed by heating and then cooling slowly after stamping and forming. This removes most of the internal mechanical stresses caused by the stamping or rolling operation(s) - these stresses reduce the magnetic properties of the material, sometimes very dramatically.

CRGO - Cold Rolled Grain Oriented Silicon Steel
Like many steels, this version is cold-rolled to obtain the required thickness and flatness needed for a transformer core. The magnetic "grain" of the steel is aligned in one direction, allowing a higher permeability than would otherwise be possible. This material is ideal for toroids and C-cores, since the grain can be aligned in the direction of magnetic flux (i.e. in a circular pattern around the core). It is less suited to E-I laminations, because the flux must travel *across* the "grain" at the ends of the lamination, reducing permeability.

CRNGO - Cold Rolled Non Grain Oriented Silicon Steel
Generally more suited to E-I laminations, this is essentially the same process as the CRGO, but the magnetic grain is left random, with no alignment of the magnetic domains. Although this reduces overall permeability, the effective permeability will be better with stamped laminations (as opposed to rolled, as with toroids and C-cores).

Powdered Iron
A soft ferrite ceramic material, used where there is significant DC in the winding. Powdered iron cores have relatively low permeability (about 90, maximum), and are designed for high frequency operation. These cores are most commonly used with no air-gap, and will not saturate easily. Typically used as filter chokes in switching power supplies, and as EMI (Electro-Magnetic Interference) filters - the toroid is the most common shape.

Ferrite
Soft ferrites are the mainstay of switching power supplies, and low level high speed transformers (such as might be used for network interface cards and small switching transformers. Ferrites are available with outstanding permeability, which allows small cores with very high power capability. Flyback (a type of switchmode operation) transformers in particular are usually gapped because of the DC component in the primary current.

High permeability ferrites are also very common in telecommunications and for other small audio frequency transformers where very high inductance and small size is required.

MuMetal
Named after the symbol for permeability, as one might expect, this material has an extraordinarily high permeability - typically in the order of 30,000. It is commonly used as magnetic shielding for cathode ray tubes in high quality oscilloscopes, screening cans for microphone transformers, and as laminations for low level transformers. The maximum flux density is quite low compared to other metallic materials. Apart from being relatively soft, if dropped, the magnetic properties may be adversely affected (MuMetal requires careful annealing to ensure that its magnetic properties are optimised).

---

15. Transformer Distortion

An ideal transformer has zero distortion, but there are zero ideal transformers. Therefore, it can be deduced that transformers do have distortion, but how much?

The answer depends entirely on how the transformer is used. When supplied from a voltage source of zero ohms impedance, the real life transformer has no distortion, but again, there is no such thing as zero ohms (actually, it *can* be done, but yields little real benefit).

Any transformer operating at low flux density, and with a low impedance source, will contribute very little distortion to the signal. As frequency decreases, and/ or operating level increases, the limits of saturation will eventually be reached in any transformer, and distortion will become a problem. This is not really an issue with mains power transformers, but is very important for valve output transformers, particularly at low frequencies.

The distortion characteristics of transformers used as valve output devices is a complex subject, and will not be covered here. Suffice to say that the normal methods of determining the turns per volt, based on the bare minimum lowest frequency response will give unacceptably high distortion levels at low frequencies.

Although interesting, the relationships are complex and will be of little real use to the majority of

hobbyists. If there is sufficient interest, I may cover this in a later article.

References

Countless different books and Web pages were researched during the compilation of this article, and although some were interesting, the majority were of minimal use. Of those who I actually remember (a daunting task in itself, considering the sheer amount of searching I had to do), I must "thank" the following Web pages (in alphabetical order) ...

- Amidon
- ATDL School (US Army)
- Jensen Transformers
- Mitchell Electronics Corporation
- Tomi Engdahl - (ePanorama.net)

In addition, I have used various other references, but notably (in order of usefulness) ...

- Radiotron Designer's Handbook - F Langford-Smith (4th Edition)
- Basic Electronics - Grob, Third Edition (McGraw-Hill)
- Miniwatt Technical Data - Philips Elcoma, 7th Edition, 1972 (Philips)
- Motorola Rectifier Applications Handbook (3rd Edition)

Download Silvio Klaic's neat little transformer calculator from his website

---

The following (edited) definitions are from Units of Measurement

Units of Measurement site copyright by Russ Rowlett and University of North Carolina at Chapel Hill.
Definitions used by permission.

Tesla (T) - flux density (or field intensity) for magnetic fields (also called the magnetic induction). The intensity of a magnetic field can be measured by placing a current-carrying conductor in the field. The magnetic field exerts a force on the conductor which depends on the amount of the current and the length of the conductor. One Tesla is defined as the field intensity generating one newton of force per ampere of current per meter of conductor. Equivalently, one Tesla represents a magnetic flux density of one Weber per square meter of area. A field of one Tesla is quite strong: the strongest fields available in laboratories are about 20 Teslas, and the Earth's magnetic flux density at its surface, is about 50 microteslas (μT). One Tesla equals 10,000 gauss. The Tesla, defined in 1958, is named after Nikola Tesla (1856-1943), whose work in electromagnetic induction led to the first practical generators and motors using alternating current (much to the annoyance of Edison, who claimed DC was "safer").

Weber (Wb) - magnetic flux. "Flux" is the rate (per unit of time) in which something crosses a surface perpendicular to the flow. In the case of a magnetic field, then the magnetic flux across a perpendicular surface is the product of the magnetic flux density, in Teslas, and the surface area, in square metres. If a varying magnetic field passes perpendicularly through a circular loop of conducting material (one turn), the variation in the field induces a electric potential in the loop. If the flux is changing at a uniform rate of one Weber per second, the induced potential is one volt. This means that numerically the flux in webers is equal to the potential, in volts, that would be created by collapsing the field uniformly to zero in one second. One Weber is the flux induced in this way by a current varying at the uniform rate of one ampere per second. The unit honours the German physicist Wilhelm Eduard Weber (1804-1891), one of the early researchers of magnetism.

Magnetic Terminology

This list is far from complete, but will be sufficient to either get you started or scare you away. I have included the symbols and units of only three of the entries below, since most are of no real interest.

Coercivity - is the field strength which must be applied to reduce (or *coerce*) the remanent flux to zero. Materials with high coercivity (e.g. those used for permanent magnets) are called *hard*. Materials with low coercivity (those used for transformers) are called *soft*. Coercivity is the "reverse" of remanence.

Effective Area - of a core is the cross sectional area of the centre limb for E-I laminations, or the total area for a toroid. Usually this corresponds to the physical dimensions of the core but because flux may not be distributed evenly the manufacturer may specify a value which reflects this.

Effective Length - of a core is the distance which the magnetic flux travels in making a complete circuit. Usually this corresponds closely to the average of the physical dimensions of the core, but because flux has a tendency to concentrate on the inside corners of the path the manufacturer may specify a value for the effective length.

Flux Density - (symbol; B, unit; Teslas (T)) is simply the total flux divided by the effective area of the magnetic circuit through which it flows.

Flux Linkage - in an ideal inductor the flux generated by one turn would be contained within all the other turns. Real coils come close to this ideal when the other dimensions of the coil are small compared with its diameter, or if a suitable core guides the flux through the windings.

Magnetomotive Force - MMF can be thought of as the magnetic equivalent of electromotive force. It is the product of the current flowing in a coil and the number of turns that make up the coil.

Magnetic Field Strength - (symbol: H, unit; ampere metres (A m^{-1})) when current flows in a conductor, it is always accompanied by a magnetic field. The strength, or intensity, of this field is proportional to the amount of current and inversely proportional to the distance from the conductor (hence the -1 superscript).

Magnetic Flux - (symbol: Φ; unit: Webers (Wb)) we refer to magnetism in terms of lines of force or flux, which is a measure of the total amount of magnetism.

Permeability - (symbol; μ, units: henrys per metre (Hm^{-1}) is defined as the ratio of flux density to field strength, and is determined by the type of material within the magnetic field - i.e. the core material itself. Most references to permeability are actually to "relative permeability", as the permeability of nearly all materials changes depending upon field strength (and in most cases - especially in ferrites - with temperature as well).

Remanence - (or remnance) is the flux density which remains in a magnetic material when the externally applied field is removed. Transformers require the lowest possible remanence, while permanent magnets need a high value of remanence. Remanence is the "reverse" of coercivity .

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 17 March 2001

The Audio Pages

Elliott Sound Products **Guide to Transformers Part 3**

Copyright © 2005 - Rod Elliott (ESP)
With Thanks to Martin Czech and Geoff Sevart for Additional Material.
Page Published 16 January 2006

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Transformer Analysis
  - 2.1 Transformer Analysis 1
  - 2.2 Transformer Analysis 2
- 3.0 Transformer Design
- 4.0 Taking Measurements
- 5.0 Conclusion
- 6.0 Downloads

1.0 - Introduction

In this section, we will look at taking measurements from existing transformers to assess their ability to be re-used, some basic calculations usable for transformer design (thanks to Geoff Sevart) and look more closely at power supply analysis (thanks to Martin Czech). In all, there are three different programs that you can download and use presented here. It is also highly recommended that you use a good simulator to verify the final design - as always, I suggest SIMetrix. Before simulating, you will need to know the parameters to build the equivalent circuit - see section 2 for more information on this

topic.

2.0 - Transformer Analysis

Transformer analysis largely involves determining if a given transformer will provide a level of performance that meets your design goals. Three completely different approaches are provided here - one is an executable program that provides a fairly accurate assessment of the transformer's performance based on a few measurements that you perform.

The other is a Microsoft Excel spreadsheet that does much the same thing. The main difference is that the spreadsheet does not have an on-line help facility, so all instructions are provided here (as well as in a 'readme' file included in the compress archive).

Finally, the third program can be used to reverse engineer a transformer, or to design one that meets your needs. Only the basics are covered, and you need to be well versed in the requirements for insulation (both inter-winding and intra-winding). You may also need to know the details of the steel used - assumptions are not always valid. Remember that in all cases, it is your responsibility to ensure that the design is safe (or remains safe if you modify an existing transformer).

There are some differences in the way the three analysis processes work, and it is up to you to decide which one is most suitable. None is intended for beginners or those with an aversion to taking measurements and experimenting - these are all serious tools to allow you to analyse or determine the suitability of any given transformer.

2.1 - Transformer Analysis 1

The transformer analysis program (xformer.exe) can be downloaded from the Downloads page or from the downloads section of this page, as a zipped archive. It has very extensive help facilities, and is complete with test circuits, a transformer equivalent circuit (with all explanations), and the complete sequence to evaluate an unknown transformer. It will provide the DC output at the specified load, and includes duty cycle calculations so that you can see what happens when the transformer is overloaded.

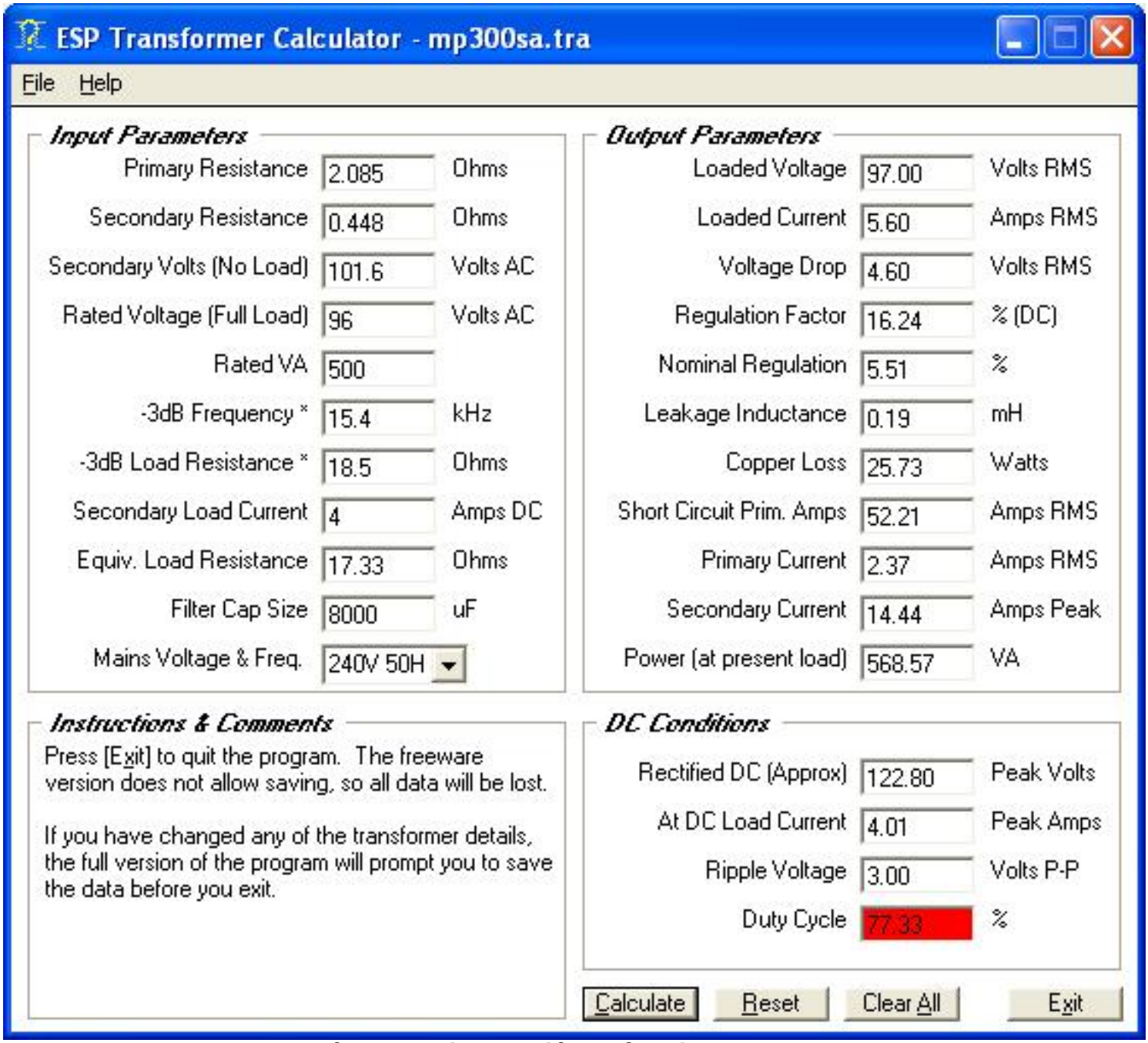


Figure 1 - Screen Shot of XFORMER.EXE

The box marked 'Instructions & Comments' is context sensitive, so it will display relevant information based on the current mouse position. To find out more about any parameter, simply position the mouse over the text box or its adjacent label. A great deal more information is available from the help screens.

As with any program, there will be some variation between what it claims and reality, so the measured values will not always be in complete agreement with those calculated. There is only so much that can be conveniently determined, but the results will be close enough for almost all applications. Where extreme accuracy is needed, you will have to build the circuit to verify exactly what it does, however the very nature of a power supply

is such that accuracy is not normally a major issue.

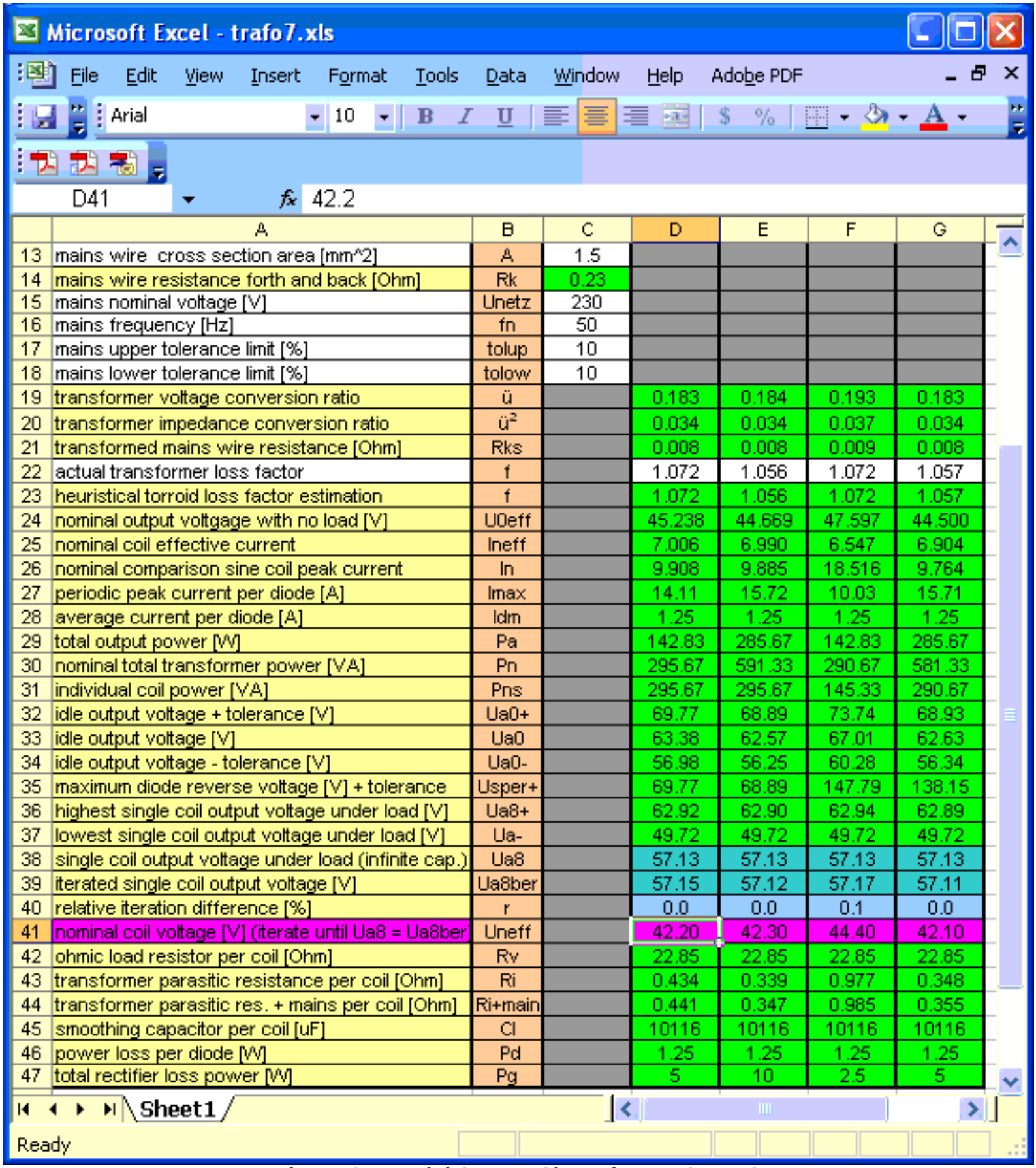
The program uses SI units only, and assumes a dual polarity full-wave bridge rectifier (conventional ± amplifier type power supply). At this stage, no other supply types are available, and probably will not be added because the default is suitable for most applications. In normal use, a single supply is still easily tested. You will need to experiment with the program to find out if it does exactly what you need.

2.2 - Transformer Analysis 2 (Martin Czech) - How to use the Excel spreadsheet

2.2.1 - Purpose

This Excel spreadsheet is made for convenient "linear PSU" design. This is a PSU made using a transformer, rectifier diodes and filtering capacitors. The load presented to the PSU is a resistor. The sheet was tested against SPICE simulations and has shown accuracy to about 2% or better.

However, since not all possible configurations can be checked it should always be verified by a subsequent SPICE simulation.


Microsoft Excel - trafo7.xls

File Edit View Insert Format Tools Data Window Help Adobe PDF

Arial 10 B I U

D41 fx 42.2

	A	B	C	D	E	F	G
13	mains wire cross section area [mm^2]	A	1.5				
14	mains wire resistance forth and back [Ohm]	Rk	0.23				
15	mains nominal voltage [V]	Unetz	230				
16	mains frequency [Hz]	fn	50				
17	mains upper tolerance limit [%]	tolup	10				
18	mains lower tolerance limit [%]	tolow	10				
19	transformer voltage conversion ratio	ü		0.183	0.184	0.193	0.183
20	transformer impedance conversion ratio	ü²		0.034	0.034	0.037	0.034
21	transformed mains wire resistance [Ohm]	Rks		0.008	0.008	0.009	0.008
22	actual transformer loss factor	f		1.072	1.056	1.072	1.057
23	heuristical torroid loss factor estimation	f		1.072	1.056	1.072	1.057
24	nominal output voltgage with no load [V]	U0eff		45.238	44.669	47.597	44.500
25	nominal coil effective current	Ineff		7.006	6.990	6.547	6.904
26	nominal comparison sine coil peak current	In		9.908	9.885	18.516	9.764
27	periodic peak current per diode [A]	Imax		14.11	15.72	10.03	15.71
28	average current per diode [A]	Idm		1.25	1.25	1.25	1.25
29	total output power [W]	Pa		142.83	285.67	142.83	285.67
30	nominal total transformer power [VA]	Pn		295.67	591.33	290.67	581.33
31	individual coil power [VA]	Pns		295.67	295.67	145.33	290.67
32	idle output voltage + tolerance [V]	Ua0+		69.77	68.89	73.74	68.93
33	idle output voltage [V]	Ua0		63.38	62.57	67.01	62.63
34	idle output voltage - tolerance [V]	Ua0-		56.98	56.25	60.28	56.34
35	maximum diode reverse voltage [V] + tolerance	Usper+		69.77	68.89	147.79	138.15
36	highest single coil output voltage under load [V]	Ua8+		62.92	62.90	62.94	62.89
37	lowest single coil output voltage under load [V]	Ua-		49.72	49.72	49.72	49.72
38	single coil output voltage under load (infinite cap.)	Ua8		57.13	57.13	57.13	57.13
39	iterated single coil output voltage [V]	Ua8ber		57.15	57.12	57.17	57.11
40	relative iteration difference [%]	r		0.0	0.0	0.1	0.0
41	nominal coil voltage [V] (iterate until Ua8 = Ua8ber	Uneff		42.20	42.30	44.40	42.10
42	ohmic load resistor per coil [Ohm]	Rv		22.85	22.85	22.85	22.85
43	transformer parasitic resistance per coil [Ohm]	Ri		0.434	0.339	0.977	0.348
44	transformer parasitic res. + mains per coil [Ohm]	Ri+main		0.441	0.347	0.985	0.355
45	smoothing capacitor per coil [uF]	Cl		10116	10116	10116	10116
46	power loss per diode [W]	Pd		1.25	1.25	1.25	1.25
47	total rectifier loss power [W]	Pg		5	10	2.5	5

Sheet1

Ready

Figure 2 - Partial Screen Shot of TRAFO7.XLS

2.2.2 - Assumptions

This Excel spreadsheet tries to model the nonlinear behaviour of such a PSU with simple equations and estimations. The transformer is modelled as a non ideal voltage source, i. e. an ideal voltage source plus series resistance. It uses only SI units (m,A,V, Ohm, etc.).

Four common circuits can be calculated ...

- Bridge - single coil transformer with diode bridge
- d. Bridge - double coil transformer with diode bridge for each coil, dual output voltage, very heavy load (1000W)
- Center - centre tap two coil circuit with two diodes for single output voltage, only useful for very low power and low voltage, less diode drop
- d. Center - centre tap two coil circuit with four diodes for dual output, for low to high output power until the highest, less diode losses than "d. Bridge"

Usually a given transformer is specified with Uneff (effective value of secondary voltage under full resistive load, and Pn (nominal output power with said resistive load) and loss factor f (ratio between U0eff and Uneff, U0eff being the effective value of output voltage with no load), and of course the primary effective voltage Unetz.

I.e.: $U0eff = f * Uneff$

Out of *Pn* and *Uneff* we get the load current *In* (*In = Pn / Un*). The series resistance is therefore

$$(U0eff - Uneff) / Ineff = (f * Uneff - Uneff) / Ineff$$
$$= Uneff * (f - 1)/In$$
$$= Uneff * 2 * (f - 1) / Pn$$

the latter is convenient because it uses only the usually given parameters. This applies also to transformers with several identical coils. It does not apply for additional auxiliary windings and taps, which can have lower power spec and therefore more series resistance than perhaps expected.

The sheet works for one coil or two identical coils, due to the symmetry of the circuit and load situation the same current and voltage values will appear. Therefore they are only given once in the sheet.

Sadly, *f* (loss factor) is not given by many vendors. So the spreadsheet has some heuristic formula to determine *f* out of the transformer total power rating. This works from about 5 VA up to 1000 VA. This is no big science, but relates to the usual tradeoffs when

winding transformer coils. The value is valid for toroidal types, *f* is often lower for EI types. However, if you can obtain a data sheet with *f*, then do not use the heuristic value.

The spreadsheet does also include mains voltage variation. In central Europe this is tending to +-10% of the nominal value, guaranteed at the house main power entry. Some additional wiring can be added to that (be it house installation or additional cables, perhaps during live shows). The transformed total mains resistance will simply add to the transformer series resistance.

The remaining interesting part of the formulas is derived by approximating the voltage sine waves with parabola, in order to get any analytic result. Unfortunately the resulting nonlinear equation can only be solved by program, I used an iteration approach in this Excel sheet.

The load is assumed to be a resistor. In reality this not always the case. Some circuits tend to be constant current drains (regulator plus electronic circuit). Others have variable current consumption (power amplifiers), but also there the current does not depend on PSU voltage. A starting point for the later cases is to make the delivered output power equal, i.e. choosing the output current and resistor in that way.

2.2.3 - How to do it

The Excel sheet has a yellow title range. Below that on the left hand side we find explanations to the parameters, the salmon colored column shows the parameter name. The next column is commonly used for all four possible circuits, the next four columns carry the results and entries for each circuit individually.

Anything shown below in *italics* indicates that it is a field in the spreadsheet.

First you have to enter data into the white common fields. Each field must be filled. Most of this is very easy. You enter ...

- The Si diode forward voltage drop [*Ud*] (recommended 1V)
- Specify peak-to-peak ripple voltage under load [*Ubrss*]
- The desired nominal output current [*Ia*] (together with output voltage this will define the delivered power)
- The current form factor [*alpha*]. The transformer has to deliver current peaks instead sine AC current, the copper losses will be higher than nominal.

This makes it necessary to use a higher transformer nominal power. Under-dimensioning can lead to ...

1. More than nominal copper loss
2. Enormous heat development in coil wires

Usually alpha is chosen to be 1.5 ... 2, even large over dimensioning is possible with modern toroid types, because idle losses are still low.

Imax >> In means that in this PSU the peak current is much larger than in the situation where the transformer drives only a resistive load to the rated power. One should compute the RMS value from simulation to estimate the loss. Based on that choose alpha to be higher.

- Your desired nominal DC output voltage [*Uaminn*] (just to remember)
- Your desired minimum nominal output voltage [*Uamin*] this has to be higher than *Uaminn* because of wiring losses and mains power variation, if the mains variation to lower is perhaps 10%, give *Uamin* 10% more, as a starting point
- The assumed wiring length from mains house entry to your transformer [*L*], in a normal house this can be 10 - 20 m, in a big hall this can be 100m
- The mains wiring cross section area [*A*], typical 1.5 mm²
- The mains nominal effective voltage [*Unetz*] (e.g. 110V or 230V)
- The mains AC frequency [*fn*] (e.g. 50Hz or 60Hz)
- The mains tolerance upper value [*tolup*] (e.g. 10% by law in Germany, meaning 230V and up to 253V)
- The mains tolerance lower value [*toloe*] (e.g. -10% by law in Germany, meaning 230V and down to 207V)

You are done with the common data. Now you can decide which circuit column you need to use. It is only necessary to fill in that column, but it is also possible to use several or all in parallel for study or comparison.

Now the Excel sheet knows the output power [*Pa*], nominal power per coil [*Pns*], and the total nominal transformer power [*Pn*]. The heuristic formula can therefore suggest a value for the toroid loss factor [*f*]. If you have no better data, enter this value in the white field "actual transformer loss factor". Now you have completed all white fields, you are nearly done.

In the last step you solve the nonlinear equation by manual iteration. You enter a nominal coil voltage [*Uneff*] (pink field) and change it, until the relative iteration difference [*r*] (light blue) is 0.0% (or near to that). In this case the single coil output voltage under load with infinitely large capacitor [*Ua8*] will match the iterated single coil output voltage [*Uaber*]

(darker blue).

Now you have solved the equation, but is the worst case voltage [*Ua-*] as expected? You should compare this lowest per coil output voltage under load [*Ua-*] with your intended desired minimum nominal output voltage [*Uaminn*]. If [*Ua-*] is too low, the circuit will have too low voltage in case of low mains voltage, a longer than expected cable can add to that.

This makes perhaps no difference for a power amp, since a few % amplitude loss is not audible. In the case of a subsequent voltage regulator a too low voltage will simply mean loss of regulation with perhaps additional artefacts, this is clearly not acceptable. Give more [*Uamin*] in such a case, and iterate again.

Take care: this can influence the power, so the current [*Ia*] may need some adjustment, too. Finally power change can change loss factor [*f*], so this has also to be adapted. If the initial guess was good enough, little or no modification needs to be made on that side.

So now you are done. You should take the relevant data from the sheet (*U0, Un, Pn, f*) and do a SPICE simulation to verify. A second computation is always good, since you are going to spend a lot of money for a PSU!

2.2.4 - What else does the sheet tell you

It not only tells you *Uneff, Pn, f, Unetz* to order your transformer, it also tells you the size of the filter capacitors *Cl*, and *Ua0+*, this idle output voltage plus tolerance tells you about the voltage stress these and other devices connected to your PSU have to take.

Imax is the periodic peak current in the transformer windings and diodes, so it will tell you something about fuse stress and diode stress, as well as capacitor ripple current.

The total rectifier loss power *Pg* and the maximum diode reverse voltage *Usper+* will also help to buy the right rectifier.

All the devices need not only cope with the rated voltage stress, but even more has to be considered because of possible mains transients. A factor of 2 over dimensioning is not too expensive in most cases, an exploded cap or whole PSU plus circuit certainly is.

If your output voltage is so high that voltage over dimensioning is not possible, primary side varistors should be used to handle transients. They should be used anyway.

2.2.5 - Explanation of Parameters

The parameters used in the spreadsheet are listed here. Although some of these may appear daunting at first, it is not difficult to figure out what they are all for. Note that the derivation of some of the terms is directly from German (netz = mains, for example). Many of the parameters used are fixed - there is no need to change them once they have been set for local conditions - examples are mains voltage and frequency, upper and lower tolerances, etc. The line length applies only if there is a significantly long extension lead used, such as setup in an auditorium where a nearby power outlet is not available. For most applications this may be set to 1 (metre) with little loss of accuracy.

3.0 - Transformer Design

This next program is thanks to 'Particle' otherwise known as Geoff Sevart. With this, you can either analyse or design a transformer, although for proper design work you need to know a bit more than the program provides from its interface. Although fairly simple, it does give good results if you know what you are trying to achieve, and know some of the basics of the core material.

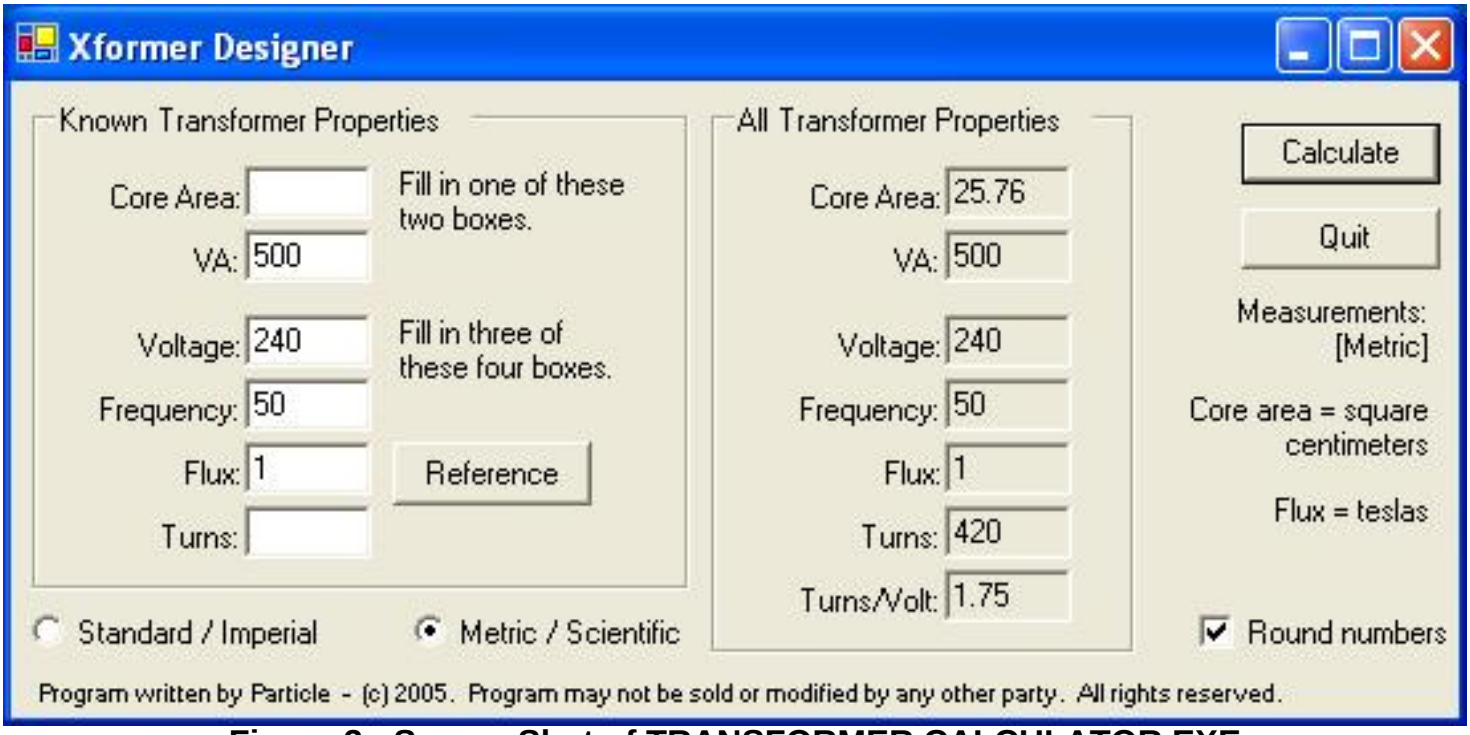


Figure 3 - Screen Shot of TRANSFORMER CALCULATOR.EXE

The screen shot above shows the complete program interface. You insert your known data into the section on the left side, and all calculated values are shown on the right. You need to input either the core area or the VA rating of the transformer, but not both. When that is done, you enter any *three* of the four remaining fields ...

- Voltage - the primary voltage for the transformer. This will nearly always be known.
- Frequency - mains frequency for a power transformer, or the lowest frequency of interest for a signal or output transformer
- Flux - the maximum flux density that the transformer core will be subjected to. *
- Turns - the number of turns on the primary. This is the least likely to be known, but it can be determined (see below)

* Click the 'Reference' button for hints. Typical power transformers will operate at around 1 Tesla. Higher flux density means the core is likely to saturate and be noisy, lower flux makes a mechanically quiet transformer, but at the expense of efficiency. Low flux density is essential for audio transformers, as even approaching saturation can cause high distortion levels.

If your transformer allows you to add some temporary turns, you can determine the number of turns/volt by adding 10 turns of thin insulated wire. Apply power to the primary, and measure the mains voltage (very carefully!). Now, measure the voltage across your extra 10 turn winding. If (for example) you measure 3.3V AC, that means the transformer has ...

TPV = *Turns / Voltage*= 10 / 3.3 = 3 Turns Per Volt

From that, you can calculate the number of turns on the primary, using the measured applied mains voltage (assume 230V) ...

Tp = *Vp* * *TPV*= 230 * 3 = 690 turns

This figure may be used instead of the assumed flux density, and you will then know the actual flux density being used for the transformer you are testing.

When you click 'Calculate', the program shows the following calculated values ...

- Core area, in cm² or inches², depending on whether you selected metric or imperial
- VA rating for the transformer
- Primary voltage
- Frequency
- Flux density
- Primary turns
- Turns per volt

Although it may be more difficult using metric calculations as a matter of course, it is

highly recommended that you do so anyway. There is nothing in the imperial system that actually makes any sense, so it is far better to use metric whenever possible.

4.0 - Taking Measurements

The equivalent circuit shown in Figure 5 allows you to see where the main losses occur with any transformer. The lumped parameter model is the most commonly used, as it gives a very good representation of a real component and is easy to manage for almost any normal load or signal.

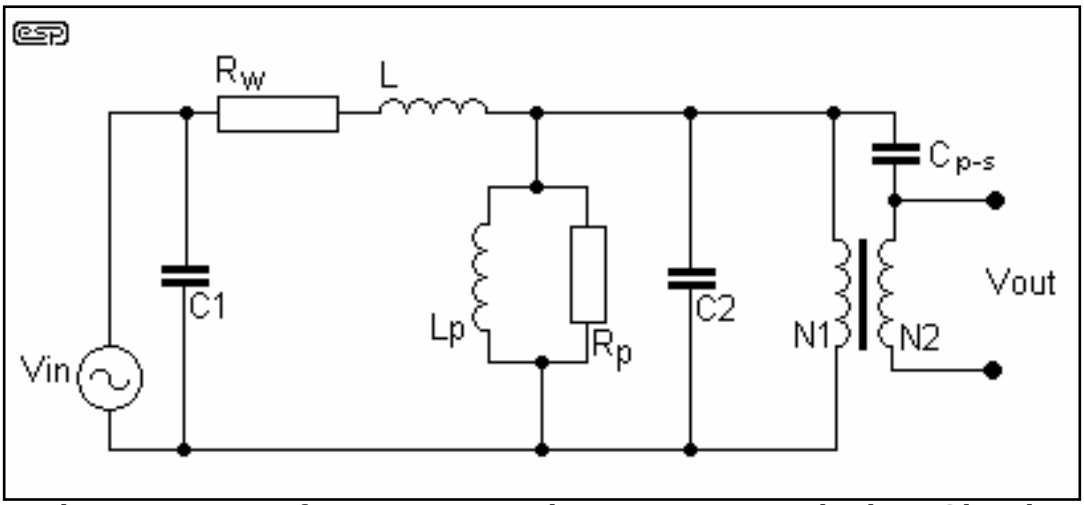


Figure 5 - Transformer 'Lumped Parameter' Equivalent Circuit

For most measurements in electronics, a multimeter is all you need. This is not good enough and will become inaccurate when working with transformer winding resistances. Because very low resistances are to be measured, few multimeters have a useful low ohms range. To get sensible results, the measurements you take *must* be accurate. Very low resistance is always hard to measure, and it can only be done using DC. Many very low ohm meters use AC, but this will give large errors because of the transformer inductance.

The easiest way to measure a very low resistance is to inject a known current, and measure the voltage across the device under test. For example, if you subject a transformer winding to (exactly) 1A DC, and measure 448mV across the winding, its resistance is 0.448Ω. A regulated DC supply and a 10 ohm 5W resistor is ideal for this. Measure its value carefully - if it measures 10.1 ohms, then 10.1V dropped across the resistor means the current is exactly 1A. This is shown in Figure 4.

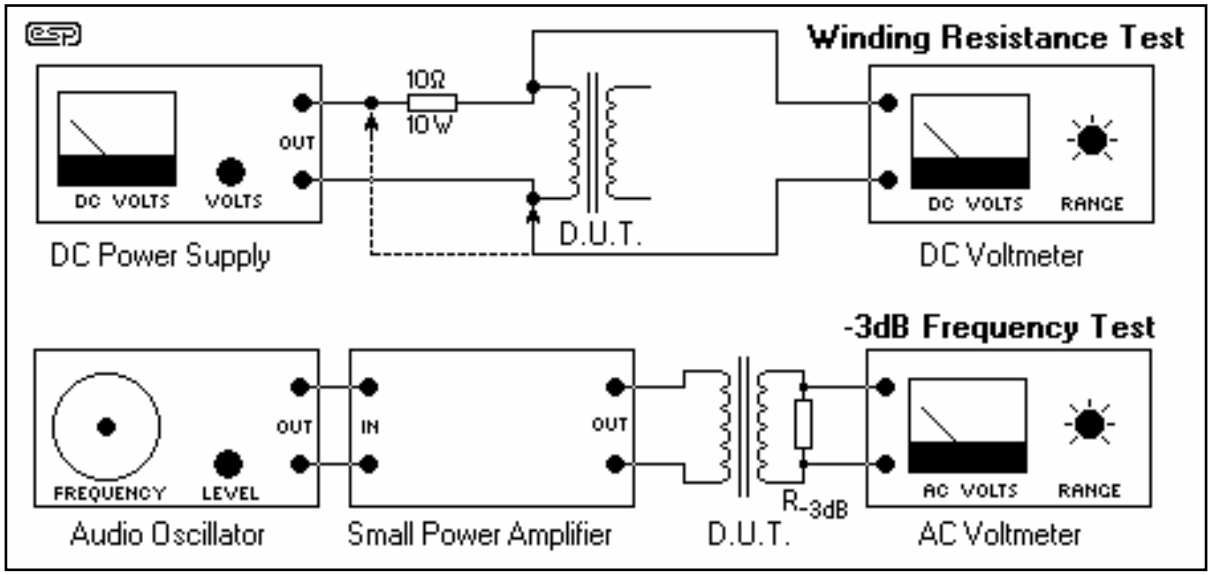


Figure 4 - Measuring Very Low Resistance & Leakage Inductance

Figure 4 also shows the basic method for measurement of leakage inductance. The load resistance across the secondary is determined by ...

$R_{-3dB} = VS / (VA / VS)$ Where R_{-3dB} = Test resistance, *VS* = Secondary Voltage, *VA* = Volt-Amp Rating

So for a 300VA transformer having a rated total secondary voltage of 70V ...

R_{-3dB}= 70 / (300 / 70) = 70 / 4.3 = 16Ω

Note that this does not have to be extremely accurate, as it is only designed to provide a representative load. This value is only used in the first program shown (TRANSFORMER.EXE), and it is used primarily to determine the transformer's short circuit (fault) current. Leakage reactance can then be determined using the -3dB frequency and the load impedance. Leakage inductance is also useful to know as a comparative figure, allowing you to make meaningful comparisons between different transformers. Transformers with high leakage inductance will inject noise into adjacent circuits, and generally have poorer regulation when high peak currents are expected.

5.0 - Conclusion

With the programs and spreadsheet detailed here, you have an excellent range of tools at your disposal to ensure that the power supply for your latest masterpiece is as good as you can get it. At last, it is possible to easily analyse an existing transformer, work out just how well (or badly) a given transformer will work in the intended circuit, or even

design your own transformer from scratch.

The performance of a transformer when loaded by a diode bridge and capacitor input filter (99% of all power supplies used) is never as good as we might hope or imagine, and these tools will allow you to predict how the supply will behave. By comparing the performance of the final supply with the predictions, you will find that there is fairly good correlation - certainly well within the accepted mains tolerance.

6.0 - Downloads

These downloads are free for personal use. They may be re-distributed, but no fee is to be charged for the software. No part of the software may be de-compiled, reverse engineered or used in any way contrary to the general principles of free software distribution. Program copyright belongs to the author of the program. No warranty is expressed or implied - the programs and spreadsheet are provided 'as is', and no responsibility is accepted by the authors for any damages howsoever caused. It is the wholly the users responsibility to determine the validity of the calculations.

All programs are tested and checked and are believed to be free of any computer virus or other 'malware'. It is the user's responsibility to scan all files to ensure that they have not been corrupted or infected.

Various support files needed may or may not be available on your PC, and if not, you will need to obtain them so the programs will run. Full details of the necessary files are shown below for each program.

TRANSFORMER.EXE © Rod Elliott
The executable file is 188,416 bytes. It is in a compressed archive, along with a 'readme' file that explains the support files and their required locations within the Windows file system. It requires the Visual Basic 6.0 runtime library vbrun60sp5.exe which is available from Microsoft, and contains all files needed to run any application written in Visual Basic 6.0

Click TRANSFORMER1.ZIP (or right click, and select 'Save [Link] Target As ...') to download the zipped archive for the program.

TRAFO7.XLS © Martin Czech

This is the Excel spreadsheet analyser (25,600 bytes). You will need Microsoft Excel™ so it will run. Alternatively, you may use OpenOffice, a free version that appears to run the spreadsheet perfectly.

Click TRANSFORMER2.ZIP (or right click, and select 'Save [Link] Target As ...') to download the zipped archive for the spreadsheet. Make sure that you read the file 'readme.txt' - this is a plain text file, and contains a full description of all fields used and their meanings.

XFORMER.EXE © Geoff Sevart
The transformer design program (61,440 bytes) requires the Microsoft '.NET' framework. Click here to find out more about the .NET system as a whole. This is normally supplied with Windows® XP, and more information (including download) is available from the Microsoft .NET Framework Developer Center.

Click TRANSFORMER3.ZIP (or right click, and select 'Save [Link] Target As ...') to download the zipped archive for the program.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott. Parts of this article are also Copyright © Martin Czech and Geoff Sevart. Downloadable program files are copyright as indicated above. These are supplied as freeware, and no fee is to be charged for their distribution.

Page created and copyright © 24 Oct 2005./ Published 16 Jan 2006.

The Audio Pages

Elliott Sound Products **Voltage Dividers & Attenuators**

Copyright © 2002 - Rod Elliott (ESP)
Page Created 18 Dec 2002

Articles Index

Main Index

Contents

- Contents
- Introduction
- Description
- Conclusion
- Copyright

Introduction

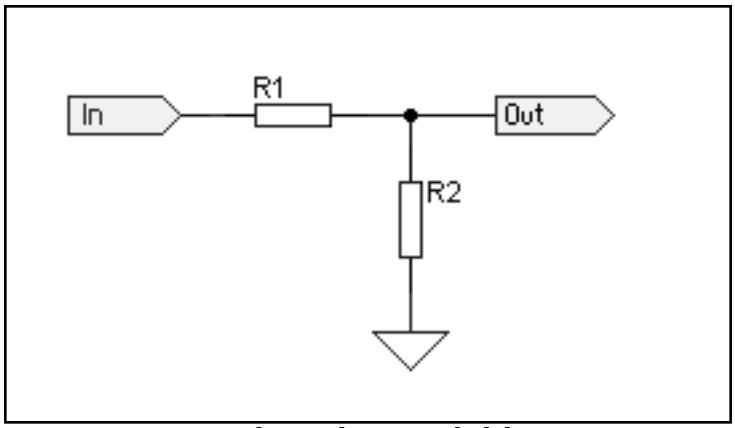
Based on the number of requests for help I receive from people wanting to know how to connect a volume control, or convert speaker level to line level, I must conclude that voltage dividers (or attenuators) are not well understood.

A volume control is, in most cases, nothing more than a variable attenuator. Exactly the same formula applies to determine the output level for any given input level, and there is nothing mysterious about any of these building blocks. This small article will de-mystify the voltage divider in any of its forms, and will be useful for the beginner and accomplished hobbyist alike.

Although simple, there are a great many uses for the humble voltage divider, and indeed, without it many of the circuits we take for granted would not exist.

Decription

A voltage divider is created whenever you have two resistors (or impedances) in series, with the signal "take-off" point between the two. Although there are literally hundreds of different possibilities, I shall only look at the standard connection for a voltage divider, and this will be all that is needed in the vast majority of cases. A traditional voltage divider is shown in Figure 1, and this is the form taken by volume and balance controls, general purpose attenuators, and similar configurations.



Basic Voltage Divider

This circuit is used for both AC and DC, and performs identically in either case. The voltage division is given by the formula ...

$$V_d = 1 + (R1 / R2)$$

Where V_d is the voltage division ratio. So, using two 1k resistors for example, voltage division is 1 + (1 / 1) = 2 [1]. 1V input will result in 0.5V output, and this holds true for DC, AC (RMS), or AC peak (as measured on an oscilloscope).

Any combination of R1 and R2 will create a voltage divider, and a pot (used for volume, for example) will still obey the same rule, except that the wiper (the moving contact in the pot) allows an infinite number of voltage division ratios.

Let's assume that you have a 25V RMS signal (the speaker output of an amp, for example), and want to reduce that to 1V RMS at maximum power from the amp. The voltage divider obviously must divide the voltage from 25V to 1V - or a 25:1 ratio. If we make R2 1k as before, R2 must be (25 - 1) * 1k = 24k [2].

Hold on a minute - how did I arrive at that? The formula (in the form I have used here) can be transposed easily - it follows from the original that ...

$$V_d - 1 = R1 / R2$$

allowing you to easily determine the values needed for any voltage division ratio you need.

To convert voltage division to dB (attenuators are commonly referred to in dB), you need to apply the dB formula ...

$$dB = 20 * Log_{10} (V_{d)}$$

... where V_d is the voltage division ratio determined as above. The two voltage dividers we used as examples above will give ...

$$dB = 20 * Log_{10} (2) = 6dB$$ [1] and ...

$$dB = 20 * Log_{10} (25) = 28dB$$ [2]

... respectively (close enough).

The only other point to consider is the loading on the previous circuit, and the power dissipation in the voltage divider resistors. A voltage divider used to convert speaker level to line level (as shown in the second example) could just as easily use 1 Ohm or 1 Megohm for R2 (instead of 1k). The voltage divider / attenuator will still work exactly as before, so why would 1k be "better" than any other value? The answer is not especially simple, and it comes down to a compromise (all too common in electronics).

Let's look at the case of R2 = 1 ohm first. R1 will be 24 ohms, and there will be 24V across it (you *must* understand this concept - draw it out on paper to make sure that you do!). Power dissipation will be 24^2 / 24 = 24W. This is power that the amp must supply, and the resistor will get hot. Using 1 ohm for R2 is obviously not a good idea. Ok, how about 1M?

R2 will now be 24 Megohms - not an easy value to find! We will also come up against another issue - output impedance.

It is essential that any attenuator or voltage divider is driven from a low impedance source, or the load of the divider itself will reduce the available voltage (the formula

shown will appear to be in error). Likewise, the load (the impedance connected to the output) must be high compared to the divider output impedance. It is generally considered that the signal source should have an impedance of at most 1/10 that of the attenuator, and the load should have an impedance (at least) 10 times the attenuator's output impedance.

The output impedance is the parallel combination of R1 and R2, so again using the examples above, we can determine input and output impedances.

Z_{in} = R1 + R2 = 1k + 1k = 2k	[1]
Z_{in} = R1 + R2 = 24k + 1k = 25k	[2]
Z_{out} = (R1 * R2) / (R1 + R2) = (1k * 1k) / (1k + 1k) = 500 ohms	[1]
Z_{out} = (R1 * R2) / (R1 + R2) = (24k * 1k) / (24k + 1k) = 960 ohms	[2]

These figures tell us that the maximum source impedance should be 2k / 10 = 200 ohms [1] and 25k / 10 = 2.5k [2], and the minimum load impedances should be 500 * 10 = 5k [1] and 960 Ohms * 10 = 9.6k [2]. As it turns out, these are easily achieved by all common circuits used in audio.

It must be understood that even with a "safety factor" of 10 as described, there will still be an error when the voltage divider is driven from any source impedance above zero, or is loaded by any circuit whose impedance is less than infinite - i.e. all voltage dividers will be in error to some (usually insignificant) degree, unless the source and load impedances are included in the calculation. The load impedance is effectively in *parallel* with R2, and the source impedance is in *series* with R1.

Returning to the R1 = 1M, R2 = 24M impedances, it is obvious that the divider will be very easy to drive from any common circuit (due to the minimal loading), but the output impedance is much too high. The final voltage divider will be loaded excessively by any load impedance less than around 9.6 Megohms, and because of the high impedance, high frequency losses will be excessive if a cable is used at the output of the attenuator (the capacitance of the cable will be sufficient to shunt HF signals to earth, instead of allowing them to reach the source). Even stray capacitance in the attenuator itself will have an effect!

It is outside the scope of this little article to cover capacitive voltage dividers (or Resistor Capacitor dividers), but they are commonly used in high impedance circuits. Project 16

(Audio Millivoltmeter) does show a perfect example of this technique, which eliminates stray capacitance effects (at the expense of higher than normal input capacitance).

In fact, a voltage divider can be made using capacitors or inductors - but only for AC. These are much less common than resistive dividers, but still work in much the same way (they are a lot harder to design though).

Conclusion

As is obvious from the above, the humble attenuator or voltage divider is not so humble after all. The maths are simple, and it is easy to convert any high voltage to a lower voltage. The divider technique is not suitable for *power* circuits however. It is used to reduce voltages, *not* current or power (although both are affected, that is a side effect, and not the real purpose).

Make sure that the resistance values you use are "sensible", and do not impose excessive loading or introduce excessive output impedances, and it is hard to go wrong. Sensible (in this context) is something that comes with experience, but the guidelines given here should be more than enough to get you under way.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 18 Dec 2002

Elliott Sound Products — Power Amplifier Design Guidelines

Rod Elliott - Copyright ©1999-2006
Page Last Updated 16 Jan 2006

Main Index

Articles Index

Contents

- Introduction
- Input Stages
  - RF Interference Suppression
- Gain Stages (Class-A Amplifier)
  - Active Current Source or Bootstrap?
- Output Stages
  - Thermal Stability
  - Output Stage Linearity
  - Bias Servos
  - Output Stage Stability
  - Output Current
- Some Notes on Power Supply Design
- Measurements Vs. Subjectivity
  - Valves Vs. Transistors Vs. MOSFETs
  - Slew Rate and Intermodulation
    - Slew Rate Nomograph
  - Frequency Response Etc.
- Further Reading
- References

Introduction

I am amazed at the number of amplifier designers who have, for one reason or another, failed to take some of the well known basics and pitfalls of amp design into consideration during the design phase. While some of these errors (whether of judgement or through ignorance is uncertain) are of no great

consequence, others can lead to the slow but sure or instantaneous destruction of an amplifier's output devices.

When I say "of no great consequence", this is possibly contentious, since a dramatic increase in distortion is hardly that, however in this context it will at least not destroy anything - other than the listener's enjoyment.

Even well known and respected designs can fall foul of some basic errors - this is naturally ignoring the multitude of "off the wall" designs (e.g. Single-ended MOSFETs without feedback (yecch! - 5% distortion, phtooey), transformer-coupled monstrosities, amplifiers so complex and bizarre that they defy logic or description, etc). This is not including valve amps, these are a "special" case and in many areas, such as guitar amps, they remain unsurpassed.

In this article, I have attempted to cover some of the areas which require their own special consideration, and the references quoted at the end are excellent sources of more detailed information on the items where a reference is given.

Reference Amplifier

My reference amplifier is shown in Project 3A, and is a hard act to follow. As I have been refining these pages and experimenting with simulations and real life, I have found that this amp is exemplary. It does need a comparatively high quiescent current to keep the output devices well away from crossover distortion, but this is easily accommodated by using decent heatsinks. Even a Class-A system (Death of Zen) fails to come close at medium power, and is barely better at low power.

This amp uses the following...

- long tailed pair input stage
- single stage bootstrapped Class-A driver
- complementary compound pair output stage
- RC Zobel network (it hates inductors)
- no current mirrors or sources (other than the bootstrap)

It is stable with all conventional loads, capable of 80W into 8 Ohms, and simple to build. Using only commonly available parts, it is also very inexpensive.

Note:

This article is not intended to be the "designers' handbook", but is a collection of notes and ideas showing the influences of the various stages in a typical amplifier. Although I have made suggestions that various topologies are superior to others, this does not mean to imply that they should automatically be used. If one were to combine all the "best" configurations into a single amp, this is no guarantee that it will perform or sound any better than one using "lesser" building blocks.

There is a school of thought that the fewer active devices one uses, the better an amp will sound. I do not believe this to be the case, but my own design philosophy is to make any given design as simple as possible, consistent with the level of performance expected of it.

Additional schools of thought will make all manner of claims regarding esoteric components, 'unexplained' phenomena, or will imply that most amplifiers as we know them are useless for audio because they do not have predictable performance at DC and/or 10GHz, cannot drive pure inductance or capacitance, etc., etc. Regardless of these claims, most amplifiers actually work just fine, and do not have to do any of the things that the claimants may imply. The vast majority of all the off-the-wall claims you will come across can safely be ignored.

Input Stages

There are two main possibilities for an input stage for a power amplifier. The most common is the long tailed pair, so we shall look at this first.

Long Tailed Pair

It has been shown [1] that failing to balance the input Long Tailed Pair properly leads to a large increase in the distortion contributed by the stage. Some designers attempt to remedy the situation by including a resistor in the "unused" collector circuit, but this is an aesthetic solution - i.e. it looks balanced, but serves no other useful purpose. (See Figure 1a) Note that the "driver" transistor is simply there to allow us to make comparisons between the circuit topologies, and to provide current to voltage conversion. It is worth noting that even though this resistor serves no purpose electronically, it can make the PCB layout easier.

Use of the long-tailed (or differential) pair in an amplifier means that the amplifier will operate with what is generally called "voltage feedback" (VFB). The feedback is introduced as a voltage, since the input impedance of both inputs is high (and approximately equal), and input current is (relatively speaking) negligible.

The feedback resistor and capacitor are selected to allow the circuit to operate at full open loop gain for the applied AC, but unity gain for DC to allow the circuit to stabilise correctly with a collector voltage at (or near) 0V. The transistors used in the simulations that follow are "ideal", without internal capacitances etc, and have an Hfe of 100 in all cases. The simulated circuits were operated at a voltage of +/- 12.

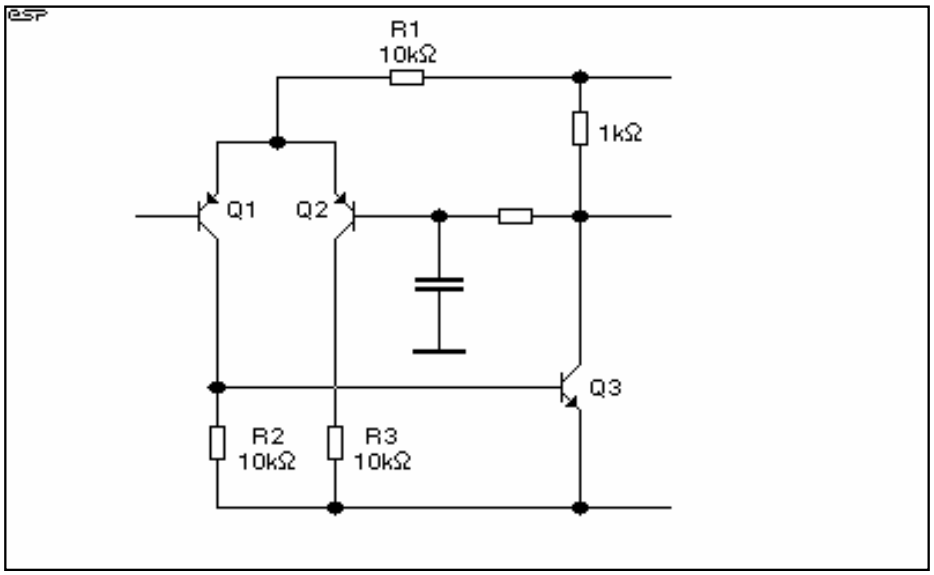


Figure 1a - Aesthetic addition of resistor to balance the collector load.

As shown, and with a 12mA collector current for Q3, the load imbalance at the LTP is 202uA for Q1, and 915uA for Q2. Simply by reducing the value of R2, it is possible to improve matters, but it is still not going to give the performance of which the circuit is capable. Again, as shown the gain of the LTP is a rather dismal 66 (as measured at the collector of Q2).

Changing R1 for a current source does not help (gain is actually reduced slightly), but provides an improvement in power supply hum rejection and in particular improves common mode rejection. A common mode signal is one that is applied in the same phase and amplitude to both inputs at once.

The overall gain of this configuration (measured at the collector of Q3) is 590, but by reducing R2 to 1.8k, it can be raised to 910. This also improves collector current matching in the LTP, but the value will be device dependent, and is not reliable for production units..

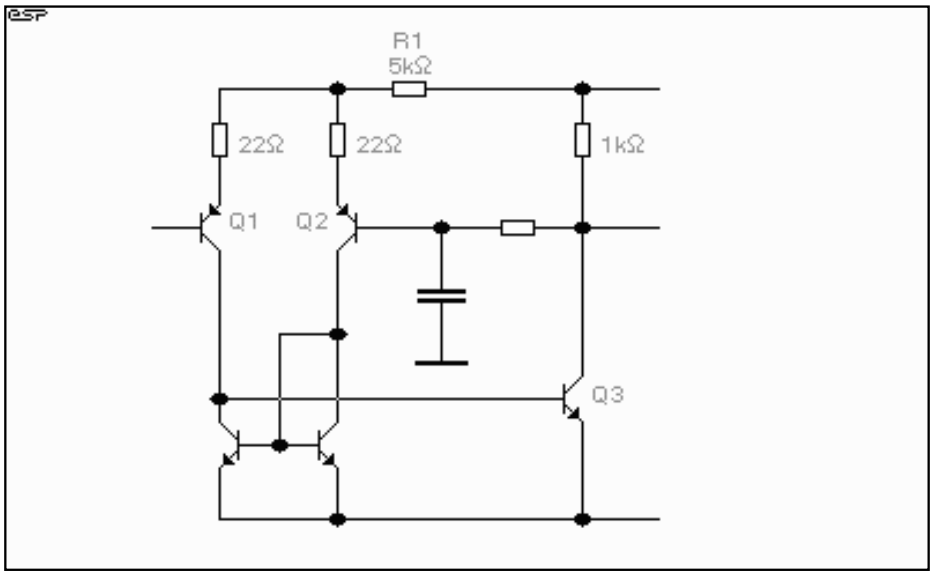


Figure 1b - A current mirror and local feedback applied to the LTP

The circuit shown in Figure 1b has improved overall gain to 2010, a fairly dramatic improvement on the earlier attempt. A further improvement in linearity is to be had by adding resistors (100 Ohm or

thereabouts) into the emitter circuits of the current mirror transistors. This will swamp the base-emitter non-linearities, and provide greater tolerance to device gain variations.

Proper selection of the operating current will improve matters considerably, and also help to reduce distortion, especially if local negative feedback (as shown in Figure 1b) is applied. This has been discussed at length by various writers [1], and a bit of simple logic reveals that benefits are bound to accrue to the designer who takes this seriously.

Since the value of the transistor's internal emitter resistance (r_e) is determined by the current flow -

$$r_e = 26 / I_e (\text{in mA})$$

at very low operating currents this value can be quite high. For example, at 0.5 mA, r_e will be about 52 ohms, increasing further as the current is reduced. Although this will introduce local feedback (and reduce the available gain), it is non-linear, resulting in distortion as the current varies during normal operation. Increasing the current, and using resistors (which are nice and linear) to bring the gain back to where it was before will reduce the distortion, since the resistor value - if properly chosen - will "swamp" the variations in the internal r_e due to signal levels.

At small currents (where the current variation during operation is comparatively high), this internal resistance has a pronounced effect on the performance of the stage. Simple solutions to apparently complex problems abound.

Use of a current mirror as the load for the long-tailed pair (LTP) again improves linearity and gain, allowing either more local feedback elsewhere, or more global feedback. Either of these will improve the performance of an amplifier, provided precautions are taken to ensure stability - i.e. freedom from oscillation at any frequency or amplitude, regardless of applied load impedance.

Single Transistor

There is another (not often used these days) version of an amplifier input stage. This is a single transistor, with the feedback applied to the emitter. It has been claimed by many that this is a grossly inferior circuit, but it does have some very nice characteristics.

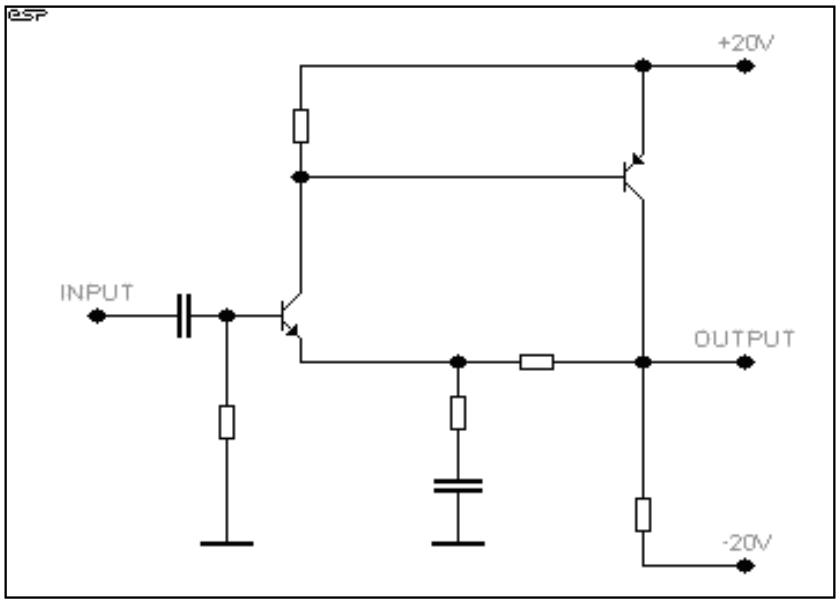


Figure 2a - Single Transistor Input Stage

So what is so nice about this? In a word, stability. An amplifier using this input stage requires little or no additional stabilisation (the "Miller" cap, or "dominant pole") which is mandatory with amps having LTP input stages.

An amplifier using this input stage is referred to as a "current feedback" (CFB) circuit, since the feedback "node" (the emitter of the input transistor) is a very low impedance. The base circuit is the non-inverting input, and has a relatively high input impedance - but not generally as high as the differential pair. The +ve and -ve inputs are therefore asymmetrical. CFB amplifiers are used extensively in extremely fast linear ICs, and are capable of bandwidths in excess of 300MHz (that is not a misprint!).

This is the input stage used in the 10W Class-A amp (John Linsley-Hood's design, which is no longer a part of The Audio Pages), and also in the "El-Cheapo" amp described in my Projects Pages. "Well if it is so good, why doesn't anybody use it?" I hear you ask (you must have said it pretty loudly, then, because Australia is a long way away - from everywhere).

There is one basic limitation with this circuit, and this was "created" by the sudden requirement of all power amplifiers to be able to faithfully reproduce DC, lest they be disgraced by reviewers and spurned by buyers.

> (I remain perplexed by this, since I know for a fact that I cannot hear DC, my speakers cannot reproduce it, I know of no musical instrument that creates it, and it would probably sound pretty boring if any of the above did apply. If you don't believe me, connect a 1.5V torch cell to your speaker, and let me know if I'm wrong. I seem to recall something about phase shift being bandied about at the time, but given the acoustics involved in recording in the studio and reproducing in a typical listening room - not to mention the "interesting" phase shifts generated by loudspeaker enclosures as the speaker approaches resonance - I feel that the effects of a few degrees of low frequency phase shift generated in an amplifier are unlikely to be audible. This is of course assuming that human ears are capable of resolving absolute phase anyway - which they have been categorically proven to be unable to do.)

This input stage cannot be DC coupled (at least not without using a level shifting circuit), because of the voltage drop in the emitter circuit and between the emitter-base junction of the transistor. Since these cannot be balanced out as they are with an LTP input stage, the input must be capacitively coupled.

In addition, some form of biasing circuit is needed, and unfortunately this will either have to be made adjustable (which means a trimpot), or an opamp can be used to act as a DC "servo", comparing the output DC voltage with the zero volt reference and adjusting the input voltage to maintain 0V DC at the output. The use of such techniques will not be examined here, but can provide DC offsets far lower than can be achieved using the amplifier circuit itself. There is no sonic degradation caused by the opamp (assuming for the sake of the discussion that opamps cause sonic degradation anyway), since it operates at DC only (it might have some small influence at 0.5Hz or so, but this is unlikely to be audible).

It has also been claimed that the single transistor has a lower gain than the LTP, but this is simply untrue. Open loop gain of the stage is - if anything - higher than that of a simple LTP for the same device current.

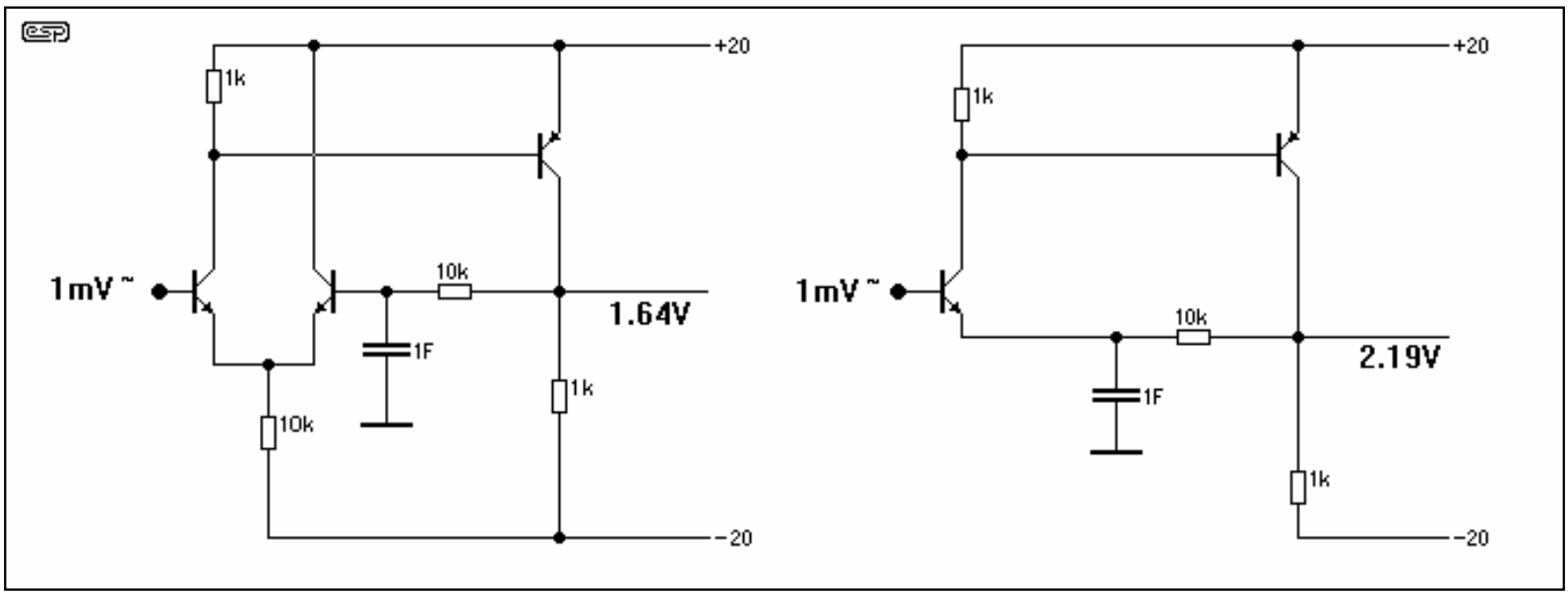


Figure 2b - Voltage Gain Comparison Of Input Stages

I simulated a very simple pair of circuits (shown in Figure 2b) to see the difference between the two. Collector current is approximately 1mA in each, and the output of the LTP shows a voltage gain of 1,640 from the combined circuit (the input stage cannot properly be measured by itself, since it operates as a current amp in both cases). In neither case did I worry about DC offset, since the effects are minimal for the purpose of simply looking at the gain - therefore offset is not shown. (Did you notice that the gains obtained in this simulation are completely different from those obtained earlier for the simple LTP circuit - I used a different voltage (the previous example used +/- 12V). This in no way invalidates anything, they are just different.)

By comparison, the open-loop gain of the single transistor stage is 2,190 - considering that all other things were maintained equal, this is considerably better. Admittedly, the addition of a current mirror would improve the LTP even more dramatically, but do we really need that much more gain? A quick

test indicates that we can get a gain of 3,570. This looks very impressive, but is only an increase of a little over 4.2dB compared to the single transistor. By the same logic, the single transistor only has a 2.5dB advantage over the simple LTP, however the difference may be moot ....

Because the single transistor stage requires no dominant pole Miller capacitor for stability, it will maintain the gain for a much wider frequency range, so in the long run might actually be far superior to the LTP. Further tests were obviously required, so I built them. Real life is never quite like the simulated version, so there was a bit less gain from each circuit than the simulator claimed. The LTP came in with an open loop gain of 1000, while the single transistor managed 1400. The test conditions were a little different from the simulation, in that +/-15 volts was used, so the gain difference is about what would be expected, and is very close to the +/-12V results obtained in the first set of simulations on the LTP.

Distortion was interesting, with the LTP producing 0.7% which was predominantly 3rd harmonic. The single transistor was slightly worse for the same output voltage with 0.9%, and this had a dominant 2nd harmonic.

As expected, the LTP was unstable without a Miller capacitor, and 56pF managed to tame it down. Quite unexpectedly, the single transistor also required a Miller cap, but only when running open-loop. When it was allowed to have some feedback the oscillation disappeared. The LTP could not be operated without the Miller capacitor at any gain, and as the gain approached unity, more capacitance was needed to prevent oscillation.

The next step was a test of each circuit providing a gain of about 27, since this is around the "normal" figure for a 60W power amp. Here, the LTP is clearly superior, with a level of distortion I could not measure. The single transistor circuit had 0.04% distortion, and again this was predominantly 2nd harmonic. In this mode, no Miller capacitor was needed for the single transistor, and it showed a very wide frequency response, with a slight rise in gain at frequencies above 100kHz. This was also noticeable with a 10kHz square wave, which had overshoot, although this was reasonably similar for positive and negative half-cycles. The LTP was well behaved, and showed no overshoot (it had the 56pF Miller cap installed), but it started to run out of gain at about 80kHz, and there was evidence of slew-rate limiting. This effect was not apparent with the single transistor.

All in all, I thought this was a worthwhile experiment, and the use of a simple resistor for the collector load of the gain stage allowed the final circuit to have a manageable gain. Had a current source or similar been used as the load, I would not have been able to measure the gain accurately, since the input levels would have been too small. As it was, noise pickup proved to be a major problem, and it was difficult to get accurate results without using the signal averaging capability on the oscilloscope.

Conclusions

Based on the tests, there are pros and cons to both approaches - and I bet that came as a surprise. The LTP in its simple form is a clear loser for gain, but use of a current mirror allows it to "blow away" the single transistor, which cannot capitalise on this technique since there is nothing to mirror.

Stability is very important to me, and I tend towards an amp which absolutely does not oscillate, even at the expense of a little more distortion, but my own 60W reference amp is unconditionally stable with

normal loads, and it uses an LTP for the input.

Protection From Radio Frequency Interference

A favourite pastime of many designers is to connect a small capacitor as shown in Figure 3 directly to the base of the input transistor. This is supposed to prevent detection (rectification) of radio frequency signals picked up by the input leads. Well, to a certain degree this is true, as the Resistor-Capacitor (RC) combination forms a low pass filter, which will reduce the amount of RF applied to the input. As shown this has a 3dB frequency of 159kHz.

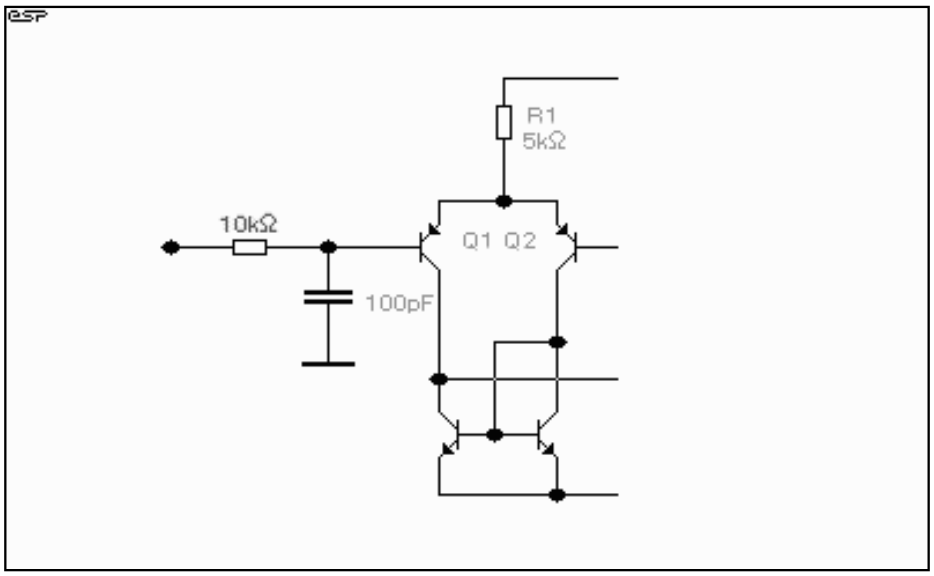


Figure 3 - The Ttraditional Method for Preventing RF Detection

Where things get really sneaky, is when the levels of RF energy are so high that some amount manages to get through anyway. I once had a workshop/lab which was triangulated by three TV transmission towers - very nasty.

The traditional method not only did not work, but made matters worse by ensuring that the transistor base was fed from a very low impedance (from an RF perspective). A vast number of commercial amplifiers and other equipment which I worked on in that time picked up quite unacceptable amounts of TV frame buzz, caused by the detection of the 50Hz vertical synchronisation pulses in the TV signal. As the picture component of TV is amplitude modulated RF, this was readily converted into audio - of the most objectionable kind.

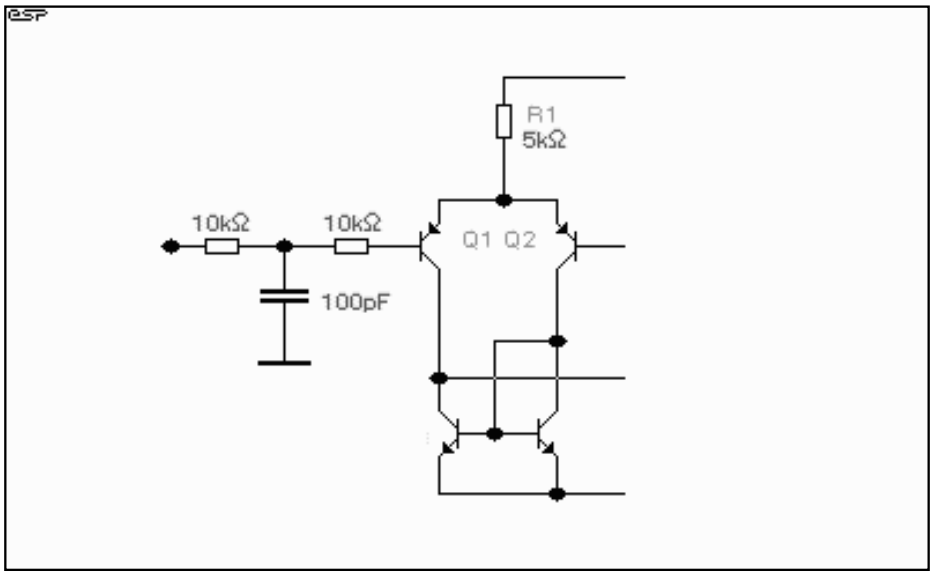


Figure 4 - Use of a Stopper Resistor to Prevent RF Detection

Figure 4 shows the remedy - but to be effective the resistor must be as close as possible to the base, or the performance is degraded. How does this work? Simple, the base-emitter junction of a transistor is a diode, and even when conducting it will retain non-linearities. These are often sufficient to enable the input stage to act as a crude AM detector, which will be quite effective with high-level TV or CB radio signals. Adding the external resistance again swamps the internal non-linearities, reducing the diode effect to negligible levels. This is not to say that it will entirely eliminate the problem where strong RF fields are present, but will at least reduce it to "nuisance" rather than "intolerable" levels.

UPDATE: I have been advised by a reader who works in a transmitting station that connecting the capacitor directly between base and emitter (in conjunction with the stopper resistor) is very effective. He too found that the traditional method was useless, but that when high strength fields are encountered, the simple stopper is not enough.

With opamps, the equivalent solution is to connect the stopper resistor in series with the +ve input, and the capacitor between the +ve and -ve inputs, with no connection to earth.

In all both cases it is essential to keep all leads and PCB tracks as short as possible, so they cannot act as an antenna for the RF. Needless to say, a shielded (and grounded) equipment case is mandatory in such conditions.

Gain Stage (Class-A Amplifier Section) The Class-A amp stage is also commonly known as the Voltage Amplification Stage (VAS), but both terms are common, and are generally interchangeable. There are a number of traps here, not the least of which is that it is commonly assumed that the load (from the output stage) is infinite. Oh, sure, every designer knows that the Class-A stage must carry a current of at least 50% more than the output stage will draw, and this is easily calculated ...

$$I_A = Peak_V / Op_R / Op_Gain * 1.5$$

where I_A is the Class-A current, Peak_V is the maximum voltage across the load Op_R, and Op_Gain

is the current gain of the output transistor combination.

For a typical 100W / 8 Ohm amplifier this will be somewhere between 5 and 10mA. Assuming an output transistor combination with a current gain of 1000 (50 for the driver, and 20 for the power transistor), with an 8 Ohm load, the impedance presented to the Class-A stage will be about 2k Ohms, which is a little shy of infinity.

Added to this is the fact that the impedance reflected back is non-linear, since the driver and output transistors change their gain with current - as do all real-life semiconductors. There are some devices available today which are far better than the average, but they are still not perfect in this respect.

The voltage gain is typically about 0.95 to 0.97 with the compound pair configuration. It must be noted that this figure will only be true for mid-range currents, and will be reduced at lower and higher values. Figure 5 shows the basic stage type - the same basic amplifier we used before, with the addition of a current source as the collector load. Also common is the bootstrapped circuit (not shown here, but evident on many ESP designs).

There is not a lot of difference between current source and bootstrap circuits, but the current source has slightly higher gain. With either type, there are some fairly simple additions which will improve linearity quite dramatically. Figure 5 shows the typical arrangement, but with the omission of the dominant pole stabilisation capacitor, which would normally be connected between the Class-A transistor's collector and base.

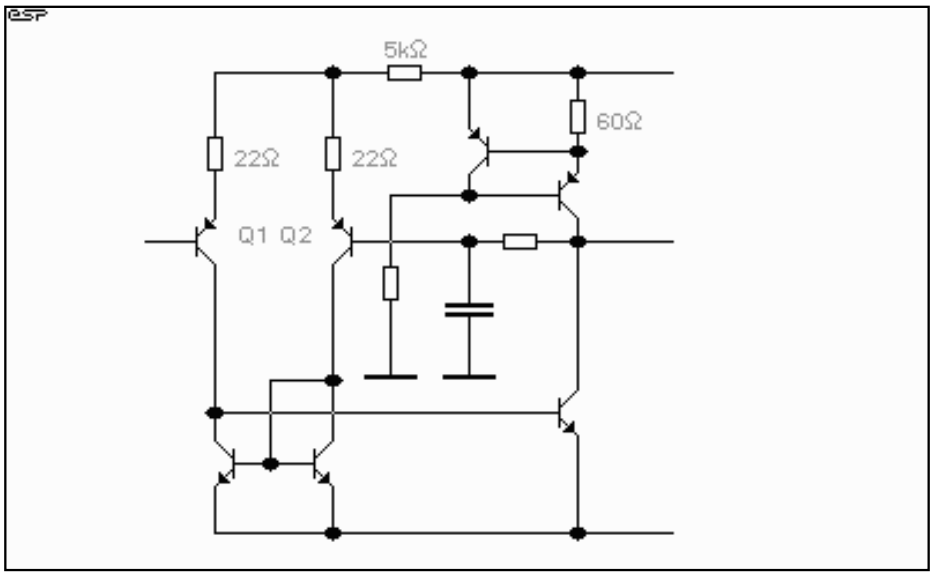


Figure 5 - Typical Class-A Driver Configuration

It is therefore important to try to make the Class-A stage capable of high gain, even when loaded by the output stage. There have been many different methods used to achieve this, but none is completely successful. The biggest problem is that many designers seem completely oblivious to this problem area, or create such amazingly complex "solutions" as to make stabilisation almost impossible.

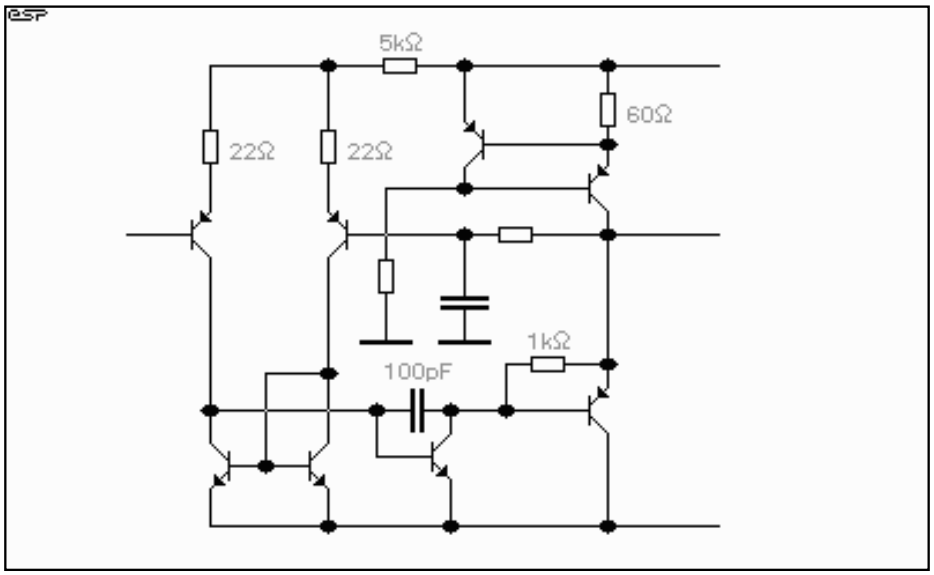


Figure 6 - Improving Open Loop Output impedance of Class-A Driver

The above is simple and very effective. This straightforward addition of an emitter follower to the Class-A driver (with the "bootstrap" resistor) has increased the combined LTP and Class-A driver gain to 1,800,000 (yes, 1.8 million!) or 125dB (open loop and without the dominant pole capacitor connected). Open loop output impedance is about 10k, again without the cap. Once the latter is in circuit, gain is reduced to a slightly more sensible 37,000 at 1kHz with the 100pF value shown. Output impedance at 1kHz is now - comparatively - very low, at about 150 Ohms.

Note that in the above, I have used a 5k resistor instead of the more usual current source. This is for clarity of the drawing, and not a suggestion that the current source should be forsaken in this position.

A special note for the unwary - If one is to use a single current control transistor for both the LTP and Class-A driver, do not use the Class-A current as the reference, but rather the LTP. If not, the varying current in the Class-A circuit will cause modulation of the LTP emitter current, with results that are sure to be as unwelcome as they are unpredictable [4].

I have often seen amplifier designs where the circuit is of such complexity that one must wonder how they ever managed to stop it from becoming high power radio frequency oscillators. The maze of low value capacitors sometimes used - some with series resistance - some without, truly makes one wonder what the open loop frequency and phase response must look like. Couple this with the fact that many of these amps do not have wonderful specifications anyway, and one is forced to ponder what the designer was actually trying to accomplish (being 'different' is not a valid reason to publish or promote a circuit in my view, unless it offers some benefit otherwise unattainable).

UPDATE: Having carried out quite a few experiments over a period of a few weeks, I am not completely certain that vast amounts of gain from the input stage and Class-A amplifier stage are necessary. As long as the circuit is linear (i.e. has low distortion levels before the addition of feedback), the final result is likely to be satisfactory. I have seen many circuits with far more open loop gain than my reference amp (Project 3A), and that in theory should be vastly superior - yet they are not.

Active Current Source or Bootstrap?

There are essentially two ways to create a constant current feed to the Class-A driver stage. The active current source (as shown above) is one method, and this is very common. It does introduce additional active devices, but it is possible to make a current source that has an impedance so close to infinity that it will be almost impossible to measure it without affecting the result just by attaching measurement equipment. For more detailed information on current sources, see the article Current Sources Sinks and Mirrors.

A simpler way is to use the bootstrap circuit, where a capacitor is used from the output to maintain a relatively constant voltage across a resistor. If the voltage across a resistor is constant, then it follows that the current flowing through it must also be constant. Figure 6a shows the circuit of a bootstrap constant current source. Unlike a true current source, the current through the bootstrap circuit will change with the supply voltage. This is a gradual change, and is outside the audio spectrum - or at least it should be if the circuit is designed correctly.

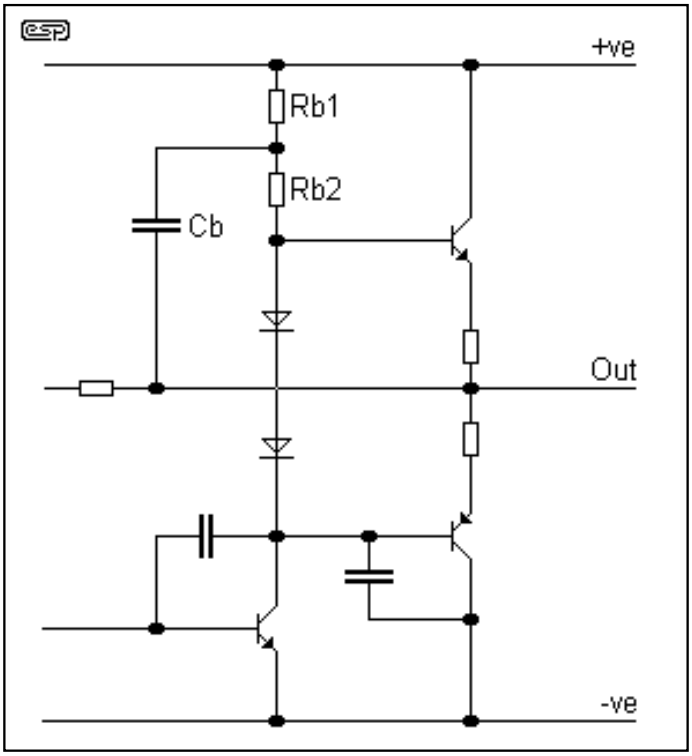


Figure 6a - Bootstrap Current Source

This circuit works as follows. Under quiescent conditions, the output is at zero volts, and the positive supply is divided by Rb1 and Rb2. The base of the upper transistor will be at about +0.7V - just sufficient to bias the transistor. As the output swings positive or negative, the voltage swing is coupled via Cb, so the voltage across Rb2 remains constant. The current through Rb2 is therefore constant, since it maintains an essentially constant voltage across it. Note that this applies only for AC voltages, as the capacitor will charge if there is a DC variation.

The overall difference is not great in a complete design. Although the current source is theoretically better, a bootstrap circuit is simpler and cheaper, and introduces no additional active devices.

Output Stage

There are countless amplifiers which still use the Darlington type configuration, even though this was shown by many [2] to be inferior to the complementary pair. Both configurations (in basic form, since there are many variations) are shown in Figure 7. There are two main areas where the Darlington configuration is inferior, and we shall look at each ...

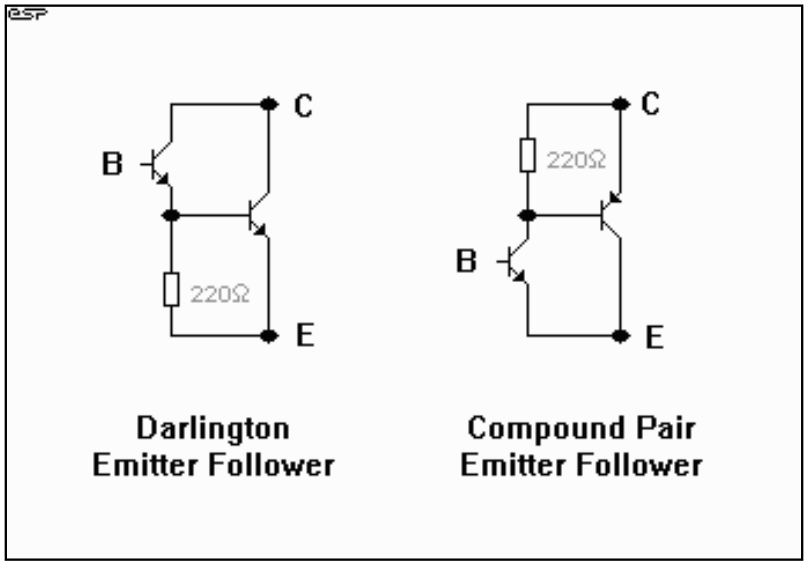


Figure 7 - The basic configurations of output stages in common use

Of the two shown, it will be apparent that I have ignored MOSFET output stages - this is because MOSFETs require no driver transistor as such - they are normally driven directly from the Class-A amplifier (or a modified version - often a modified long-tailed-pair. As can be seen, the component count is the same for those shown, but instead of using two same polarity (both PNP or both NPN), the compound pair (also called a Sziklai pair) uses one device of each polarity. The final compound device assumes the characteristics of the driver in terms of polarity, and the Emitter, Base and Collector connections for each are shown. The resistor is added to prevent output transistor collector to base leakage current from allowing the device to turn itself on, and also speeds up the turn-off time. Omission of this resistor is not a common mistake to make, but it has been done. The results are degraded distortion figures - especially at high frequency - and poor thermal stability.

The value must be selected with reasonable care, if it is too low, the output transistor will not turn on under quiescent (no signal) conditions, the driver transistor(s) will be subject to excessive dissipation, and crossover distortion will result. If too high, turn-off performance of output devices will be impaired and thermal stability will not be as good. The final value depends (to some extent) on the current in the Class-A driver stage and the gain of the driver transistor, but the final arbiter of quiescent is the Vbe multiplier stage. These comments apply equally to the Darlington and compound pairs.

Values of between 100 Ohms up to a maximum of perhaps 1k should be fine for most amplifiers, with lower values used as power increases, which creates higher currents throughout the output stage and makes the transistors harder to turn off again, especially at high frequencies..

Thermal Stability

It can be seen that in the Darlington configuration, there are two emitter-base junctions for each output device. Since each has its own thermal characteristic (a fall of about 2mV per degree C), the combination is notoriously difficult to make thermally stable. In addition, the gain of transistors often increases as they get hotter, thus compounding the problem. The bias 'servo', typically a transistor Vbe multiplier, must be mounted on the heatsink to ensure good thermal equilibrium with the output devices, and in some cases can still barely manage to maintain thermal stability.

If stability is not maintained, the amplifier may be subject to thermal runaway, where after a certain output device temperature is reached, the continued fall of Vbe causes even more quiescent current to flow, causing the temperature to rise further, and so on. A point is reached where the power dissipated is so high that the output transistors fail - often with catastrophic results to the remainder of the circuit and/or the attached loudspeakers.

The compound pair has only one controlling Vbe, and is thus far easier to stabilise. Since the single Vbe is that of the driver (which in many cases will not be mounted on the main heatsink, and in some will have no heatsink at all), the requirements for the Vbe multiplier are less stringent, mounting is far simpler and thermal stability is generally very good to excellent.

I have used the compound pair for over 20 years, and when I saw it for the first time, it made too much sense in all respects to ignore. Thermal stability in a fairly basic 100W/4 Ohm amplifier of my design (of which hundreds were built -it was the predecessor of the P3A design in the projects section) was assured with a simple 2-diode string - no adjustment was ever needed.

Linearity

Numerous articles have been written on the superior linearity of the compound stage (Otala [3], Self, Linsley Hood among others) and I cannot help but be astonished when I see a new design in a magazine, still using the Darlington arrangement. The use of the compound pair requires no more components - the same components are simply arranged in a different manner. It was with great gusto that an Australian electronics magazine proudly announced (in 1998) that "this is the first time we have used this arrangement in a published design" (or words to that effect). I don't know the reason(s) they may have had for not using the complementary pair in *every* design they published (this magazine is a lot younger than I). Words fail me. The magazine in question is not the only one, and the Web abounds with designs old and new - all using the dreaded Darlington emitter-follower.

This is not to say that the Darlington stage shouldn't be used - there are many fine amplifiers that use it, and with a bit of extra effort to get the bias servo right, such amps will give many years of reliable service. It is particularly suited to very high power amps, because of its simplicity - especially with multiple paralleled output devices.

Darlington		
Driver	O/P Transistor	Total Gain
50	25	1310

Compound		
Driver	O/P Transistor	Total Gain
50	25	1290

Figure 8 - Relative Forward Current Gain of Compound Pair vs. Darlington Emitter Follower.

The lower gain of the compound pair indicates that there is internal local negative feedback inherent in the configuration, and all tests that have been performed indicate that this is indeed true. Although the gain difference is not great, much of the improved linearity can be assumed to result from the fact that only one emitter-base junction is directly involved in the signal path rather than two, so only one set of direct non-linearities .is brought into the equation. The second (output) device effectively acts as a buffer for the driver.

Having said that, there are some very well respected amplifiers using Darlington emitter-follower output stages. There are no hard and fast rules that can be applied to make the perfect amplifier (especially since it does not yet exist), and with careful design it is quite possible to make a very fine sounding amplifier using almost any topology.

Design of Bias Servo

It would seem (at first glance at least) that there is nothing to this piece of circuitry. It is a very basic Vbe multiplier circuit, and seemingly, nothing can go wrong. This is almost true, except for the following points.

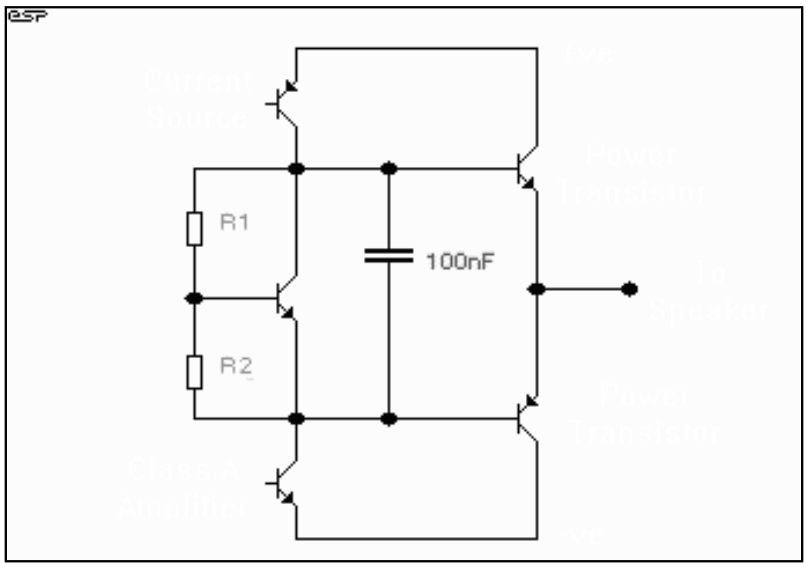


Figure 9 - The basic bias servo circuit

The design of many amps (especially those using a Darlington output stage) requires that the bias servo be made adjustable, to account for the differing characteristics of the transistors. If resistor R1 (in Fig 9) is instead a trimpot (i.e. variable resistor), what happens when (if) the wiper decides (through age, contamination or rough handling) to go open-circuit?

The answer is simple - the voltage across the bias servo is now the full supply voltage (less a transistor

drop or two), causing both the positive and negative output devices to turn on as hard as they possibly can. The result of this is the instantaneous destruction of the output devices - this will happen so fast that fuses cannot possibly prevent it, and even the inclusion of sophisticated Load-Line output protection circuitry is unlikely to be able to save the day.

The answer of course is so simple that it should be immediately obvious to all, but sadly this is not always the case. By making R2 the variable component, should it happen to become open-circuited the bias servo simply removes the bias. This will introduce crossover distortion, but the devices are saved. To prevent the possibility of reducing the pot value to 0 ohms (which will have the same effect as described above!), there should be a series resistor, whose value is selected to allow adequate adjustment while retaining a respectable safety margin.

An additional precaution must be taken here, in that if the resistor values are too low, the offset voltage seen by the output transistors is simply the voltage drop across the resistors, with the transistor having little or no control over the result. This is easily avoided by ensuring that the resistor current is 1/10 (or thereabouts) of the total Class-A bias current.

UPDATE: It is also quite possible to make the resistance too large, so the bias servo will amplify the temperature change too much. This will cause the amplifier's quiescent current to fall as it gets hotter. While this is a good thing from the reliability point of view, if it causes crossover distortion to appear when the amp is hot, the audible effect will obviously be a disaster. It will generally be necessary to experiment with the values to ensure that stability is maintained - there is no way to calculate this that comes to mind, although I am sure it is possible. The base-emitter voltage falls at 2mV / degree C, but the variation in gain with temperature is not as readily calculated.

As a secondary safeguard, using a suitable diode string in parallel with the servo is suggested. These should be chosen to prevent destructive current, but some method of over temperature protection will be needed.

Output Stage Stability

It is a simple fact of life that an emitter follower (whether Darlington or compound) is perfectly happy to become an oscillator - generally at very high frequencies. This is especially true when the output lead looks like a tuned circuit. A length of speaker cable, while quite innocuous at audio frequencies, is a transmission line at some frequency determined by its length, conductor diameter and conductor spacing. A copy of the ARRL handbook (from any year) will provide all the formulae needed to calculate this, if you really want to go that far.

All power amplifiers (well, nearly all) use emitter follower type output stages, and when a speaker lead and speaker (or even a non-inductive dummy load) are connected, oscillation often results. This is nearly always when the amp is driven, and is more likely when current is being drawn from the circuit. It is a little sad that the compound pair is actually more prone to this errant behaviour than a Darlington, possibly because the driver is the controlling element (and its emitter is connected to the load), and has a higher bandwidth.

Some of the 'super' cables - much beloved by audiophiles - are often worse in this respect for their ability to act as RF transmission lines than ordinary Figure-8, zip cord or 3-core mains flex, and are therefore more likely to cause this problem.

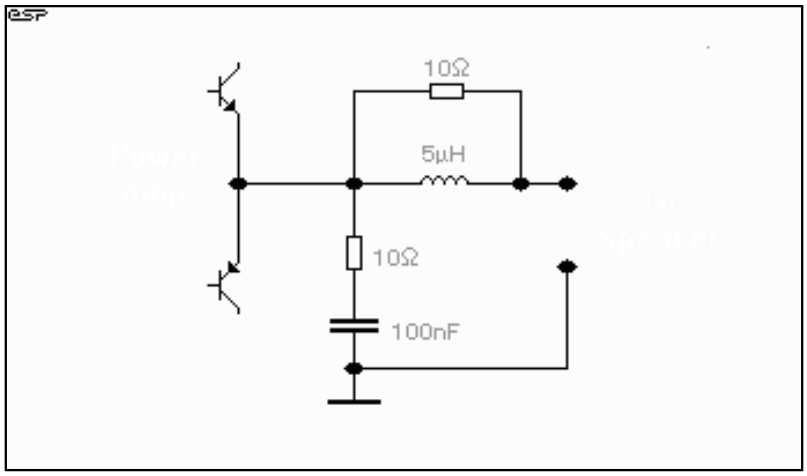


Figure 10 - The standard output arrangement for power amp stability

The conventional Zobel network (consisting of the 10 Ohm resistor and 100nF capacitor) generally swamps the external transmission line effect of the speaker cables and loudspeaker internal wiring, and provides stability under most normal operating conditions.

With capacitive loading (as may be the case when a loudspeaker and passive crossover are connected), the Zobel network has no additional effect whatsoever. The only sure way to prevent oscillation or severe ringing is to include an inductor in the output of the amplifier. This should be bypassed with a suitable resistor to reduce the Q of the inductor, and the typical arrangement is shown in Fig 10. For readers wishing to explore this in greater depth, read "The Audio Power Interface" [2]. Of late (see the note below), I have been suitably perplexed by the sudden appearance of distortion in an amp, simply by adding the inductor. On the basis of this, it might be better to use a far lower resistance than the 10 Ohms specified above - I am thinking 1 Ohm or maybe even less !

Another alternative is to include a resistor in series with the output of the amplifier, but this will naturally have the dual effect of reducing power output and reducing damping factor. At resistor values sufficient to prevent oscillation, the above losses become excessive - and all wasted power must be converted into heat in the resistor.

The choice of inductor size is not difficult - for an 8 Ohm load it will be typically a maximum of 6uH, any larger than this will cause unacceptable attenuation of high frequencies. A 6uH inductor as shown in Figure 10 will introduce a low frequency loss (assuming 0.03 Ohm resistance) of 0.03dB and will be about 0.2dB down at 20kHz. These losses are insignificant, and will not be audible. In contrast, ringing (or in extreme cases, oscillation) of the output devices will be audible (even at very low levels) as increased distortion, and in extreme cases may destroy the transistors.

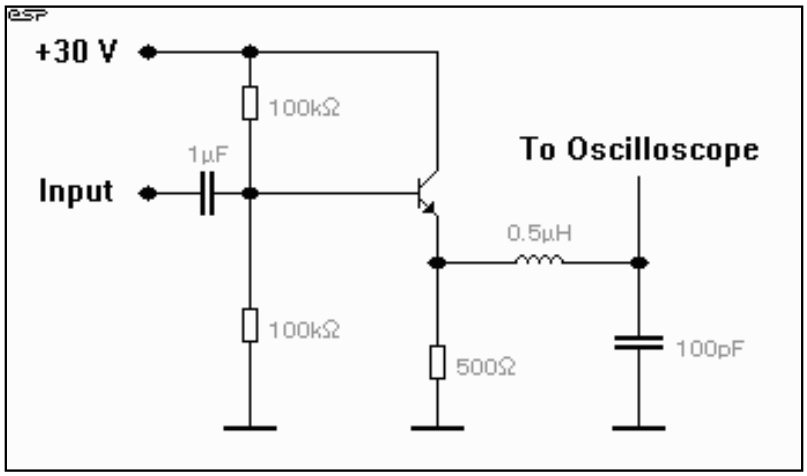


Figure 11 - Lumped component transmission line causes emitter follower to oscillate

In simulations and on the lab bench, I have been able to make a single transistor emitter follower circuit oscillate quite happily, with a real transmission line (such as a length of co-axial cable), or a lumped component equivalent of a transmission line, consisting of a 5nH inductor and 100pF as a series tuned circuit. This is shown in Figure 11.

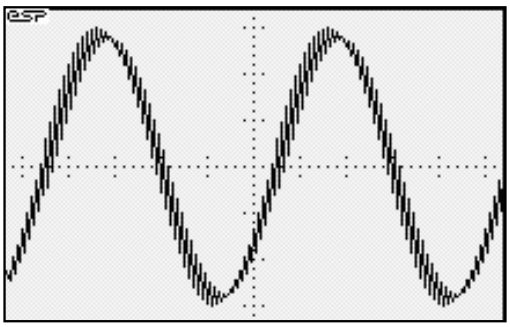


Figure 12 - Simulator oscilloscope display of oscillation in emitter follower

This effect is made worse as the source impedance is lowered, but even a base stopper resistor will not prevent oscillation - only the swamping effect of the transmission line by a Zobel network succeeds. In case you were wondering why the oscilloscope take-off point is at the junction of the L and C components, this allows series resonance to amplify the HF component, making it more readily seen.

In my own amp (the one that hundreds of were made), I did not use an output inductor, but instead made the dominant pole (the capacitance from the collector to base of the Class-A driver) somewhat larger than normal. This managed to keep the amp stable under all operating conditions, but at the expense of slew rate (and consequent slew rate limited power at high frequencies). This was largely an economic decision, since a couple of ceramic capacitors are much cheaper than an inductor, and the amp was used largely for musical instrument amplification so an extended high frequency response was actually undesirable. Full power bandwidth - the ability of an amp to supply full power over its entire operating frequency range - is a sure way to destroy hearing, HF horn drivers (etc) in a live music situation, so the compromise was not a limitation. However ….

There is another reason that a series output inductor may be helpful. It has been suggested (but by whom I cannot remember) that radio frequencies picked up by the speaker leads may be injected back

into the input stage via the negative feedback path. When one looks at a typical circuit, this seems plausible, but I have not tested the theory too deeply.

The basics behind it are not too difficult to work out, however. Since it is known that there must be a dominant pole in the amplifier's open-loop frequency response (the capacitor shown in Figures 5 and 6) if it is to remain stable when feedback is applied, it follows that as internal gain decreases with increasing frequency then the output impedance must rise (due to less global feedback). Indeed this is the case, and by the time the frequency is into the MHz regions, there will be negligible loading of any such frequencies by the output stage.

If appropriate precautions are not taken (as in Figure 4) for the negative feedback return path, then it is entirely likely that RF detection could occur. In my own bi-amped system (which uses the amplifier described above, still without an output inductor), I recently had problems with detection of a local AM radio station. Fitting of RF "EMI" suppression chokes (basically, loop the speaker cable through a ferrite ring 3 or 4 times) completely eliminated the problem, so I must conclude that it is indeed possible or even probable.

Output Current

The maximum output current of a power amplifier is often thought to be something that affects the output transistors only, and that adding more transistors will automatically provide more current to drive lower impedances. This is only partially true, because bipolar transistors need base current, and this must come from the driver stage.

It is common to bias the Class-A driver stage so that it can provide between 1.5 to 5 times the expected base current needed by the output transistors. As the current in this stage is lowered, there is likely to be a substantial increase in the distortion, since the current will change by a larger percentage. If the Class-A driver current is too high, there will be too much heat to get rid of, and it is possible to exceed the transistor's maximum ratings. I normally work to a figure of about double the expected output device base current, but in some cases it will be less than this.

As an example, let's look at a typical power amplifier output stage. Assuming a power supply of +/-35V, the maximum output current will be 35 / 8 = 4.375 Amps (an 8 ohm load is assumed). Since we know that there will be some losses in the driver / power transistor combination, we can safely assume a maximum current of 4A. A suitable power transistor may be specified for a minimum gain (Hfe) of 25, with a collector current of 4A. The driver transistors will generally have a higher gain - perhaps 50 at a collector current of 250mA. The product of the two current gains is accurate enough for what we need, and this gives a combined Hfe of 1,000. The peak base current will therefore be 4mA.

If we choose to use a Class-A driver current of double the expected output device base current, this means that the driver will operate at about 8mA. This could be achieved with a current source, or a bootstrapped circuit using a pair of 2.2k resistors in series. At the maximum voltage swing (close to +/-35V), the driver current will be increased to 12mA or decreased to 4mA, depending on the polarity. The current source or bootstrap circuit will maintain a constant current, but the driver has to deal with a current that varies by +/-4mA as the current into the load changes.

If the load impedance is dropped to 4 ohms, the current source will still only be able to provide 8mA, so output current will be limited to 8A - the driver at this point in the cycle has zero current. At the opposite extreme, the driver will have to cope with 16mA when it is turned on fully. At lower impedances, the driver will be able to supply more current, but the current source will steadfastly refuse to provide more than the 8mA it was designed for, so the peak output current will be limited to 8A in one direction (when the current source provides the drive signal and the Class-A driver is turned off), or some other (possibly destructive) maximum current in the opposite polarity.

But hang on! A Class-A driver is called a Class-A driver because it never turns off - we now have a Class-AB driver, which is not the desired objective! The amplifier will clip asymmetrically, and is no longer operating in the linear range - it is distorting.

Adding more power transistors will provide a very limited benefit, since the maximum base current is still limited by the current source supplying the Class-A driver. In order to obtain maximum power at lower impedances requires that either the gain of the output stage is increased, or the Class-A driver current must be increased. Increasing the gain of the output stage devices is not trivial - you must either use a different topology or higher gain power and driver transistors.

The design phase of an amplifier follows similar guidelines, regardless of topology. From Amplifier Basics ...

- **Power Output vs. Impedance**
  The power output is determined by the load impedance and the available voltage and current of the amplifier. An amplifier that is capable of a maximum of 2A output current will be unable to provide more, just because you want it to. Such an amp will be limited to 16W peak (or 8W RMS) into 8 ohms, regardless of the supply voltage. Likewise, and amp with a supply voltage of +/-16V will be unable to provide more than 8W RMS into 8 ohms, regardless of the available current. Having more current available will allow the amp to provide (for example) 16W into 4 ohms (4A peak current) or 32W into 2 ohms (8A peak current), but will give no more power into 8 ohms than the supply voltage will allow.

- **Driver Current**
  Especially in the case of bipolar transistors, the driver stage must be able to supply enough current to the output transistors - with MOSFETs, the driver must be able to charge and discharge the gate-source capacitance quickly enough to allow you to get the needed power at the highest frequencies of interest.

  For the sake of simplicity, if bipolar output transistors have a gain of 20 at the maximum current into the load, the drivers must be able to supply enough base current to allow this. If the maximum current is 4A, then the drivers must be able to supply 200mA of base current to the output devices.

- **Class-A Driver Stage**
  The stages that come before the drivers must be able to supply sufficient current for the load imposed. The Class-A driver of a bipolar or MOSFET amp must be able to supply enough current to satisfy the base current needs of bipolar drivers, or the gate capacitance of MOSFETs.

Again, using the bipolar example from above, the maximum base current for the output transistors was 200mA. If the drivers have a minimum specified gain of 50, then their base current will be

200 / 50 = 4mA.

Since the Class-A driver must operate in Class-A (what a surprise), it will need to operate with a current of 1.5 to 5 times the expected maximum driver current, to ensure that it never turns off. The same applies with a MOSFET amp that will expect (for example) a maximum gate capacitance charge (or discharge) current of 4mA at the highest amplitudes and frequencies. For the sake of the extercise, we shall assume a Class-A driver (VAS) current of double the base current needs of the drivers ... 8mA.

- **Input Stages**
  The input stages of all transistor amps must be able to supply the base current of the Class-A driver. This time, a margin of between 2 and 5 times the expected maximum base current is needed. If the Class-A driver operates with a quiescent current of 8mA, the maximum current will be 12mA (quiescent + driver base current. Assuming a gain of 50 (again), this means that the input stage has to be able to supply 12 / 50 = 240uA, so it must operate at a minimum current of 240uA * 2 = 480uA to preserve linearity.

- **Input Current**
  The input current of the first stage determines the input impedance of the amplifier. Using the above figures, with a collector current of 480uA, the base current will be 4.8uA for input devices with a gain of 100. If maximum power is developed with an input voltage of 1V, then the impedance is 208k (R = V/I).

  Since the stage must be biased, we apply the same rules as before - a margin of between 2 and 5, so the maximum value of the bias resistors should be 208 / 2 = 104k. A lower value is preferred, and I suggest that a factor of 5 is more appropriate, giving 208 / 5 = 42k (47k can be used without a problem).

These are only guidelines (of course), and there are many cases where currents are greater (or smaller) than suggested. The end result is in the sound of the amp, and the textbook approach is not always going to give the expected result.

Some Notes on Power Supply Design

When specified, regulation is based upon a resistive load over the full cycle, but when used in a capacitor input filter (99.9% of all amplifier power supplies), the quoted and measured figures will never match.

Since the applied AC spends so much of its time at a voltage lower than that of the capacitor, there is no diode conduction. During the brief periods when the diode conducts, the transformer has to replace all energy drained from the capacitor in the intervening 7 to 8ms or so (assuming 50Hz and full wave rectification).

Consider a power supply as shown in Figure 13. This is a completely conventional full-wave capacitor input filter (it is shown as single polarity for convenience). The circuit is assumed to have a total effective series resistance of 1 Ohm - this includes transformer winding resistances (primary and secondary) and diode losses. The capacitor C1 has an initial value of 4,700uF.

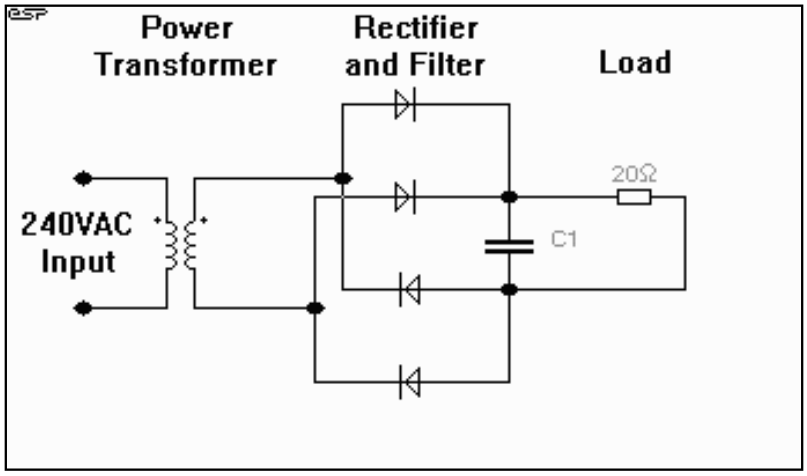


Figure 13 - Full Wave, Capacitor Input Filter Rectifier

The transformer is rated at 60VA and has a primary resistance of 15 Ohms, and a secondary resistance of 0.5 Ohms. This calculates to an internal copper loss resistance of 0.75 Ohm.

With a 20 Ohm load as shown, and at an output current of about 1.7 amps (1.69A actually), diode conduction is about 2.8ms, and the peak value of the current flowing into the capacitor is 6A - 100 times per second (10ms interval). Diode conduction is therefore 28% of the cycle. (To check that these figures are valid, we can take the peak capacitor current and the duty cycle, and calculate the average

$$I_a = (I_p / 100) * \text{duty cycle so, } I_a = (6 / 100) * 28 = 1.68 \text{ (close enough for our purposes)}$$

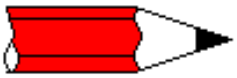
Ripple across the load is 2.4V peak-peak (or about 740mV RMS), and is the traditional sawtooth waveform. Average DC load voltage is 33.8V. The no-load voltage of this supply is 41V, so at a mere 1.7A load, the regulation is ...

$$\text{Reg (\%)} = ((V_n - V_l) / Vn) * 100$$

Where V_n is the no-load voltage, and V_l is the loaded voltage

For this example, this works out to 17.5% which is terrible. By comparison, the actual transformer regulation would be in the order of 5% for a load current of 2A at 30V. Note that the RMS current in the secondary of the transformer will be 2A AC (i.e. DC current multiplied by √2), for a load current of 1.7A DC - this must be so, since otherwise we would be getting something for nothing - a practice frowned upon by physics and the taxman.

Output power is 33.8V * 1.7A = 57.46W, and input power is 240V * 0.25A = 60W (VA actually, but we won't split hairs here). The losses are caused by the diode voltage drop and winding resistance of the transformer.



Capacitor Ripple Current - It is well known that bigger transformers have better efficiency that small ones, so it is a common practice to use a transformer that is over-rated for the application. This can improve the regulation considerably, but also places greater stresses on the filter capacitor due to higher ripple current. This is quoted in manufacturer data for capacitors intended for use in power supplies, and must not be exceeded. Excessive ripple current will cause overheating and eventual failure of the capacitor.

Large capacitors usually have a higher ripple current rating than small ones (both physical size and capacitance). It is useful to know that two 4,700uF caps will usually have a higher combined ripple current than a single 10,000uF cap, and will also show a lower ESR (equivalent series resistance). The combination will generally be cheaper as well - one of the very few instances where you really can get something for nothing.

For further reading on this topic, see the Linear Power Supply Design article.

Measurements Versus Subjectivity

If I *never* hear someone complaining that "distortion measurements are invalid, and a waste of time" again, it will be too soon. I am so fed up with self-proclaimed experts (where 'x' is an unknown quantity, and a 'spurt' is a drip under pressure) claiming that 'real world' signals are so much more complicated than a sinewave, and that static distortion measurements are completely meaningless.

They ***are not*** meaningless, and real world signals ***are*** sinewaves! The only difference is that with music, there is usually a large number of sinewaves, all added together. There is not a myriad of simultaneous signals passing through an amp, just one (for a single channel, naturally).

Since physics tells us that no two masses can occupy the same physical space at the same time, so it is with voltages and currents. There can only ever be one value of voltage and one value of current flowing through a single circuit element at any instant of time - if it were any different, the concept of digital recording could never exist, since in a digital recording the instantaneous voltage is sampled and digitised at the sampling rate. This would clearly be impossible if there were say 3 different voltages all present simultaneously.

So, how do these x-spurts determine if an amplifier has a tiny bit of crossover distortion (for example). I can see it as the residual from my distortion meter, and it is instantly recognisable for what it really is, and I can see the difference when I make a change to a circuit to eliminate the problem. If I had to rely on my ears (which although getting older, still work quite well), It would take me much longer to identify the problem, and even longer to be certain that it was gone. I'm not talking about the really gross crossover distortion that one gets from an under-biased amp, I am referring to *vestiges* - miniscule amounts that will barely register on the meter - I use my oscilloscope to see the exact distortion waveform. I suspect that this dilemma is 'solved' by some by simply not using the push-pull arrangement at all, thereby ensuring that power is severely limited, and other distortion is so high that they would not dare to publish the results.

These same x-spurts will wax lyrical about some really grotty single ended triode amp, with no power and a highly questionable output transformer, limited frequency response and a damping factor of unity if it is lucky.

Don't get me wrong - I'm not saying that this is a definition of single-ended triode amps (for example), there are some which I am sure sound very nice - not my cup of tea, but 'nice'. I have seen circuits published on the web that I would not use to drive a clock radio speaker (no names, so don't ask), and 'testimonials' from people who have purchased this rubbish, but there are undoubtedly some that do use quality components and probably sound excellent at low volume levels.

Sorry if I sound vehement (vitriolic, even), but quite frankly this p****s me off badly. There are so many people waving their 'knowledge' about, and many of them are either pandering to the Magic Market, or talking through their hats.

The whole idea of taking measurements is to ensure that the product meets some quality standard. Once this standard is removed, and we are expected to let our ears be the judge, how are we supposed to know if we got what we paid for? If the amp turns out to sound revolting (to us), should we accept this, or perhaps we should listen to it for long enough that we get used to the sound (this will happen - eventually). I am not willing to accept this, and I'm sure that many others feel the same.

Please don't think that I am advocating specsmanship, because I'm not. I just happen to think that consumers are entitled to some minimum performance standard that the equipment should meet (or exceed). I have yet to hear ***any*** amplifier with high distortion levels and/or limited bandwidth sound better than a similar amplifier with lower distortion and wider bandwidth. This implies that we compare like with like - a comparison between a nice valve amp and a nasty transistor amp will still show the transistor amp as having better specs, but we can be assured that it will sound worse.

Valves Vs. Transistors Vs. MOSFETs

This is an argument that has been going for years, and it seems we are no closer to resolving the dilemma than we ever were. I have worked with all three, and each has its own sonic quality. Briefly, we shall have a look at the differences - this is not an exhaustive list, nor is it meant to be - these are the main points, influenced by my own experiences (and I must admit, prejudices). Please excuse the somewhat random order of the comparisons :

Valves :
Valves are Voltage to Current Converters, so the output current is controlled by an input voltage. It is necessary to apply the varying output current to a load (the anode resistor or transformer) to derive an output voltage.

- Valves themselves are inherently quite linear, and can operate with no feedback at all within a restricted range, and still provide a high quality signal. The range is usually more than sufficient for preamps, but is pushed to its limits in power amplifiers.
- Relatively low gain per device, meaning that more are needed, or less feedback can be used.
- 'Soft' distortion characteristics, meaning that most of the distortion is low order (including

crossover distortion and clipping) - this is not as obtrusive or fatiguing as 'hard' distortion.
- Distortion onset is gradual, and effectively warns the listener that the limits are being approached by losing clarity, but in a manner that is not too obtrusive.
- Distortion is usually measurable at nearly any power level, but is low order (mainly 2nd, with some 3rd harmonics - vestigial amounts of additional harmonics are usually also present).
- Limited feedback, mainly due to the fact that the output transformer introduces low and high frequency phase shift, so large amounts of global feedback are generally not possible without oscillation. This results in a (comparatively) limited bandwidth.
- High output impedance, meaning that damping factor in power amps is generally rather poor. Extremely low values of output impedance are very difficult to achieve.
- Valves have a perfect dielectric (mainly a vacuum, with some mica), leading to a highly linear Miller capacitance - it is unknown if this contributes any audible benefit.
- Inefficient output stage, allowing the amp to sound louder than it really is on a watt for watt basis. This may sound like a contradiction, but a valve amp has a 'compliant output, that allows it to provide a larger voltage swing to high impedance loads (such as a loudspeaker driver at resonance).
- Very rugged, and can withstand short circuits without damage - BUT open circuits can cause the output transformer to create high flyback voltages that can cause insulation breakdown in the transformer windings or the valve sockets (short circuits are OK, open circuits are bad)
- Usually quite tolerant of difficult loads, such as electrostatic loudspeakers.
- A wonderful nostalgia value, which allows people to accept the shortcomings, and truly believe that the amp really does sound better than a really good solid-state unit. Proper double-blind testing will usually reveal the truth - provided that the solid-state equivalent is modified to match the output impedance of the valve unit ***!!!***

Transistors :

By default, transistors are Current to Current Converters. This means that they use an input current change to derive an output current change that is greater than the input (therefore amplification occurs). Again, it is necessary to use a resistor or other load to allow an output voltage to be developed.

- Transistors are also quite linear within a restricted range, but due to the lower operating voltages cannot successfully be used without feedback if a quality signal is desired, even in preamp stages.
- High to very high gain per device, allowing local feedback to linearise the circuit before the application of global feedback.
- Onset of distortion is sudden and without warning. This is worsened as feedback is increased.
- Low to very low distortion, provided clipping is not introduced. This creates both the low order harmonics of the valve amp, plus high order harmonics which are very fatiguing, and generally sound hard and nasty.
- Wide to very wide bandwidth, and low phase shift, largely due to the elimination of the output transformer. The wide bandwidth is obviously an advantage, the phase response is highly debatable as to its overall value to the listener.
- Usually large amounts of global feedback, which is needed to linearise the output stage, especially at the crossover point between output devices (0 Volts).
- Completely oblivious to open circuit loads, but must be protected against instant damage with short circuited outputs (open circuits are OK, short circuits are bad - i.e. the opposite of valves)

- The Miller capacitance of transistors has an imperfect dielectric, and varies with applied voltage. This might be the reason that some transistor amps can be seen to oscillate at a specific voltage level (small bursts of oscillation on the waveform, but only above a certain voltage across the device). Tricky.
- Intolerant of difficult loads, unless extensive measures are taken to ensure stability. This can increase complexity quite dramatically.

MOSFETs :
MOSFETs, like valves, are voltage to current converters, and rely on a voltage on the gate to control the output current. As before, a resistor or other load converts the varying current into a voltage.

- Similar to most of the comments about transistors, with the following differences:
- Onset of (clipping) distortion is (usually) not quite as savage as transistors, but is much more sudden than valves. This is a very minor difference, and can safely be ignored.
- Not as linear as valves or transistors, especially near the cutoff region. Big differences between different types (lateral/ vertical)
- More efficient than valves, but not as efficient as transistors. There will always be less output voltage swing available from a MOSFET amp than a transistor amp (for the same supply voltage), unless an auxiliary power supply is used.
- Gain is (usually) higher than valves, but lower than transistors - limited ability to apply local feedback, and even overall (global) feedback may not produce distortion figures as good as transistors - especially with vertical MOSFETs.
- Low distortion (lateral types), but may require more gain in the preceding stages to allow sufficient feedback to eliminate crossover distortion.
- Very wide bandwidth (better than transistors), allowing less compensation and full power operation up to 100 kHz in some amps - the value of this is debatable.
- More rugged than transistors, and do not suffer from second breakdown effects - fuses can be used for short circuit protection, and no open circuit protection is needed.
- Reasonably tolerant of difficult loads without excessive circuit complexity.

To make matters worse, there are two main types of MOSFET as hinted at above - lateral and vertical. This applies to the internal construction. Lateral MOSFETs are well suited to audio (see Project 101), while vertical (e.g. HEXFETs) are not really suitable at all. Despite this, it is possible to make an amplifier using HEXFETs that performs very well, and this has been achieved by many hobbyists and manufacturers.

Because of the differences outlined above, it is very important to compare like with like, since each has its own strengths and problems. Also, each of the solid state amp types has its niche area, where it will tend to outperform the other, regardless of specifications. The valve amp is the odd man out here, as it is far more likely to have devoted fans who would use nothing else - most solid state amp users are (or should be) a pragmatic lot, using the most appropriate configuration for the intended application.

There is no such thing at the time of writing as the much sought after (but elusive) 'straight wire with gain'. But wait - there's more ....

Slew Rate and Intermodulation

Another aspect of amplifier design is slew rate. This term is not well understood, and the possible effects even less so.

Slew Rate Nomograph

It has been claimed by many writers on the subject that a slew-rate limited amplifier will introduce transient intermodulation distortion, or TIM. In theory, this is perfectly true, provided that the slew rate is sufficiently low as to be within the audible spectrum (i.e. below 20 kHz), and the program material has sufficient power at such high frequencies to cause the amplifier to limit in this fashion.

The following nomogram is helpful in allowing you to determine the required slew rate of any amplifier, so that it can reproduce the required audio bandwidth without introducing distortion components as a result of not being fast enough.

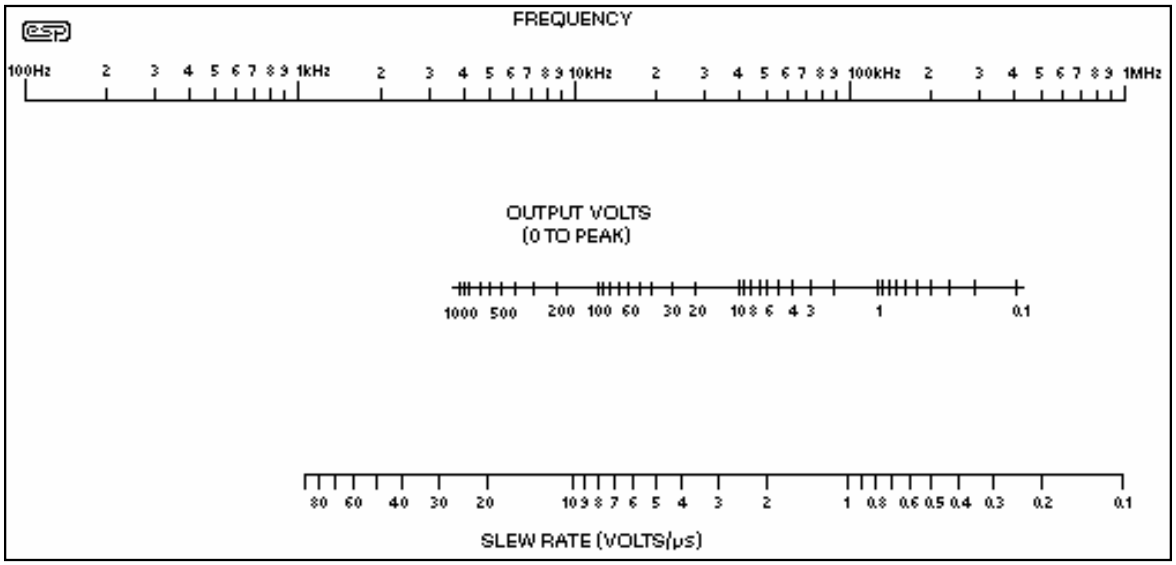


Figure 14 - Slew Rate Nomograph

To use this nomograph, first select the maximum frequency on the top row. Let's assume 30kHz as an example. Next, select the actual output voltage (peak, which is RMS * 1.414) that the amplifier must be able to reproduce. For a 100W 8 Ohms amp, this is 28V RMS, or 40V peak. Now draw a line through these two points, and read the slew rate off the bottom row. For the example, this is 8V/us. This is in fact far in excess of what is really needed, since it is not possible for an amp reproducing music to have full power at 30kHz.

By 20kHz, our 100W amp will need a maximum of 10W (typically much less), and this is only about 12V peak. Using the nomograph with this data reveals that a slew rate of about 2V/us is quite sufficient. Such an amp will go into what is known as slew-rate limiting at full power with frequencies above 10kHz or so, converting the input sinewave into a triangular wave whose amplitude decreases with increasing frequency.

Some claim that this is audible, although this is purely subjective. That a typical audio signal is a complex mixture of signals is of no real consequence, because an amp has no inherent concept of 'complex' any more than it has an opinion about the date. At any given point in time, there is an

instantaneous value of input voltage that must be increased in amplitude and provide the current needed to drive the loudspeaker. As long as this input voltage does not change faster than the amplifier is capable of doing, then no degradation is should occur.

Although this is a fine theory, there seems to be too much entrenched prejudice against 'slow' amplifiers, which may indeed sound different from another that is not constrained by slew rate limiting - at least within the full audible range.

As shown above, the slew rate of an amplifier is usually measured in Volts / microsecond, and is a measure of how fast the amplifier's output can respond to a rapidly changing input signal. Few manufacturers specify slew rate these days (mainly because few buyers understand what it is), but it is an important aspect of an amplifier's design.

As can be seen from the above, for an amplifier (of any configuration) to reproduce 28V at 20 kHz (about 100W / 8 Ohms) requires a slew rate of 4.4 V / us. This is to say that the output voltage can change (in any direction) at the rate of 4.4 Volts in one microsecond. This is not especially fast, and as should be obvious, is dependent upon output voltage. A low power amp need not slew as fast as a higher powered amp. There is no real requirement for any amp to be able to slew faster than this, as there is a significantly large margin provided already. This can be calculated or measured - I measured it at 22V maximum rate of change in 5us (give or take a little), which works out to 4.4V/us again.

Doubling the amplifier's output voltage (four times the power) requires that the slew rate doubles, and vice versa, so a 400W amp needs a slew rate of 8.8 V / us, while a 25W amp only needs 2.2 V /us. This is a very good reason to use a smaller amplifier for tweeters in a tri-amped system, since it is much easier to achieve a respectable slew rate when vast numbers of output devices are not required.

Essentially, if the amplifier's output cannot respond to the rapidly changing input signal, an error voltage is developed at the long-tailed pair stage, which tries to correct the error. The LTP is an amplifier, but more importantly, an error amplifier, whose sole purpose is to keep both of its inputs at the same voltage. This is critical to the operation of a solid state amp, and the LTP output will generally be a very distorted voltage and current waveform, producing a signal that is the exact opposite of all the accumulated distortions within the remainder of the amp (this also applies to opamps).

The result is (or is supposed to be) that the signal applied to the inverting input is an inverted *exact* replica of the input signal. Were this to be achieved in practice, the amp would have no distortion at all. In reality, there is always some small difference, and if the Class-A driver or some other stage enters (or approaches) the slew rate limited region of operation, the error amp (LTP) can no longer compensate for the error.

Once this happens distortion rises, but more importantly, the input signal is exceeding the capabilities of the amplifier, and the intermodulation products rise dramatically. Intermodulation distortion is characterised by the fact that a low frequency signal modulates the amplitude (and / or shape) of a higher frequency signal, generating additional frequencies that were not present in the original signal. This also occurs when an amplifier clips, or if it has measurable crossover distortion.

Sounds like ordinary distortion, doesn't it? That too creates frequencies that were not in the original,

but the difference is that harmonic distortion creates harmonics (hence the name), whereas intermodulation distortion creates frequencies that have no harmonic relationship to either of the original frequencies. Rather, the new frequencies are the sum and difference of the original two frequencies. (This effect is used extensively in radio, to create the intermediate frequency from which the audio, video or other wanted signal can be extracted.) The term 'harmonic' basically can be translated to 'musical', and 'non-harmonic' is mathematically derived, but not musically related .... if you see what I mean.

Whenever the LTP loses control of the signal, intermodulation products will be generated, so the bandwidth of an amplifier must be wide enough to ensure that this cannot happen with any normal audio input signal. There is nothing wrong or difficult about this approach, and it is quite realisable in any modern design. Although unrealistic from a musical point of view, it is better if an amplifier is capable of reproducing full power at the maximum audible frequency (20 kHz) than if it starts to go into slew rate limiting at some lower frequency.

The reason I say it is unrealistic musically is simply because there is no known instrument - other than a badly set up synthesiser - that is capable of producing any full power harmonic at 20 kHz, so in theory, the amp does not have to be able to reproduce this. In reality, inability to reproduce full power at 20 kHz means that the amp will very likely suffer from some degree of transient intermodulation distortion with some program material.

This is not a problem that affects simple amps with little or no feedback - they generate enough harmonic distortion to more than make up for the failings of more complex circuits with lots of global feedback. This fact tends to annoy the minimalists, who are often great believers in no feedback under any circumstance, which relegates them to listening to equipment that would have been considered inferior in the 1950's.

Frequency Response, etc.

Few sensible people would argue that measurements of frequency response are unimportant or irrelevant, and this is one of the simplest measurements to take on an amplifier. Again, the subjectivists would have it that these fail to take into account some mysterious area of our brain that will compensate for a restricted response, and allow us to just enjoy the experience of the sound system. This is true - we will compensate for diminished (or deranged) frequency response, but it need not be so.

If you listen to a clock radio for long enough, your brain will think that this is normal, and will adjust itself accordingly. Imagine your surprise when you hear something that actually has real low and high frequencies to offer - the first reaction is that there is too much of everything, but again, the brain will make the required allowances and this will sound normal after a time.

There are so many standard measurements on amplifiers that are essential to allow us to make an informed judgement (is this amp even worth listening to?). I really object to the attitude that "it does not matter what the measurements say, it sounds great". In reality this is rarely the case - if it measures as disgusting, then it will almost invariably sound disgusting. There is no place for hi-fi equipment that

simply does not meet some basic standards - and I have *never* heard an amp that looked awful on the oscilloscope, measured as awful on my distortion meter, but sounded good - period.

Further Reading

For further reading, I strongly recommend The Self Site, and in particular 'Science and Subjectivism in Audio' and also my own article on the subject Cables, Interconnects & Other Stuff - The Truth. There is also an article called Amplifier Sound - What Are The Influences? that goes a little deeper into the measured and subjective performance of amplifiers, and suggests a couple of new tests that might be applied.

References

1. Refer to The Self Pages (available from links in the ESP Home Page).
2. The Audio Power Interface, Douglas Self, Electronics World September 1997, p717
3. Intermodulation at the amplifier-loudspeaker interface, Matti Otala and Jorma Lammasneimi, Wireless World, December 1980, p42
4. Douglas Self - actual source unknown (but I did read it in one of his papers!)

Main Index

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999, 2000, 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Last Update: 09 Apr - added VFB and CFB info./ 18 Feb - added some minor details (various)./ 09 Apr-Added a few small extras and a correction./ 26 Feb-Added reference to Amp Sound article./ 29 Jan 2000-Fixed a couple of confused statements and typos, added missing component references to Figures 1a and 1b./ 17 Dec-added slew rate nomograph./ 15 Dec 99-added extra info about reference amp, use of inductor in O/P stage, bias servo, and RF stoppers./ 16 Jan 06 - additional comments on MOSFETs, minor reformatting + spelling corrections.

The Audio Pages

Elliott Sound Products **Using HEXFETs in High Fidelity Audio**

Using HEXFETs in High Fidelity Audio

Mitch Hodges
Edited and Updated by Rod Elliott (ESP)

Last Update: 13 March 2006

Articles Index

Main Index

Contents

- 1 - Introduction
- 2 - Description
- 3 - Conclusion
- 4 - References
- 5 - Footnote
  - 5.1 - Update (March 2006)

Introduction

When we build linear power amplifiers, we always need to choose some device for the output stage. This could be any power device including valves, BJTs, IGBTs, and MOSFETs. Each has its own strengths and weaknesses which forces us to choose between them. If, perchance, we wanted to build a very simple and accurate amplifier, we can safely ignore valves, since they all need heating circuitry and are not simple for a true hi-fi amplifier. That it is possible to build a valve amp to a high specification is not in doubt, but they tend to be complex and expensive.

BJTs are often used, but they do not respond well to even momentary overloads. This is because they suffer from second-breakdown - an instantaneous and catastrophic failure mode. IGBTs (Insulated Gate Bipolar Transistors) are seldom used, and will be very similar to a BJT, only with an insulated gate. They still need thermal compensation and a suitable gate drive design, and can suffer from a 'latch-up' condition in some cases. Lastly there is the MOSFET, which does not suffer any second-breakdown effects. MOSFETs (Metal Oxide Semiconductor Field Effect Transistors) come in two primary types - vertical and lateral.

These devices are extremely rugged, yet they do have a large nonlinear gate capacitance to deal with. If driven incorrectly, they show high distortion levels, especially vertical types - most commonly these days, HEXFETs. This is why I wrote this article - to show how to use HEXFETs properly in audio applications.

The update below has some important information that I recommend you read thoroughly and make sure you understand before settling on the use of HEXFETs in your next amp project. While there appear to be many advantages to their use over BJTs, HEXFETs may often suffer from exactly the same problems - thermal runaway and a failure mode that is suspiciously similar to second breakdown. On top of this, there is a much larger voltage loss ... 2-4V is needed to bias the HEXFETs to the on condition, vs. 0.65V (nominal) for BJTs. This voltage is usually taken from the main supplies, so for a given supply voltage, expect less output power.

2 - Description

Many people say (including IRF) that HEXFETs are not suited for linear audio circuits and should be avoided. Well, that is the easy route to take for designing an output stage. *Any* device can be used for audio and give great performance if a proper design is found. It is just easier to use more linear devices. Lateral MOSFETs are usually specified for audio, but there are relatively few different devices on the market. When found, they tend to be quite expensive. On the other hand, HEXFETs are very common, reasonably priced, and only need a good design to do well.

This article is intended for Class-AB designs. HEXFETs will run Class-A with barely any problems besides driving the gate. If you are designing a class A amplifier, the first trick (see below) should be used (the second is not needed since the bias is already quite high).

Alright, now for some explanations. Comparatively, HEXFETs usually have a low gate capacitance than other vertical MOSFETs, yet have a higher gate capacitance than their lateral counterparts. There is not only one capacitance to deal with, but two (one from the gate to source and the other from the gate to drain). This is the main problem: to find a way to drive the gate capacitance of the HEXFETs. Through a lot of time and molten breadboards, I found the best two things to design for are the following:

1. Fix: Drive the gates with as much current as possible. This may include adding a class AB driver stage.
   Why: HEXFETs have a nonlinear transfer curve up to about an amp or two, depending on the device(s) used. In a class AB amplifier, this characteristic is the cause for a majority of the THD. When driven with enough current, the device will follow the 'new' linear curve, since it is balancing out the nonlinear gate capacitance. The lower impedance of the driver stage the better.
2. Fix: HEXFETs like to run hot. This does not mean use an inadequate heatsink, but the bias between devices should be a bit more than many are used to. 250mA of idle current is not a bad bias figure for these devices.
   Why: To balance out the nonlinear curve, we can simply cut if off where it seems too bad by using bias. This will increase dissipation, though.

For the design of the amplifier, I will assume a single LTP input stage. Better performance can be seen by using multiple LTPs, but this will not be a simple design (in fact it will be quite complex with high frequency stability issues needing attention).

When choosing the complementary output components, one can obviously choose the IRFPXXX and its IRFP9XXX complement. If we look at these complementary device data sheets, we will see very different figures for current capability, on resistance, and, most importantly, gain (or forward transconductance). But if we use a matching tool, we will find that the gain varies considerably from the actual devices vs. the data sheet. That is why we need to buy a few extra and match them together. Since the gain varies a bit from batch to batch, it is quite easy to find a IRFPXXX and IRFP9XXX that are very similar, at least with gain factors.

Also take note that HEXFETs will require a Vbe multiplier for thermal compensation, since the negative temperature coefficient does not come into play until the device has about 10 amps through it (at least for the IRFP240). The exact values around the Vbe multiplier (also known as a bias servo) are critical to ensure that the thermal performance is matched as closely as possible.

In every practical design I have tried I had to use a class AB driver stage. A class-A driver will work fine if you really want an electric heater, as you will see in the next calculation. Now, in order to size-up the proper driver for the FETs, we need to do a little maths. I promise it is not hard. An example would work nicely here ... if we wanted to design a class AB driver stage with five IRFP240 and five IRFP9240 devices, how much current will we need at *minimum* for full functionality up to 50kHz? For a better understanding, a simplified output stage circuit is shown below.

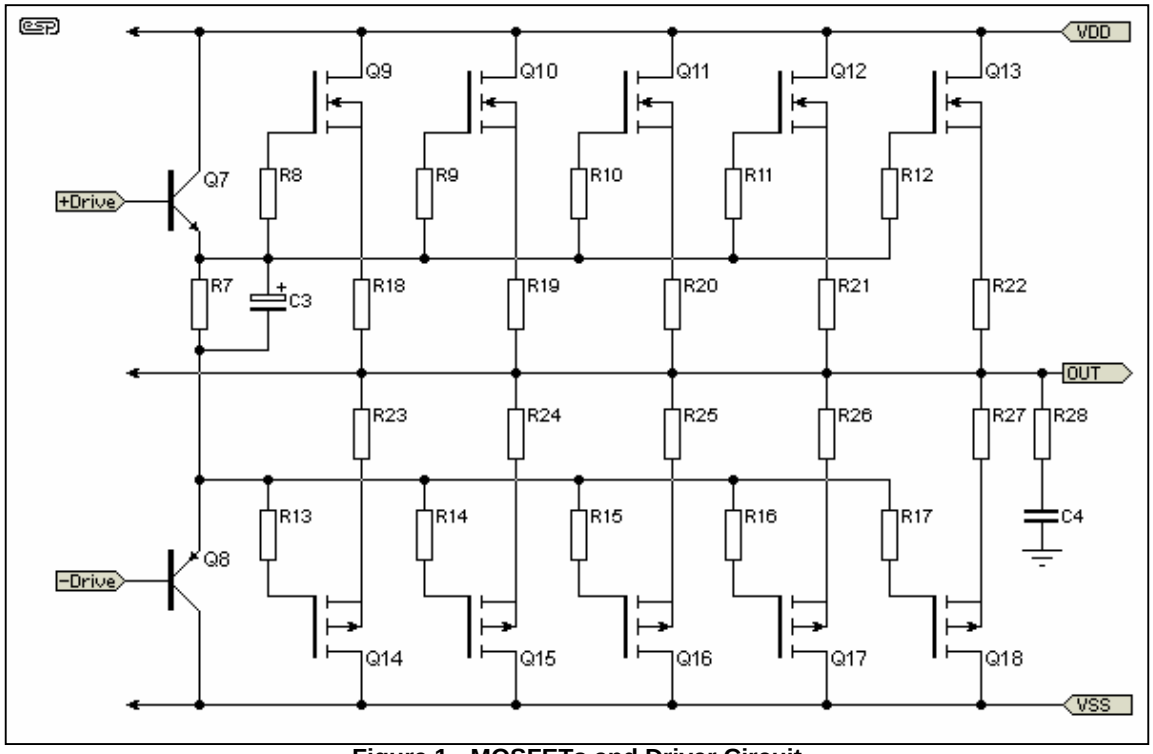


Figure 1 - MOSFETs and Driver Circuit

We will do calculations using the gate charge method, which IRF recommends (AN-944). Looking at the data-sheets, we find the IRFP240 has a total gate charge (Qg) of 70nC and the IRFP9240 has a Qg of 44nC. Don't add these yet! We will find each device's needs individually. The general formula to determine gate current is ...

$I = 100Q_g f$ where I is current needed, Q_g is the total gate charge in Coulombs, and f is frequency of operation.

The multiplication factor of 100 gives the headroom needed for accurately reproducing a square wave (or high frequency sinewave), since the gate driver needs a lot of current to quickly switch the MOSFET from OFF to ON. Although the requirement for this is minimal (the CD format is incapable of anything even approaching a square wave above a couple of kHz), it has become an expectation that power amps should be able to reproduce perfect square waves at 10kHz as a minimum.

When we plug our figures in for the IRFP240, we get $I = 100 * (70^{E-9} * 50{,}000)$ = 350mA per device

For the IRFP9240 we get $I = 100 * (40^{E-9} * 50{,}000)$ = 200mA per device

Multiplying each figure by five (because there are five devices of each polarity) gives us 1.75 amps for the upper driver and 1 amp for the lower driver. So a Class-A driver would need bias set to 3.5 amps to get the job done with a reasonable safety margin.

The value for R7 will depend on the linearity of the driver transistors. I had to guess and check with my ammeter to get a good value. This can range anywhere from 100 Ohms up to perhaps 5k. Make sure you check the idle current before calling the design done! These drivers (Q7 and Q8) may need a heatsink. Also note the capacitor in parallel with R7. This should be of a high value (i.e. 100uF or more), and 470uF works fine for my 10 MOSFET stage shown here. It helps with discharging the MOSFET gates by providing a path for the gate current.

These current figures seem quite high, but keep in mind this current will only last a very short time compared to the signal, and virtually no current is needed to keep the devices either in the OFF or ON state. The current to reproduce a sinewave will be a bit lower, since it is a smooth curve, but this much headroom will drastically lower distortion. This is why we cannot practically use a class A driver, unless, of course we use one pair of output devices.

For some comparison, below is a HEXFET setup driven by a class A driver at 13mA bias:

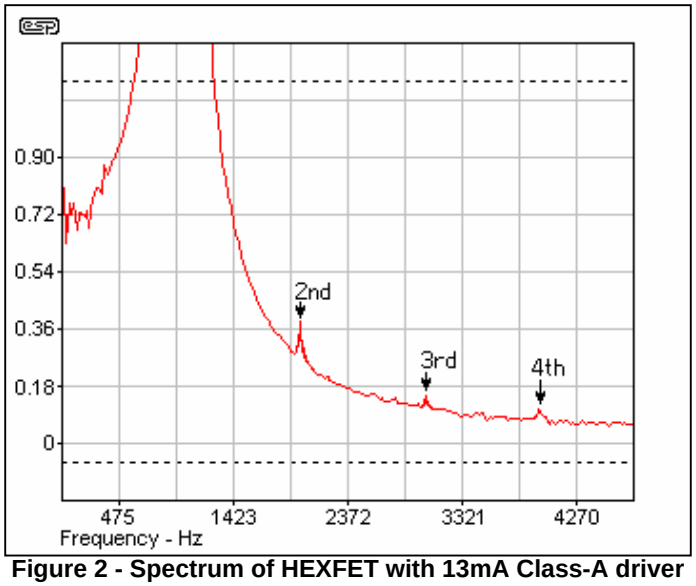


Figure 2 - Spectrum of HEXFET with 13mA Class-A driver

The large notch is at the second harmonic, and the small bump to the right is the fourth harmonic. Barely any third harmonic is seen. This shows 0.25% distortion at the second harmonic at ¾ power, and 250mA amp bias. Not very good for a true hi fi, unless we are making a valve-like amplifier. Even this will not show the same effects as a true valve amp - the nature of the distortion components will almost certainly be different.

Adjusting the bias to 1 amp removes nearly all distortion, yet now we are approaching a heater... I mean class A.

After fixing the problem by adding a class AB driver, distortion was greatly decreased ...

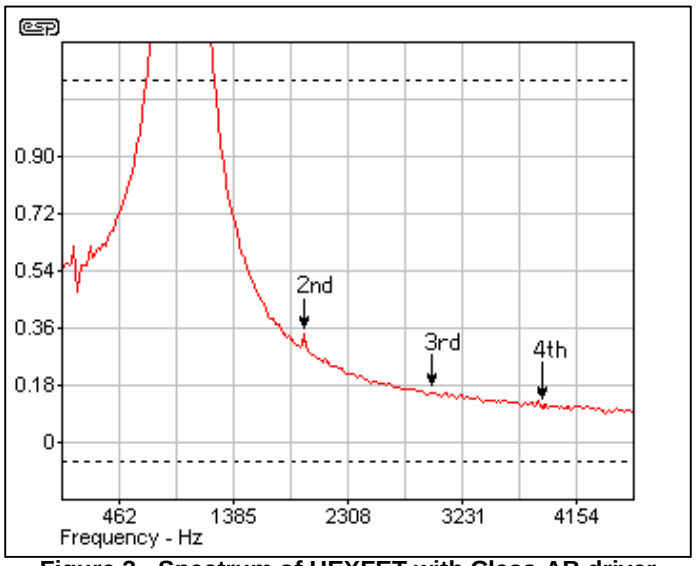


Figure 3 - Spectrum of HEXFET with Class-AB driver

As the picture shows, the second harmonic was reduced considerably, while the fourth harmonic is below the noise floor. This shows 0.04% distortion solely on the second harmonic at ¾ power and still with 250mA bias. This greatly improved the amplifier. At one watt, the distortion is not measurable at all, unlike with the class A driver. Reducing the gate resistors to 4.7 Ohms to get more current through does nothing noticeable, so the use of 10 Ohm resistors is fine. There is no evidence of “notch” distortion or any other nasty odd harmonic, only a “nice” second harmonic added in. Also note that this amp was built on a breadboard. A compact and nicely wired PCB should decrease distortion even further.

Below is the final simplified schematic of the entire amplifier ...

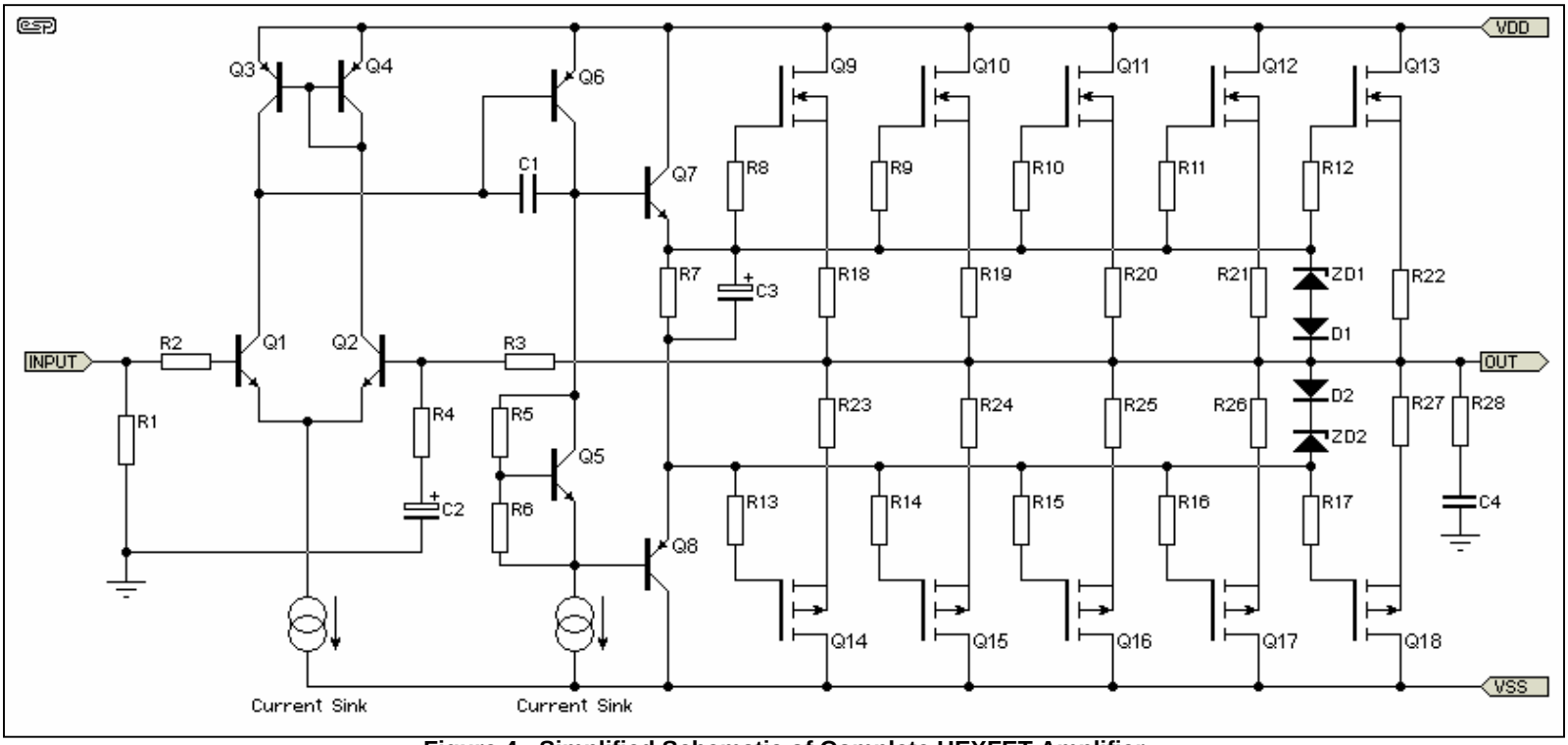


Figure 4 - Simplified Schematic of Complete HEXFET Amplifier

It looks very simple, and includes the class AB driving stage to improve gate driving. Very simple compared to the multiple LTP amps. The minimum stability network (Zobel) shown is always needed, and a series inductor (with parallel resistor) may also be required. The values of these components will be found by experiment.

For some further reductions in distortion, the following work quite well:

1. Design a compact PCB layout with feedback taken as close as possible to the load terminal. We want to measure what the load is doing, not what it should be doing.
2. Try not to run any signal wire close to a power supply wire or the power transformer.
3. Make big traces on the PCB with decent sized spacing between other traces.
4. Make sure your measuring equipment is decently calibrated!
5. For any output device, use some sort of emitter/source resistors (as shown) to help balance the load when parallelling.
6. Use a current mirror with the LTP input to get as much gain as possible to allow for negative feedback, but do not solely rely on the feedback to make the amplifier better.
7. Match the output devices as close as possible, especially for transconductance at low values of V_{GS}.

---

3 - Conclusion

HEXFETs are decent devices once the gates are driven correctly. They are much more rugged than BJTs as my burned parts pile shows, and sound very good when a class AB driver is added. I hope this short article with aid others in using these “switching” and “not linear enough for audio” devices to get distortion figures below many good amplifiers with 'very linear' devices. And remember one thing - *any* output device can be precise if a proper design is found. Finding the correct design parameters becomes more complex with non-linear devices.

---

4 - References

The following are all PDF files, and are direct links to the International Rectifier web site ...

IRFP240 data sheet
IRFP9240 data sheet

AN-936 application note, “The Do's and Don'ts of Using MOS-Gated Transistors”
AN-937 application note, “Gate Drive Characteristics and Requirements for HEXFET® power MOSFETs”
AN-941 application note, “Paralleling HEXFET®® power MOSFETs”
AN-944 application note, ”Use Gate Charge to Design the Gate Drive Circuit for Power MOSFETs and IGBTs"
AN-948 application note, “Linear Power Amplifier Using Complementary HEXFET Power MOSFETs”

... And many pages from “The Audio Pages”

---

5 - Footnote

The above article is a contribution from Mitch Hodges, and ESP has not verified all aspects of the design process described. While the circuit can be (and has been) simulated quite readily with good results, this is no guarantee that everything will work as expected. I added diodes and zeners to protect the MOSFET gates from excessive voltage. It may be possible to select the zeners to achieve basic current limiting, giving the amp some protection from overload conditions. Because of the high gain of HEXFETs, this simple protection scheme will not be particularly effective. Also, remember to that a series inductor may also be required.

It will be noted that there are no component values supplied - this is quite deliberate, and is not an omission. The article is intended to describe the design process and how to work around the inherent non-linearity of vertical MOSFETs, and is not intended to be a construction project. Requests for component values will *not* be fulfilled.

It must be understood that at the suggested current (250mA per MOSFET pair), the total quiescent current will be in the order of 1.25A - at a typical supply voltage of perhaps ±50V, this represents a total quiescent dissipation of 125W! This is a formidable amount of heat to dispose of, and will require very large heatsinks (and/ or forced air cooling). It is probable that the constructor will be forced to compromise, using a significantly lower quiescent current than suggested just to maintain a sensible heatsink size and temperature. Reducing your expectations of the maximum frequency that needs to be passed at full power will reduce the loading on the Class-AB drivers, but does nothing for the MOSFET low current linearity. Compromise will be almost essential (IMHO).

Finally, I'd like to thank Mitch for his contribution, since it describes the issues and how to solve them in an easy to follow manner, keeping complexity to the absolute minimum in the final design example.

Rod Elliott

5.1 - Update (March 2006, ESP)

I have had occasion to build a HEXFET based power amp as a test for a client. While it worked well enough, giving the expected power output and with fairly low distortion, as noted above the required bias current is quite high to reduce crossover distortion to an acceptable figure. While the circuit I built is actually quasi-complementary (using only N-Channel MOSFETs), the basic principles apply regardless.

Bias stability is definitely an issue as discussed above. It is commonly (and erroneously) stated that MOSFETs are 'safe' because they have a positive temperature coefficient, so as a device gets hotter, its drain-source ($R_{DS(on)}$) resistance increases. This much is true, but is actually completely useless in a linear circuit.

What is *not* commonly noted is that all MOSFET devices have a fairly high negative temperature coefficient for the gate-source threshold voltage (V_{th}). At the gate-source voltages needed to obtain typical bias currents, even a small temperature increase causes a large drain-source current increase, so the use of a carefully designed bias servo (Q5, R5 and R6 in the Figure 4 schematic) is absolutely essential. This point is made above, but is sufficiently important that repetition will not go astray.

To illustrate this, Figure 5 shows the graph from the IRF540 data sheet, and although it does not continue down to the levels we are interested in, the trend is clearly visible. At a V_{GS} of 4.5V, we see I_D of around 12A at T_j = 25°C, rising to a little over

20A at T_j = 175°C. While the graph might seem to indicate that the effect will be greatly exaggerated at gate lower voltages and drain-source current, the initial tests that I did indicate that the effect is roughly similar. The graph was taken from the IRF540 data sheet, but has been colour coded to make identification of each graph easier. The use of source resistors to help force current sharing is essential, and these should be as high as practicable. While 0.1 ohms is common for BJT amplifiers, I would recommend values not less than 0.47 ohms for HEXFETs. Higher values provide more stable quiescent current with temperature variations. For example, with 1 ohm source resistors, the current can increase at a maximum of 1mA/mV or 1A/V. Should V_{th} fall by 100mV, Iq can only increase by 100mA. This eases the design of the bias servo.

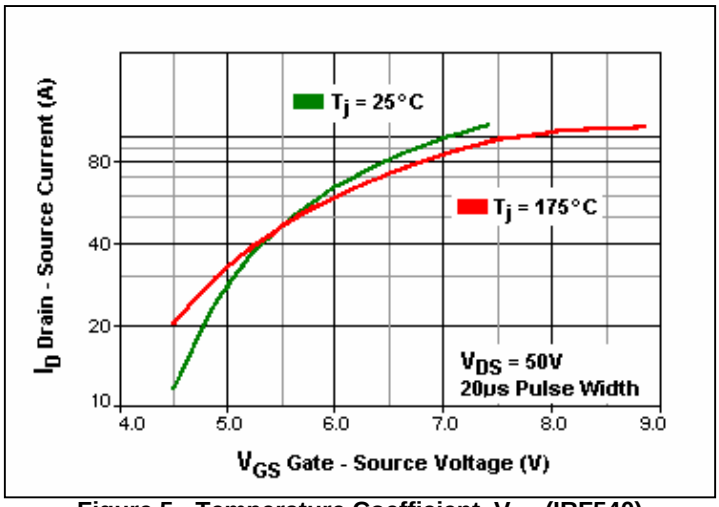


Figure 5 - Temperature Coefficient, V_{GS} (IRF540)

The test I ran was very simple. Apply a suitable voltage to the drain, then carefully adjust the gate voltage until a suitable measurement current was drawn. The current started at a relatively low value (around 1A in my test), and as the device heated up this was seen to rise. It stabilised fairly quickly because the heatsink prevented further (possibly damaging) heat levels, but with two MOSFETs in parallel, the current between them was different, and (more importantly) it remained different, even as they became hotter. The claims for better (and 'automatic') current sharing apply only to devices operated as switches, or where the two curves shown in Figure 5 cross over each other. Only lateral MOSFETs provide a crossover point on their transfer characteristics that is low enough for linear operation.

Something that is missing from nearly all MOSFET data sheets I have looked at is gate threshold voltage vs. temperature (it *is* shown in the data I have for the IRF840). This will show that the threshold voltage falls as T_j increases - a *negative* temperature coefficient. The positive coefficient of $R_{DS(on)}$ is insignificant at the current levels needed for setting quiescent current accurately. Note that the two curves cross over, but the point where the temperature curves cross is when V_{GS} is at a current of over 40A and the gate-source voltage is 5.5V - this is not useful. Lateral MOSFETs (as used in P101) have exactly the same issue, but the curve changes from negative to positive at a much lower current (around 100mA), and this is visible on the transfer characteristic graph (but you will need to look for it carefully - it is not specified in a useful manner IMO).

In an application note [1], OnSemi describe this transition as the 'inflection' point, and it is determined by V_{GS}, although it appears to be related more to the drain current than gate voltage. However, the two are directly related, so the point is moot. Lateral MOSFETs are usually quite safe here, because the inflection point is at such a low voltage and current, but vertical MOSFETs (HEXFETs and similar) are prone to thermal runaway. There is also the possibility of a failure mode very similar to second breakdown when HEXFETs used in linear circuits, where V_{GS} is usually (well) below the inflection point. This must remain a very good reason to stay clear of these devices for audio, unless you are fully aware of the potential risks, and how to avoid them. Note that the article above does not address this potential failure mode, (nor do any others I have seen), and your only choice is to find MOSFETs where the thermal 'changeover' occurs at the lowest possible drain current. Lateral devices are almost unbeatable in this respect.

A careful examination shows that the 'inflection' point is actually the region where the negative temperature coefficient of V_{GS} exactly compensates for the positive temperature coefficient of $R_{DS(on)}$. That's the reason it's at such a high current for HEXFETs (because of the usually low $R_{DS(on)}$ value), and it is known that the inflection point is inversely proportional to R_{DS}

$_{(on)}$. Note that this only applies if the device is used in linear mode. When switching, V_{th} and its temperature coefficient is not relevant because the gate voltage is invariably selected to provide 'hard' switching to minimise losses.

In contrast, a typical lateral MOSFET (such as the 2SK1058) has an $R_{DS(on)}$ of around 1 ohm at 10A, compared to an IRF540N with 0.044 ohms at 16A.

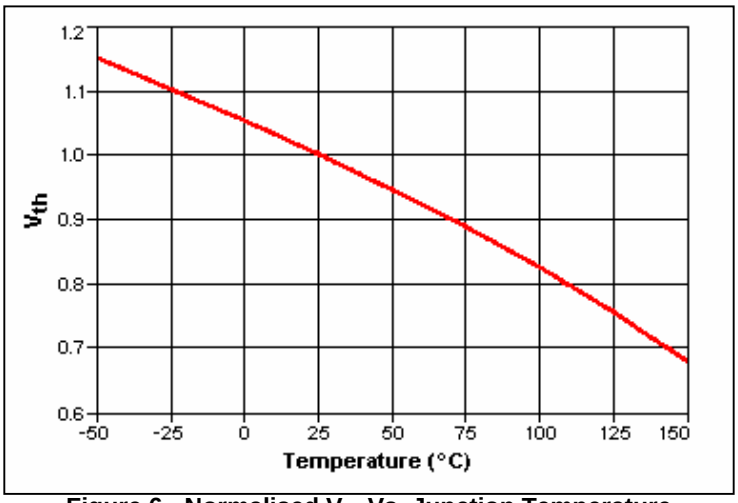


Figure 6 - Normalised V_{th} Vs. Junction Temperature

From Advanced Power Devices, their application note [2] provides the graph shown in Figure 6. Normalising simply means that everything is taken back to a reference of unity, so simply multiply the claimed V_{th} by the figure shown along the left side, for the temperature at which your device will operate. If your MOSFET will undergo a temperature change from 25° to 100°, then V_{th} will fall to 0.83 of the ambient temperature value at 100°C. This application note also mentions the possibility of a failure mode similar to second breakdown when operating switching MOSFETs as linear amplifiers. Indeed, two of the articles listed below refer to the fact that using a HEXFET much bigger than needed (to provide a safety margin) has exactly the *opposite* effect. Rather than increasing the safety margin, the larger device is *more* likely to fail if it working well below the 'inflection' current in a linear circuit.

Do not be mislead by claims that MOSFETs are immune from thermal runaway, although lateral MOSFETs are much better in this respect than vertical MOSFETs (HEXFETs and similar switching devices). Based on the above, it is quite apparent that vertical MOSFETs can easily get into thermal runaway if the bias servo is not set up correctly. Using just a pot (as shown in P101) is absolutely verboten with vertical FETs - a matched bias servo is essential to prevent both thermal runaway and crossover distortion.

Further searching reveals a document from Solid State Optronics [3], where the temperature coefficient for V_{GS} is said to be -1.5mV/°C (the above chart shows it as 1mV/°C for V_{GS} of 4.5V at 25°C). It is claimed to be 'insignificant', and for switching applications this is true. It is definitely not insignificant for a linear circuit (as shown in Figure 4), and especially so because of the relatively high transconductance of vertical MOSFETs. A few (tens of) millivolts of gate voltage is the difference between acceptable quiescent current and overheating.

What of the second breakdown effect? HEXFETs, (and indeed many (most?) other MOSFET devices) are made using a multiplicity of individual small MOSFETs called cells. If the device as a whole exhibits a negative temperature coefficient for V_{GS}, so must each internal cell. If one cell has a slightly lower V_{th} than the others (due to microscopic impurities in the silicon substrate) it will take more of the total current. This will cause it to get hotter, so it will then draw more current, causing it to get still hotter. This process continues until the cell fails due to over temperature, at which point the whole MOSFET is useless.

While this scenario is not common, it does exist, and has caused problems in the past. It will continue to cause problems if designers are unaware that this failure mode even exists - after all, most comments seen describe MOSFETs as almost indestructible. While this is true to an extent, it is now obvious that it is not a general rule upon which one should rely in all circumstances.

Now you know *why* International Rectifier and other vertical MOSFET manufacturers don't recommend HEXFETs or their equivalent for linear applications - they are simply not designed for the purpose. Yes, they most certainly will work, but just be aware of the limitations. I suggest that high voltage, relatively low current devices are preferable to the reverse, as they will have an inherently higher $R_{DS(on)}$, and therefore a lower inflection point.

As it transpires, the design I was looking at is unsuitable for the intended purpose, because the quiescent current needed to remove crossover distortion is too high to be practical. In many cases, the lowest heat output possible is highly desirable, and HEXFETs are simply unsuited to any application where very low Iq is desirable or necessary. Lateral MOSFETs would be fine, but are too expensive for the application (in case anyone was wondering).

---

1 On Semiconductor - AND8199
2 Advanced Power Technology - AN0002
3 Solid State Optronics - Application Note 50

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott and Mitch Hodges, and is Copyright © 2006. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Mitch Hodges) and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allow that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Mitch Hodges and Rod Elliott.

Page created and copyright © 15 Jan 2006./ Updated 13 March 2006 - Added update with thermal characteristic data and additional caveats.

The Audio Pages

Elliott Sound Products **Balanced Line Driver with Floating Output**

Copyright © 2002 - Uwe Beis (Edited by Rod Elliott)
Page Created 30 March 2002

Articles Index

Main Index

Contents

- Introduction
- Why Use Balancing?
- Balanced Line Receivers
- Balanced Line Drivers vs. Floating Balanced Line Drivers
- A Balanced Line Driver with Floating Output
- Performance Measurements
- Simulation Results
- Practical Assembly and Adjustment
- Static Performance
- Dynamic Performance
- Further Improvements - Input Options
- Conclusion
- Justification (Why isn't it perfect?)

Transformerless Balanced Line Driver with Floating Output

Introduction

The article presented is a contributed design by Uwe Beis. Uwe has built and tested the unit as shown, and although more complex than the balanced line drivers and receivers presented in the original versions in the Project Pages (see Project 51 for details), it also has better performance.

However (and this is important), the sensitivity to component tolerance is high, and thermal drift will cause the circuit's characteristics to change. It is very important that Uwe's recommendations are followed closely with regard to the tolerance and type of resistors used. I have simulated the circuit, and also built one using 5% carbon film resistors to see just how far off the circuit would be. Although it worked (after a fashion at least), the balance and frequency response were quite different for each output. After trimming (using a pot, and experimenting until I found the best spot for it), the performance was excellent, but just heating any of the 10k resistors with my fingers was enough to disturb the balance.

Metal film resistors are very much better than carbon in this respect, and are essential in a circuit such as this.

Additional testing and simulations of the circuit have revealed that there is a sensible trade-off that can be made, simply by increasing the values of 2 resistors. Although the modification degrades the "perfect" balance, in reality there is only about a 0.4dB loss of signal when the circuit is connected in unbalanced mode, and this is well within the acceptable range (IMO).

The simplified version requires no adjustments. Mine uses an unbalanced input, but a balanced input can be added in the same way as with Uwe's circuit. Generally, an unbalanced input will be the most common, since if the internal circuitry is already balanced, there is little need to do it again. See Project 87 for all the details.

The following is Uwe's original material with a very small number of changes. Note that I have made minimal corrections to the grammar - Uwe is from Germany, and his English is extremely good. You will still see some evidence of his origins, but to remove all of these would be to lessen the value of the original text (IMHO anyway).

Why Use Balancing?

In the professional audio industry analogue information is transmitted as a balanced signal over symmetrical cables. The big advantage of this method is that interferences of the signals by ground loops can be totally eliminated. To balance device-internal signals, where they are usually unbalanced, transformers are often used. Transformers have another very welcome side effect: They can completely galvanically decouple source and destination, i.e., there is no ground loop they are not just compensated, in fact they do not exist any more. One transformer, either at the output or at the input, is sufficient for that, but professional equipment usually uses transformers at its inputs and outputs.

It is well known to use electronic circuits instead of transformers to balance signals at the outputs or to unbalance them at the inputs. The advantage is that this is cheap, small and works easily from DC up to very high frequencies. One major disadvantage is that they cannot

handle voltages - particularly common mode voltages - greater than their operating voltage. Rod has written some more explanation about balanced signals in his project "Balanced Line Driver".

From here on we should discriminate the difficulties between balanced inputs and outputs.

Balanced Line Receivers

The line receiver Rod describes is a single op-amp one. It is derived from the basic differential amplifier circuit. Its inputs are not totally symmetrical, i.e. signals at the -In "see" a lower impedance than those at +In. Furthermore signals at +In are "visible" at -In while vice versa this is not the case. For higher performance the following circuit is often used ...

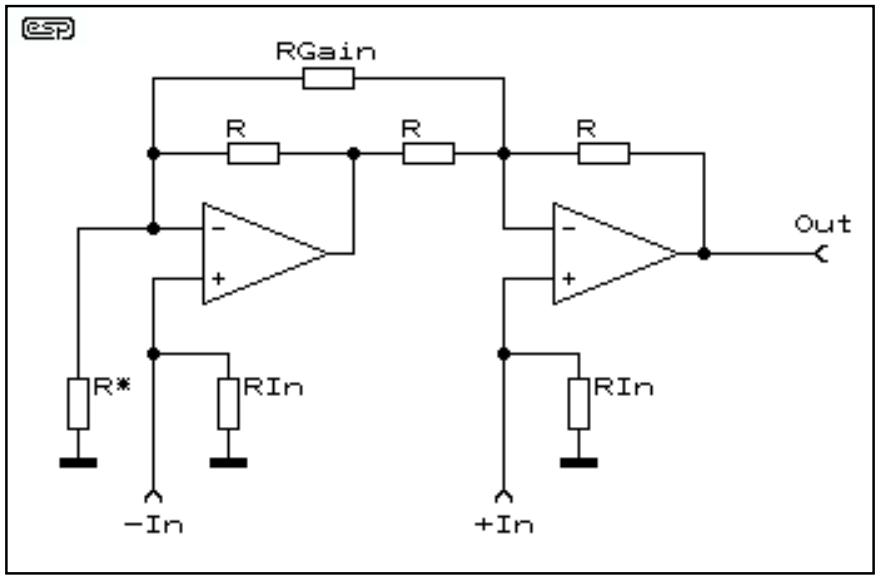


Figure 1 - Improved Balanced Line Receiver

Both inputs may be directly connected to the incoming signal - there are no resistors in between. This improves noise performance so that this circuit is well suited for microphone preamps. Furthermore there is no feed-back from the output to any input so that the inputs may have any input impedance up to nearly infinite. The common mode rejection ratio is determined by the precision of the resistors R and will be about 40dB with 1% resistors. With only one additional resistor RGain the gain can be increased without decreasing the common mode rejection ratio. Or, in other words, the ratio between differential and common mode gain is further improved.

Another improved version can be found in project 66, which is based on a dual transistor differential input.

All in all, balanced line receivers are quite easy to realise with little effort. With the exception of the limited common mode range they easily can come close to the performance of transformer

coupled inputs. They are no challenge - so I do not want to discuss them here further.

Balanced Line Drivers vs. Floating Balanced Line Drivers

The line driver Rod describes in his article is quite simple too. There is nothing to grumble at (;-)), but there is one essential difference between an output like the described one and a transformer coupled one: The common mode output resistance. A common voltage applied to the output will cause a considerable common mode current into the outputs - the output is not floating. Example: The differential output resistance in his project is 2 x 220 Ohm = 440 Ohm. If you apply a common mode voltage U at both outputs, a current of 2 x U / 220 Ohm will flow, i. e., the common mode output resistance is 110 Ohm (1/4 of the differential output resistance).

Or, another and more important point of view: With transformer coupled outputs you may short either output to ground or apply any voltage there - you will still have the full output voltage across its output ports. If you do this on one output of the simple push-pull amplifier it will not affect the other one. You can even connect several transformer coupled outputs in series and by this way sum all output voltages. This, of course, is not possible with a push-pull amplifier output stage which "looks" like a balanced output with a centre tapped earth reference.

I call transformer coupled outputs "floating outputs", because by nature they have a high or infinite common mode output resistance. So I was looking for a transformerless amplifier with floating outputs, and this is where my story starts. (Even though I'm afraid there is nobody out there raising his hand and shouting "Yes that's it! That's what I was always looking for!")

A Balanced Line Driver with Floating Output

I must admit I never really investigated in circuits to be found in literature or in the www, because on one hand I very soon had an idea myself for an appropriate circuit and on the other hand I cannot remember ever having seen anything like this before. I have since seen the schematics for the Analog Devices SSM2142 which uses almost exactly the principle described here and the complement from THAT corporation, the THAT1420/30. I should also mention the Maxim MAX435, a so-called "Wideband Transconductance Amplifier with Differential Output", which is somewhat similar but somewhat different too. At least it has a floating balanced output.

My basic circuit not only has balanced outputs but balanced inputs too and looks like this ...

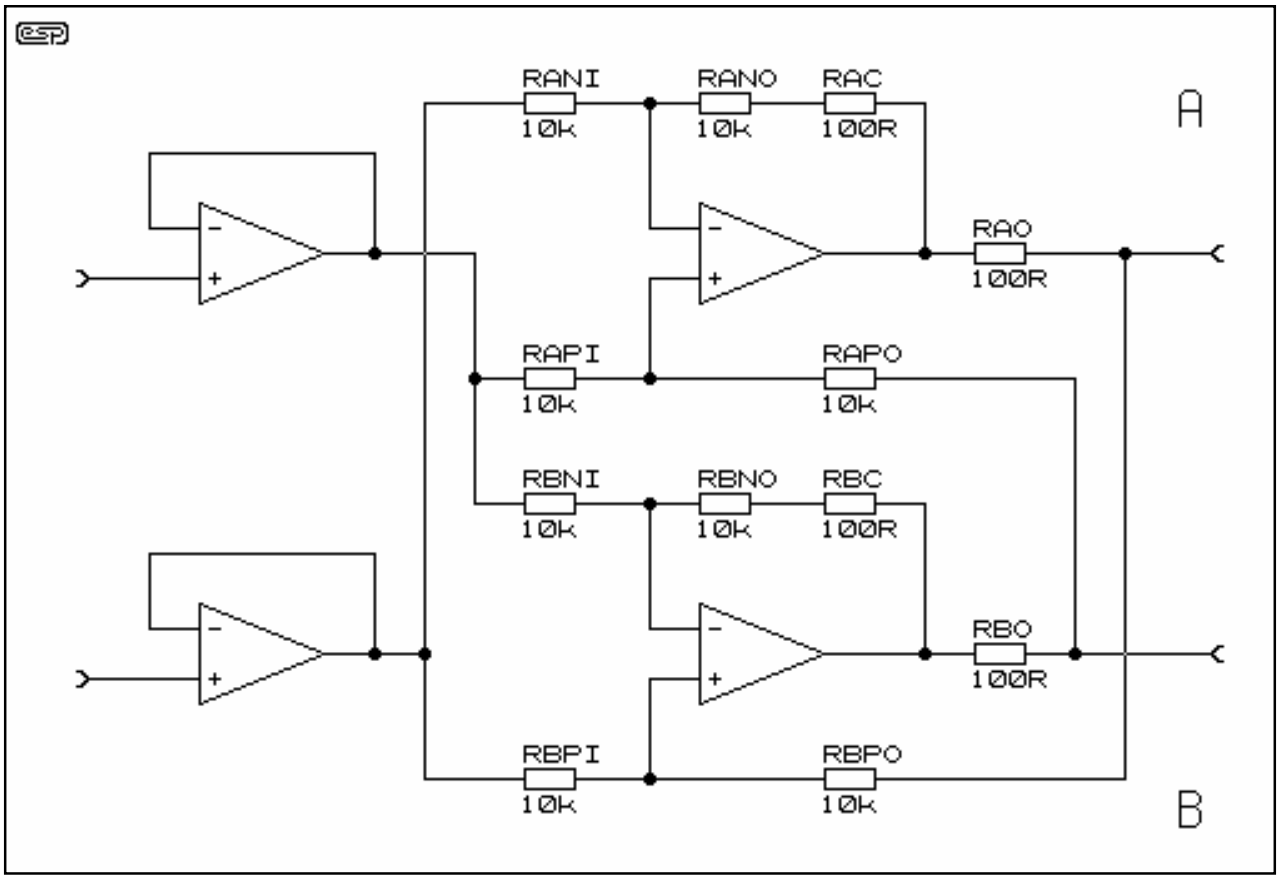


Figure 2 - Basic Form of Balanced Line Driver

Explanation of the indices used:

A, B ($R_{\mathbf{A}NI}$):	A: "Upper" ("positive") resp. B: "lower" ("negative") half
N, P ($R_{A\mathbf{N}I}$):	Connected to the negative (N) resp. positive (P) input
I, O (R_{ANI}):	Connected to the input (I) resp. output (O) op-amps
C ($R_{A\mathbf{C}}$):	Used to compensate R_{AO} resp. R_{BO}

No bias resistors are shown. The input stage is used to buffer the input signal so that the inputs may be dimensioned for almost ideal properties while the actual line drive is supplied with the required low impedance source.

The whole amplifier, as it is dimensioned here, has a gain of 1. The same amount of voltage across the input terminals appears across the output terminals. This remains true if any output terminal is supplied with any voltage - like transformer coupled outputs do (provided both output voltages stay within the supply voltage area of course). This is a consequence of the fact that the common mode output resistance (R_{COut}) is high.

The differential output resistance (R_{DOut}) is low, but not zero. It is determined by the resistors R_{AO} and R_{BO} and equal to it. In fact, $R_{COut}$and R_{DOut} (the output resistance in differential mode) are coupled at each other by the precision of the applied resistors. As R_{DOut}does not need to be very low and R_{COut} should be as high as possible, it seems obvious to start with a reasonable differential mode output resistance ($R_{DOut} = R_{XO}$) of 100 Ohm.

As mentioned, it is obvious that the precision of the resistors must have a certain influence on the performance of the whole circuit. But what is "the performance" of such a circuit?

The primary target obviously is a high R_{COut}at a given R_{DOut}. But at least three side effects must be kept in mind:

A common mode input voltage shall neither be visible at the outputs in common output mode nor in differential output mode (just like with transformers). The "Common to Common Mode Rejection Ratio" (CCMRR) and the "Common to Differential Mode Rejection Ratio" (CDMRR) tells about that. Furthermore a differential input signal shall not be visible at the outputs in common output mode. I call this "Differential to Common Mode Rejection Ratio" (DCMRR).

Of course differential input signals shall be visible at the outputs in differential output mode - this is simply called gain and by design always so close to the desired value that it does not need to be discussed.

Now we know one "primary target" (the common mode output resistance R_{COut}) and three "side effects" (CCMRR, CDMRR and DCMRR) to characterise the performance. I presume two more properties "gain" and "output resistance" as sufficiently precise given by design. I neglect other side effects and all input properties.

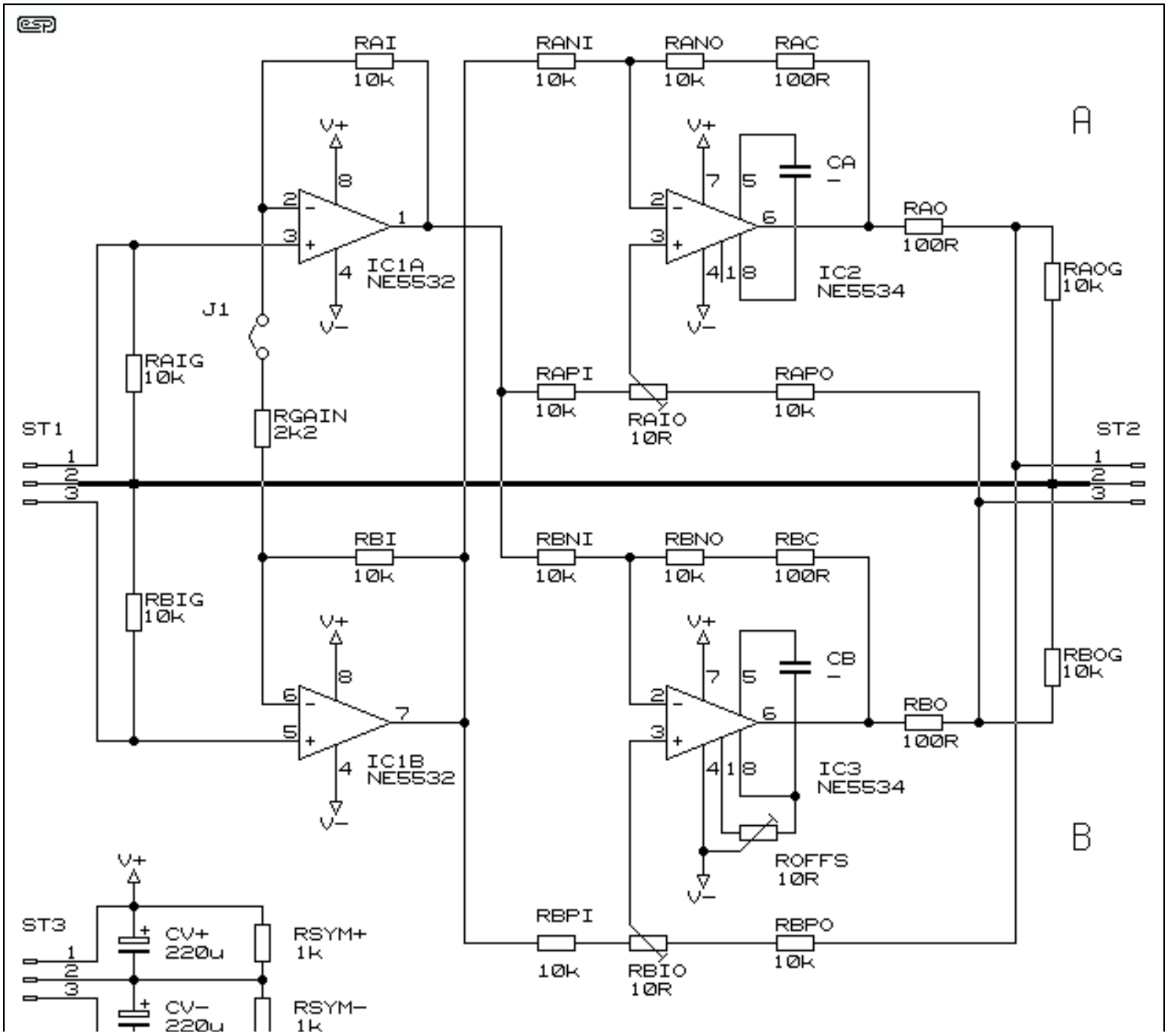
Now we can restart with the perception that it is obvious that the precision of the resistors must have a certain influence on the performance of the whole circuit and to find out this performance.

Performance Measurements

I could have tried to put the whole circuit into formulas and to solve the formulas in a way, that the influence of varying resistors on our target R_{COut} is directly visible, or I could have done this by simulation. Unfortunately my Pspice evaluation version does not support circuits as complex as this one. So at my first approach I preferred to measure the performance in a real circuit. Later I learned more by doing simulations, but I will come back on this later.

For my experimental circuit I thought about how to select the resistors. It seemed obvious that all ratios R_{XYI} / R_{XYO} had to be the same (replace X by A or B and Y by P or N). Furthermore all sums R_{XYI} + R_{XYO} had to be equal. At last all R_{XO} and R_{XC} should be the same.

But you cannot buy parts with equal values, so you have to live with similar ones. I bought 12 resistors with 10 kOhm and 4 with 100 Ohms, all specified with 0.1%. The cost of each is approx. $US0.20. I intended to select them so that the sums R_{XYI} + R_{XYO} are as equal as possible. Fortunately I have a 4 1/2 digit multimeter so that I could measure out 4 pairs of 10 kOhm resistors with difference of the summed resistances less than 2 Ohm. In order to meet the required equality of the ratios R_{XYI} / R_{XYO} I added three 10 Ohm 15-turn trimpots. The result was the following circuit:



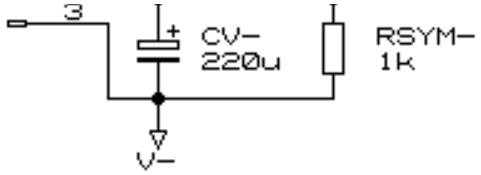


Figure 3 - Balanced Line Driver

Why R_{AOG} and R_{BOG}? With ideal parts the output resistance R_{COut} becomes infinite. With real ones it may become either positive or negative. If R_{COut} is negative, the outputs run to one of the supply rails and remains there - the circuit does not work properly, it is unstable. Thus R_{AOG} and R_{BOG} are introduced: R_{AOG} parallel to R_{BOG} must at least be lower than $-R_{COut}$ to guarantee stability. But R_{AOG} and R_{BOG}also symmetrise the output signal, and the lower R_{XOG} / 2 is compared to R_{COut}, the better the output is balanced. At the end, an extremely high R_{COut} (without R_{XOG}) is only good for a reasonable R_{COut} (including R_{XOG}) combined with reasonable common mode rejection ratios. Which explains why I declare a high R_{COut} as the primary target. Symmetrising the output means that both outputs have to carry the same voltage (but with opposite signs of course) in case they are open.

The trimpot R_{OFFS} is introduced to reduce the remaining common mode DC output current to zero. Particularly without R_{AOG} and R_{BOG} the remaining common mode DC output voltage caused by the current is extremely high, and just "a sharp look" at the trimpot may cause a difference of several volts.

Simulation Results

During our discussions Rod gave me the hint to try SIMetrix Intro as a simulator. SIMetrix was able to simulate this circuit and with this means I discovered an essential relationship between R_{COut}, R_{DOut} and the resistor tolerances:

Have a look at the schematic in Figure 2 and note that both op-amps are connected in a loop. Under ideal conditions the loop gain G_{Loop} is exactly 1. By the simulation I found out that the common mode output resistance R_{COut} can be calculated as $R_{COut} = 2 * R_{DOut} / (1 - G_{Loop})$. As R_{COut} shall be as high as possible, G_{Loop} has to be as close as possible to 1. A loop gain of 1.01 results in a negative R_{COut} of -20 kOhm must be compensated by $R_{AOG} || R_{BOG} < 20$ kOhm to maintain stability, but for balancing reasons $R_{AOG} ||$ Rsub>BOG should be << 20 kOhm. To guarantee $|R_{COut}| >= 20$ kOhm G_{Loop} must be between 0.99 and 1.01. Keep in mind that there are 8 resistors and their deviations that influence G_{Loop}! That looks difficult, but on the other

hand a Monte Carlo analysis teaches that with resistor tolerances of 1% (which theoretically could result in loop gains between 0,99 ^ 8 = 0,92 and 1.01 ^ 8 = 1,08) $|R_{COut}|$ will be > 20 kOhm (2 * R_{DOut} * G_{Loop}) as desired in more than 99% of all cases. Under these circumstances I would recommend R_{AOG} = R_{BOG} = 4.7 kOhm, which is a fair compromise between the resulting symmetry and the common mode output resistance.

Practical Assembly and Adjustment

The experimental board

I built the circuit in my favourite way on a Vero square pad board. I observe special rules for that and I would like to write more about it, but not here. I am quite convinced it looks professionally nice - on both sides of the board.

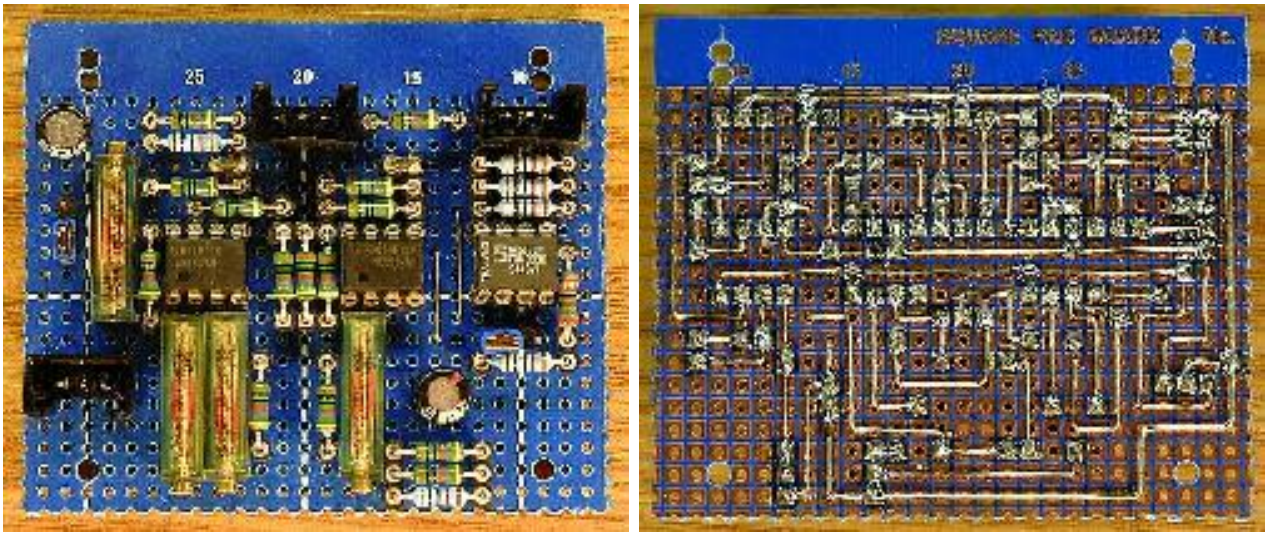


Figure 4 - Front and Rear of Veroboard Layouts

In order to be easy to change, all resistors and op-amps are pluggable.

The power supply is symmetrical. As almost no current flows out of and into the ground rail, a single, floating supply may be used with the ground rail artificially generated by a resistor divider of 2 x 1 kOhm. Actually none of the I/O- and power connector's ground pins needs to be connected externally.

The gain of the input stage may be raised with an additional resistor R_{GAIN}, similar to the resistor R_{Gain} above. The gain of the output stage may additionally be altered by varying the ratios of R_{XYI} / R_{XYO}.

In order to adjust the ratios R_{XYI} / R_{XYO} I unplugged all op-amps, connected the (former) outputs of the input op-amps (IC1, pins 1 and 7) together to V- and the (former) outputs of the

output op-amps (IC2 and 3, pins 5) together to V+. R_{AG} and R_{BG} must be removed. With the multimeter the voltage between the "untrimmed" op-amp (IC2, pin 2) input and the three trimmed op-amp inputs (IC 2, pin 3 and IC 3, pins 2 and 3) now can be minimised so that the balance between the dividers is extremely high - a resolution of 10 uV from 10V corresponds to 10E-6 or 0.0001%! But don't ask me about long term or temperature stability.

Static Performance

In order to find out more about the consequences of resistor deviations, I paralleled a resistor to 1.: R_{ANI}, 2.: R_{ANO}, 3.: R_{ANI} + R_{ANO} and 4.: R_{AO}. The resistances were reduced by 1%.

The table shows the results without R_{AG} and R_{BG}. Without these resistors I_{COut} rather than U_{COut} for or CCMRR and DCMRR have to be measured.

	Name	Formula	Units	Nominal	-1% R_{ANI}	-1% R_{ANO}	-1% (R_{ANI} + R_{ANO})	-1% R_{AO}
R_{COut}	Common Mode Output Resistance excl. R_{XOG}	U_{COut}/I_{COut}	Ohm	>> 10 M	~ -20 k	~ +20 k	-1.2 M	-2.5 M
R_{COut}	Common Mode Output Resistance incl. R_{XOG}	U_{COut}/I_{COut}	Ohm	5 k	~ 4 k	~ 6 k	5 k	5 k
CCMRR	Common to Common Mode Rejection Ratio	U_{CIn}/U_{COut}	-	> 1000 (60 dB)	~ -3 (10 dB	~ +5 (14 dB)	-120 (42 dB)	-500 (54 dB)
CDMRR	Common to Differential Mode Rejection Ratio	U_{CIn}/U_{DOut}	-	> 1E4 (80 dB)	~ 300 (50 dB)	~ 550 (55 dB)	> 1E4 (80 dB)	> 1E4 (80 dB)
DCMRR	Differential to Common Mode Rejection Ratio	U_{DIn}/U_{COut}	-	~ +400 (52 dB)	~ +6 (16 dB)	~ -10 (20 dB)	~ +140 (43 dB)	~ +160 (44 dB)

The common mode output resistance excl. R_{XOG} in this experimental circuit is extremely high. But only one of the resistors R_{XYI} or R_{XYO} changed by 1% reduces the common mode output

resistance down to +/-20 kOhm, which, in my opinion, is hardly at the edge of being acceptable. During my experiments it seemed plausible to me that with an 1% error the output resistance could not be very different from 100 or 200 times the output resistance. Later, with the simulations, I learned to know the exact relation of 200 for 1% error.

Altering either R_{ANI} or R_{ANO} imbalances the divider network at the op-amps input. Altering both by the same amount causes no imbalance here, but somewhere else where it is obviously much less critical. The same applies to an alteration of R_{AO} and R_{BO}.

Dynamic Performance

The frequency response of this amplifier reaches from DC far beyond the need of audio circuits. As two op-amps are partially connected in a loop, a danger of self-oscillation arises. In fact, with a square wave input signal some ringing can be observed at a few MHz. With an input frequency around a few MHz the amplifier becomes unstable. The NE5534 as a single op-amp is meant to be compensated by an external capacitor (C_A and C_B, ~22 pF each). Using these does not improve the situation in this special circuit: The frequency of the ringing not only lowered but it is also "un-damped", i.e. the ringing lasts for more periods. It turned out to be best to omit both capacitors and to care that higher frequencies are blocked out. I did not notice any danger of self-oscillation with various resistive or capacitive loads.

In my previous text I wrote about output resistances R rather than output impedances Z because I primarily looked at static (DC) values, i.e. the common mode output resistance R_{COut} above, which is measured as a DC value. Of course in reality it is a complex value with approx. 2 nF in parallel, so it becomes capacitive at higher frequencies. With output resistors R_{AG} and R_{BG} of 10 kOhm each, the output impedance up to 10 kHz is predominantly resistive (5 kOhm). With slower op-amps, i.e. TL072, the output impedance will remain resistive up to a few kHz only. I did not care for the dynamic properties of the rejection ratios.

Further Improvements - Input Options

The inputs of my sample circuit above are equipped with one simple buffer amplifier each. Later (unfortunately too late) I realised that I missed much room for major improvements:

Remember: R_{XOG} have been introduced not only to guarantee stability but also to symmetrise the output signal. For this reason I dimensioned R_{XOG} extra low. Stability can also be guaranteed by lowering the loop gain, i.e. by just omitting both R_{XC} (resulting in a loop gain of 0.98 typically). But just omitting both R_{XC} and both R_{XOG} results in very poor common mode

rejection ratios, i.e., an input signal applied at one input only will appear at the corresponding output only - totally unbalanced.

If both input op-amps are connected as differential amplifiers with complementary outputs, the following output stage will already be supplied with a balanced input signal and needs not to symmetrise the signal any more:

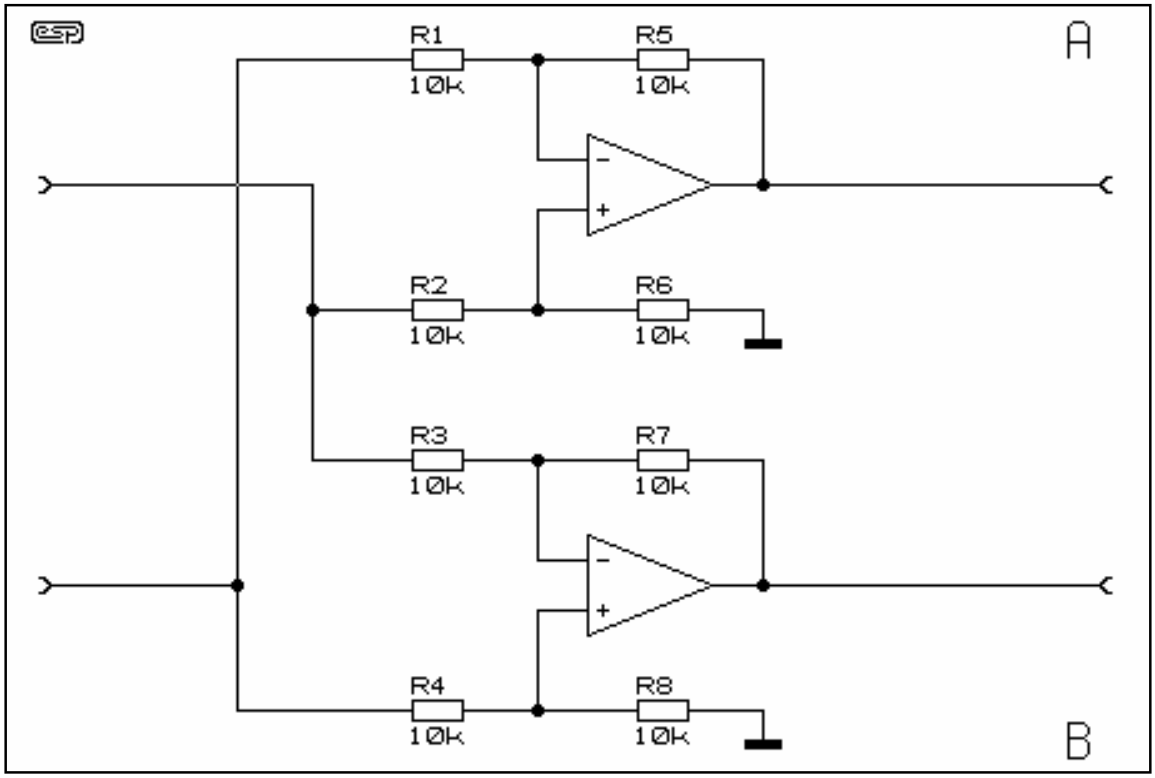


Figure 5 - Balancing Differential Preamplifier

Alternatively, if you have an unbalanced source, you may simplify the design by just inverting the input signal with a single op-amp. Omit R_{XC} and $R_{XOG}$and you have the circuit Rod has chosen in his project 87 and Analog Devices in the SSM2142 too.

But in both of these alternatives there is a snag: A reverse feedthrough from the outputs to the input(s). Therefore either a low impedance source should be used, or another op-amp connected as a buffer should precede the inverter(s). This will finally lead to "ultimate perfection" ;-)

Conclusion

Without the latter input improvement, a common mode output resistance R_{COut} of 5 kOhm, determined by R_{AG} in parallel with R_{BG}, is reasonable, but with resistor tolerances of 1% the asymmetry may become quite high. It is not likely that the amplifier becomes unstable. I prefer

to use either 0.1% resistors, or the resistors should be measured out so that R_{XYI} / R_{XYO} is quite equal or the dividers should be trimmed. With multiple of these measures R_{COut} may easily be risen far above 1 MOhm typically.

With the input improvement, R_{XC} and R_{XOG} should be omitted. By using 1% resistors R_{COut} in 99% of all cases will become between 7 and 20 kOhm, combined with very good CMRRs. With 0.1% resistors and both R_{XC} = 90 Ohm R_{COut} in 99% of all cases will even become between 70 and 200 kOhm.

But do not forget to trim at least one of the op-amps offset voltages: 1 mV offset voltage will cause 400 mV common mode offset voltage if R_{COut} is 10 kOhm or even 4 V if R_{COut} is 100 kOhm (!). In order to reduce the differential offset voltage too, the offset voltages of both op-amps must be trimmed. But this seems to me to be less important.

Justification (Why isn't it perfect?)

This project is a small, few-weekend project which does not claim to be theoretically, mathematically or scientifically correct and complete. I may have made errors, I may have overlooked better solutions. I was lucky that Rod gave the hint to do the simulation with SIMetrix Intro - as my PSpice Student was too much limited. With this simulation I learned a lot about this circuit.

Rod cared for the circuit to an extent that I really appreciated very much. He "pushed" me to have closer looks to what I was actually doing, to make simulations and lots of things that finally resulted in the current state. Without him it all would look much more amateurish. Rod, I really have to thank you. You are a great editor!

Articles Index

ain Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Uwe Beis and Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Uwe Beis) and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Uwe Beis and Rod Elliott.

Page created and copyright © 30 Mar 2002

Elliott Sound Products — Using Current Sources, Sinks & Mirrors In Audio

Rod Elliott

Last Updated 28 Jan 2000

ESP Home Page

Articles Index

Contents

- Introduction
- Current Sources (Sinks)
- Common Emitter Amplifier
- Local Feedback
- The Current Source As A Load
- Real-Life Current Sources
- Bootstrap Current Source
- Circuit Applications
- Current Mirrors
- Differential Pair (Long Tailed Pair)
- Conclusion

Introduction

Integral to most audio amplification systems, the current source and current sink are indispensable. For the purposes of this article, sources and sinks shall be treated as being the same thing - which they are, depending upon how one views the circuit. In essence, the functionality (and purpose) are identical - force the current through another device to be constant, as long as the voltage is within the boundaries of the power supply (less some voltage drop across the source / sink).

Current mirrors have the ability to greatly increase open loop gain and linearity, giving better closed loop performance. This short article looks at the various types, from the theoretical "perfect" device, through various practical examples.

The Perfect Current Source

A perfect current source can be constructed very simply, requiring only an infinite voltage and an infinite resistance. Unfortunately, this will also provide an infinite current (which will be infinitely constant), but is (infinitely?) excessive for most applications.

Sorry, I was being facetious, but the principle is nonetheless sound. In reality, a sufficiently high voltage can produce a current through a varying resistance that is essentially constant, provided that the voltage swing on the resistance is a small fraction (<100th) of the supply voltage. This will provide an accuracy of 1% at 100:1 voltage ratio.

This is a method that was common years ago, and I still have (and use) a transistor tester that uses this method to provide a constant base current to the device under test. With a supply voltage of 250V feeding the base-emitter junction (0.65V typically) via switched resistances, the current is not affected at all by any variation in the base-emitter voltage under load.

The principle works well, but no-one wants to have to use such high voltages in equipment these days.

So why mention this method at all? Simply because the operation and characteristics of current sources are not well understood by most non-engineers (and quite a few engineers, too). Figure 1 shows a resistive current source that will maintain the set current of 1mA to an accuracy of within than 1% over the range 0-10V.

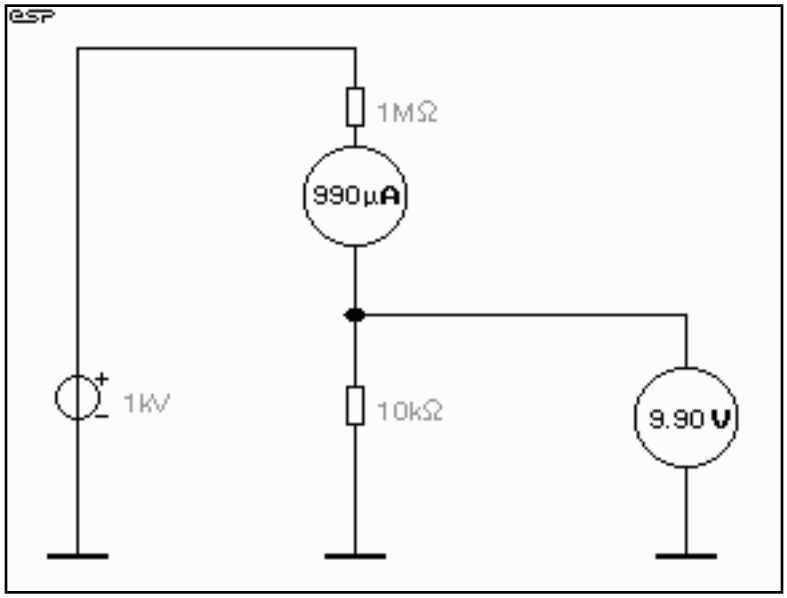


Figure 1 - Simple Resistive Current Source

The first thing that we need to be aware of is that the impedance is very high, and a perfect current source (as described above) does indeed have an infinite impedance. Although this is impossible to achieve in practice, impedances of many, many megohms are quite possible, and are used regularly in many different circuit types.

The Common Emitter Transistor Amplifier

If we look at a simple common emitter transistor amplifier (Figure 2), the output impedance is equal to the collector resistance, and the gain is equal to the collector resistance divided by r_e - the transistor's internal emitter resistance. This is equal to approximately

$$26 / I_e \text{ (mA)}$$

and will naturally vary with the current. Please note that biasing components have been omitted for clarity, but in reality they must be used.

For all examples, a transistor with a current gain of 100 is used and the bias current will be set at exactly 100th of the expected collector current (using a current source, naturally).

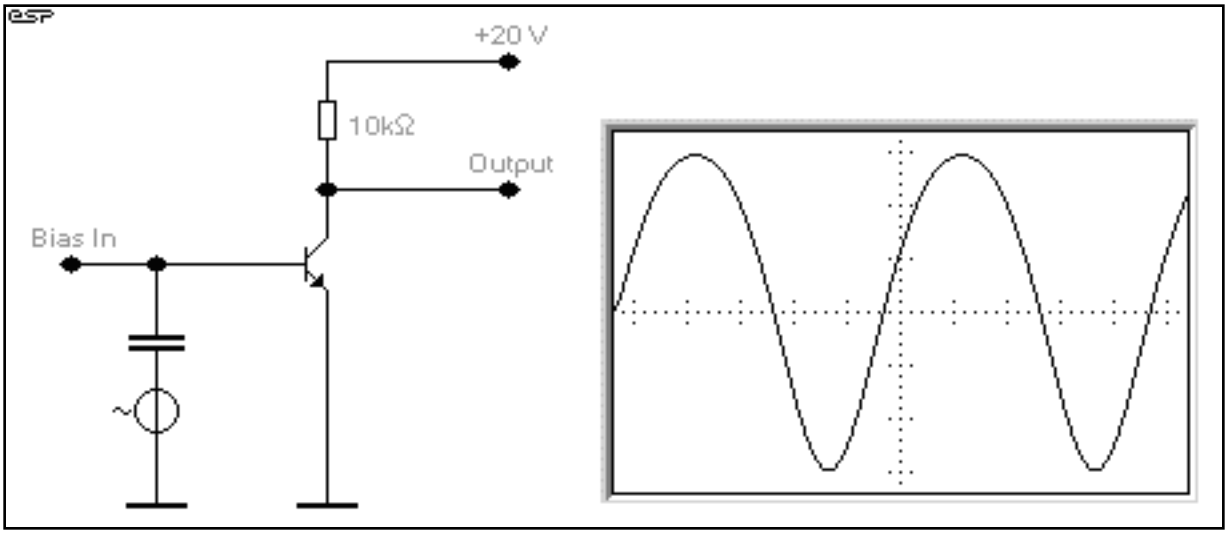


Figure 2 - Simple Common Emitter Transistor Amplifier

At low output levels, the gain should be 385, but as simulated is actually 397. This is well within the degree of accuracy we want to go to in this article, so practice correlates well with theory (one would hope so, or we would all be in trouble).

As the transistors output voltage increases, the available collector current (and hence emitter current) falls, so r_e increases (and vice versa). The higher the output voltage (relative to the supply voltage), the less emitter current flows, and the lower the gain becomes. As the transistor turns on (reducing the output voltage), the exact opposite occurs, and the gain rises. This effect is immediately evident in the output waveform, which is distorted. This is shown in the oscilloscope trace in Figure 2, and the distortion is clearly visible, with the tops of the waveform flattened out, and the bottoms "stretched". This is typical of any transistor amplifier operating without feedback, and is also true of valves and FETs.

Local Feedback

One method of reducing this distortion is to introduce a resistance into the emitter circuit (local feedback) to reduce the gain and swamp the variations in r_e. This works, but we still have a relatively low output impedance, and distortion is still evident - although much reduced. This is shown in Figure 3, where a 100 Ohm emitter resistor has been added.

A commonly held belief is that local emitter feedback reduces the output impedance - this is simply not true - the output impedance of a common emitter amplifier is equal to the collector resistance, regardless of emitter degeneration. To reduce the output impedance, it is necessary to apply feedback around the entire circuit - from the output back to the input. The proof of this is outside the scope of this article, but it is nonetheless a fact.

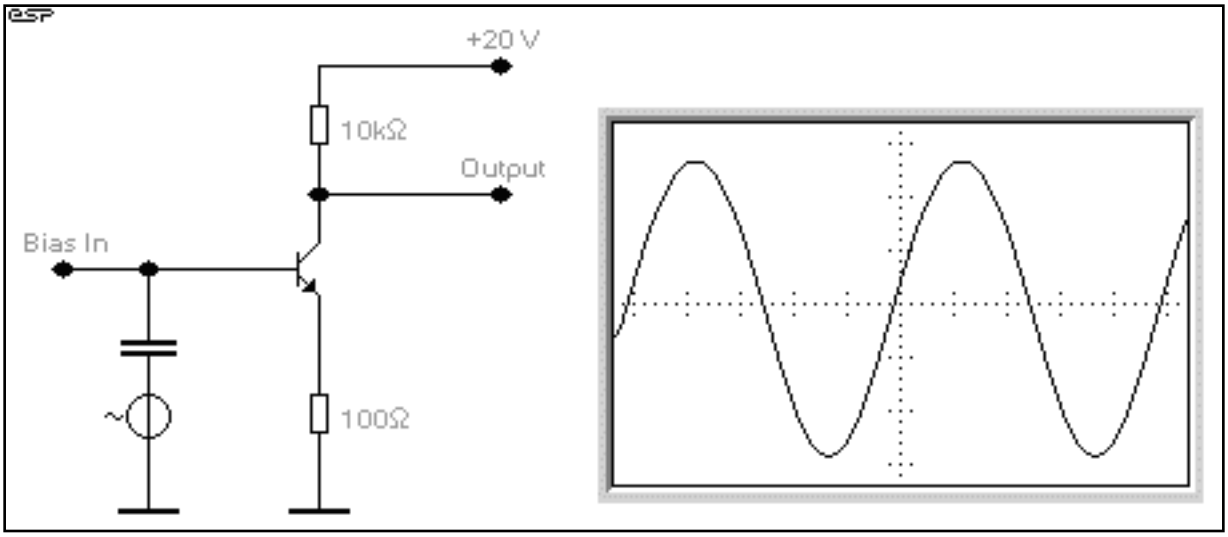


Figure 3 - Common Emitter Stage With Local Feedback

With all this talk of distortion, I can almost hear all the thermionic valve (tube) enthusiasts cheering "I knew it - and here is the proof I have been looking for!"

Sadly, this is not the case. Valves do exactly the same thing, as do FETs of all types - the mechanism is different in each, but the result is the same. Gain varies with current in all known amplifying devices to a greater or lesser degree, so the principles outlined here are valid for any device. I am staying with transistors because they are the most commonly used of all the devices currently available.

In this case, the gain of the stage works out to be just over 78, and theory would predict a gain of 79, so again the correlation is very good. Remember that with an emitter current of 1.001 mA (collector current plus base current), the value of r_e is about 26, and this must be added to the external resistance R_e.

Use of a Current Source Load

Since we have determined that the load resistance affects gain, and that variations of collector current (and hence emitter current) affect linearity, if a current source were to be used instead of the 10k resistor, we should get vastly more gain, and the distortion should be negligible.

In reality, this is exactly what will happen, and Figure 4 shows a theoretical current source as the collector load. The current source is set to provide a current of 10 mA as before. No oscilloscope trace is shown, because the linearity is so good that no distortion is visible, and a trace of a sinewave gets tedious after a while.

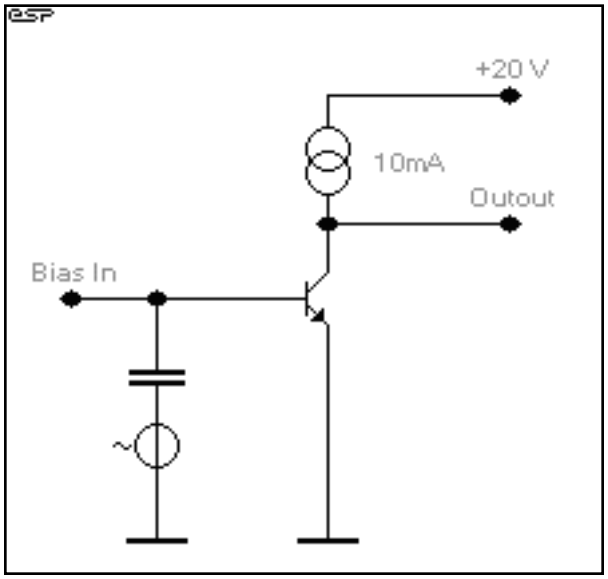


Figure 4 - Use of Current Source as Load

Figure 4 shows the circuit, which (theoretically) has a gain of about 20,000,000 (and no, there are not too many zeros, it really is 20 Million). Naturally, real life will reduce this considerably - even using a 10M Ohm oscilloscope probe will cause a dramatic loss (down to about 390,000 in fact). In reality, the gain of this circuit is actually infinite, but the "meters" I used in the simulation are "only" 1G Ohm each, which reduces the gain.

Figure 5 shows the same circuit with a real current source, and the gain from this (again, the transistors all have an Hfe of 100) is still quite astonishing. We still manage a gain of 19 Million (bearing in mind the meter load, which reduces the gain).

This form of circuit can no longer be biased using a simple current source, because the circuit gain is so high that even the smallest variation in transistor Hfe, or the tiniest variation in a resistor value will cause the output to "go to ground" or sit at the supply voltage. Consequently, an opamp has been added to apply bias, otherwise the circuit is almost impossible to stabilise. The use of a bias servo in this manner is not too common, but I have seen it used in a number of applications. Care needs to be taken to ensure that a low frequency phase shift oscillator is not created if you want to experiment with this idea.

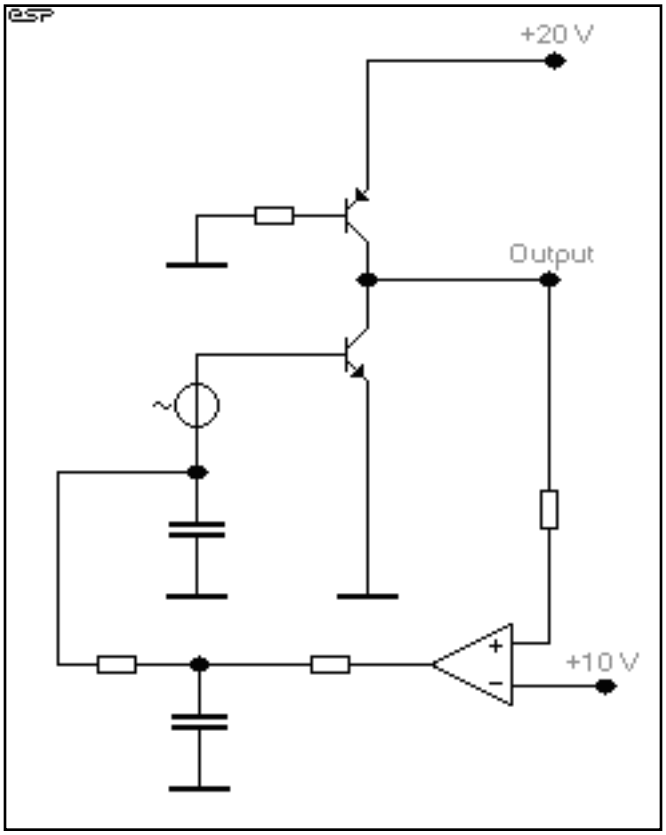


Figure 5 - Real-Life Current Source Load With Bias Servo

For a current source of this type, the current is equal to

```
I = (( V - 0.65 )  / R ) * Hfe
```

The current source shown is workable, but is highly dependent upon the gain of the transistor, and the supply voltage. Now that the basic principles have been examined, we shall look at some real current sources (or sinks, depending on how you look at it).

Where **I** is the current, **V** is the supply voltage, **0.65** is the emitter-base voltage (typical), **R** is the base resistor and **Hfe** is the forward current gain of the transistor.

Real-Life Current Sources

The "classic" current source is shown in Figure 6, and is reasonably stable with transistor gains and temperature. In audio work, absolute stability is not really necessary, since the feedback loop is used to ensure that bias voltages are maintained at the correct value. In some cases a servo circuit may be used (as shown in Figures 6 and 7), but this is not really necessary in the majority of cases.

The circuit shown uses the diode stabilised base voltage to maintain a constant voltage across the emitter resistor. As always, a constant voltage across a resistor means that the current is

also constant. The output current of this source is calculated from

```
I = 0.65 / RE
```

Where **RE** is the total emitter resistance - the internal r_e of the transistor plus the external resistance R_e. The diode current should be a minimum of 10 times the expected base current, but can be up to 100 times as great.

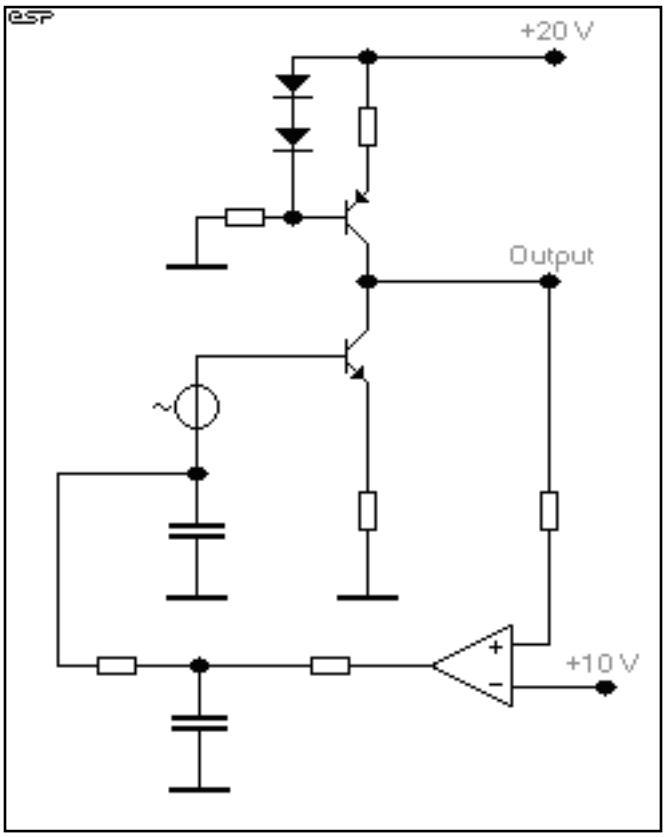


Figure 6 - The "Classic" Current Source

The gain is relatively unaffected by the change in the configuration of the current source, and the next variation (which is preferred by many designers - including me) has no significant effect either.

In any of the above cases, the inclusion of an emitter resistor (100 Ohms - as shown in Figure 6) reduces the gain to about 3.75 Million - this is still an outstanding amount of gain from a single transistor!

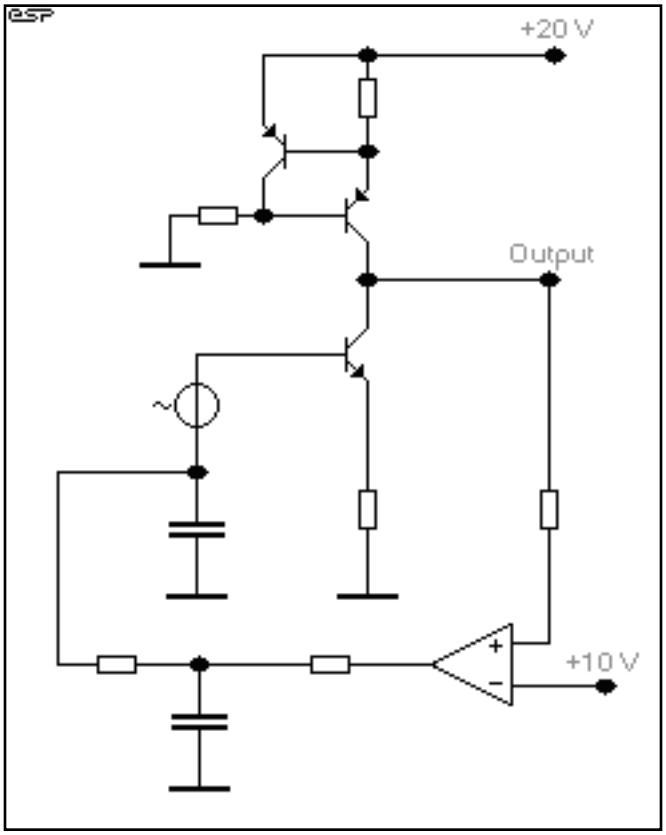


Figure 7 - The Preferred Current Source

This current source is probably one of the most commonly used, and has excellent stability, especially if the two transistors are thermally bonded. This is not needed for most applications, since as mentioned previously, absolute current accuracy / stability is not (or should not) be a requirement for an audio amplifier.

This is a very simple circuit, but will give excellent results wherever it is used. Current is calculated by

$$I = 0.65 / Re$$

In this case, the internal emitter resistance is not an issue, since the second transistor allows only that exact amount of base current for the current source to develop 0.65 Volt across the resistor. Should the current try to increase, the second transistor simply turns on harder, and steals more of the base current. The current through this second transistor should be at least 10 times as great as the expected base current of the current source.

With all the circuits shown, output impedance is very high, and by my tests, is about 500 Megohms (this happens to the the combined resistance of the two meters I used for monitoring, so the actual amplifier output impedance is effectively infinite. Inclusion of the emitter resistor (providing local feedback) has no effect whatsoever on the impedance (contrary to popular belief), but simply reduces the gain.

Bootstrap Current Source

The final (and least understood) of all the current sources is the so-called "bootstrap" amplifier. While not as good as a true current source, its performance is still very good, and in the example shown in Figure 8 manages to provide a gain of 8,860. Bear in mind that this is at a current of only 1mA (the others all used 10mA), so the gain must be reduced somewhat. However, this circuit is not in the same league as the gain of millions from the current source, but in real terms this is more than enough for the Class-A amplifier / driver stage of a power amplifier.

There are some advantages and disadvantages to this circuit, and it is dependent on the application whether it will be suitable. Where a true current source has the same (more or less) impedance at all frequencies right down to DC, the bootstrap circuit has a low frequency cut-off point determined by the resistance and capacitance used. This makes DC biasing easier (since the gain is quite low at DC), but since it reduces gain (and global feedback) at DC, there may be more DC offset at the amplifier's output as a result.

Distortion will also increase at very low frequencies, since the constant voltage across resistor R2 will not be maintained as phase shift (and LF rolloff) start to take their toll.

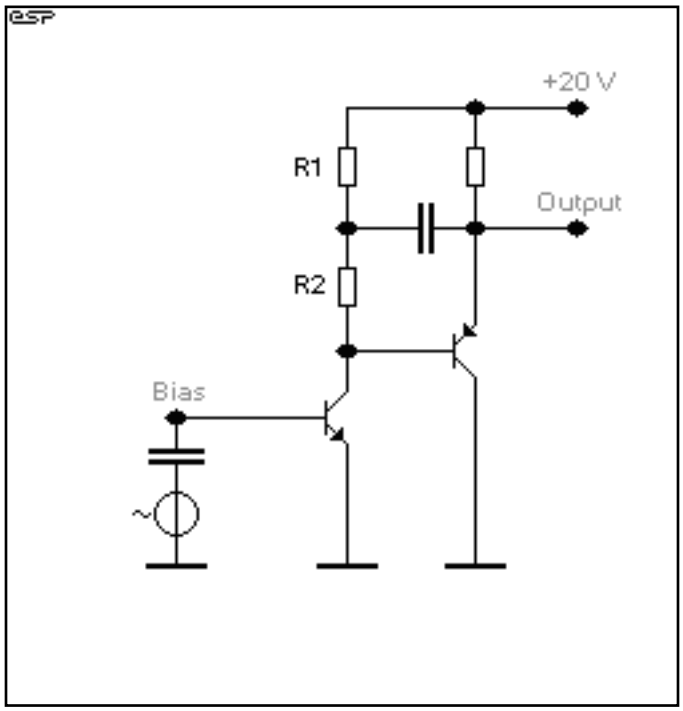


Figure 8 - Bootstrap Current Source

Since it is so misunderstood, it is worth a few words of explanation. R1 and R2 supply a nominal current of 1 mA (resistors are 5k each). Under normal (i.e. quiescent) conditions, there will be 5 Volts across each resistor. As the output varies, the capacitor couples the output signal back to the centre tap of the two resistors, and if the emitter follower were a

perfect buffer amplifier, this would exactly equal the input. As a result, the voltage across R2 is kept constant - if the voltage across a resistor is constant, then it follows that the current through the resistor must also be constant. The current through R1 varies considerably, but R2 maintains a constant current over the full output operating range.

Small differences in voltage (due to the loss through the emitter follower) or in phase (due to the capacitive reactance starting to become significant with decreasing frequency) both cause the bootstrap current source to reduce its effectiveness, so impedance and gain drop, and distortion increases. In a well designed circuit, these effects are minimal at all normal operating frequencies, but cannot be discounted.

Practical Applications

Now that we have looked at the various types of current source (there are many more, but the others will not be examined here), we can examine where they are best used, and why.

I am mainly interested in audio amplification, and will therefore concentrate on this aspect of the use of current sources and mirrors. There are naturally countless other applications, but I will leave it to the reader's imagination to explore the additional uses.

One of the most irksome loads for any amplifier is one that has inherent non-linearities, and the output stage of an audio amp definitely fits this description. Consider the worst case, where there is no forward bias applied to the output devices, which are operating in Class-B. The global feedback in an amplifier is generally considered to the be main saviour, by reducing the distortion components by using the input stage as an error amplifier.

This is actually only a small part of the story - a properly designed Class-A amplifier will have both high gain and high output impedance, both of which are useful in reducing the otherwise inevitable crossover distortion. Consider Figure 9a and 9b, showing the same output stage delivering the same peak output voltage to the same resistance.

The first is using a low impedance amplifier (an opamp) with unity gain as shown, but will normally be adjusted to give the required output, while the second uses a high impedance source having a high gain and impedance, but still providing the same output voltage.

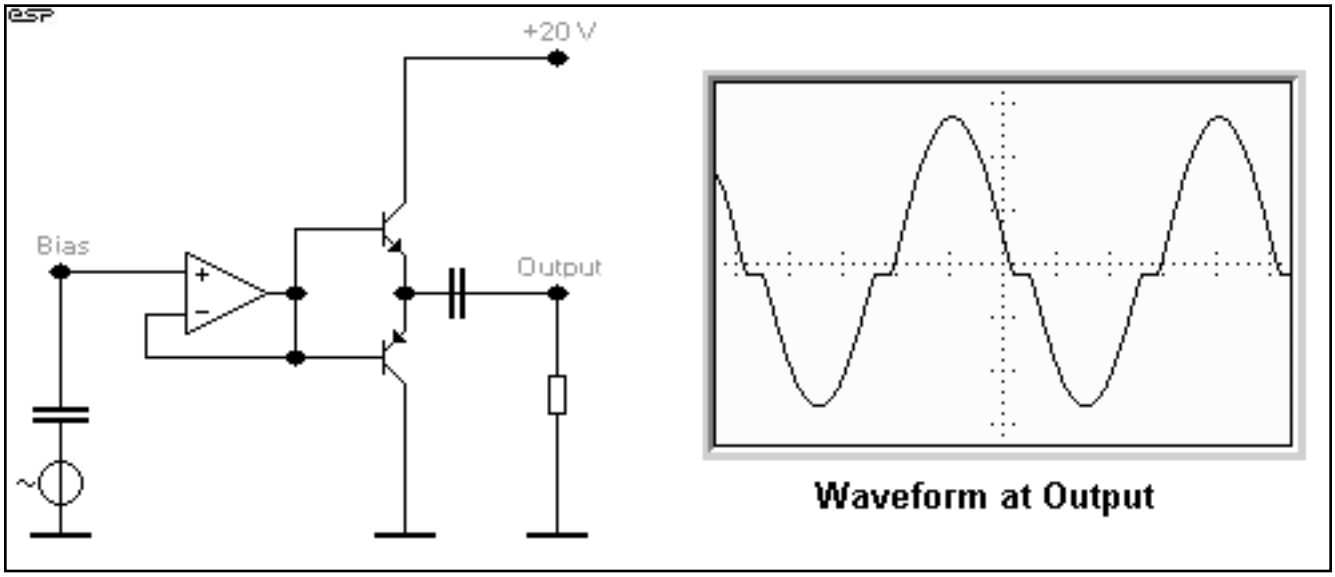


Figure 9a - Use Of A Voltage Source Driving an Unbiased Output Stage

The load for each amplifier is 100 Ohms, and it is clear from the first figure (9a) that crossover distortion is gross. Figure 9b shows how the high impedance drive to the output stage overcomes the crossover distortion - the output waveform is not shown, because there is virtually no distortion visible in the oscilloscope trace, so it looks just like a sinewave.

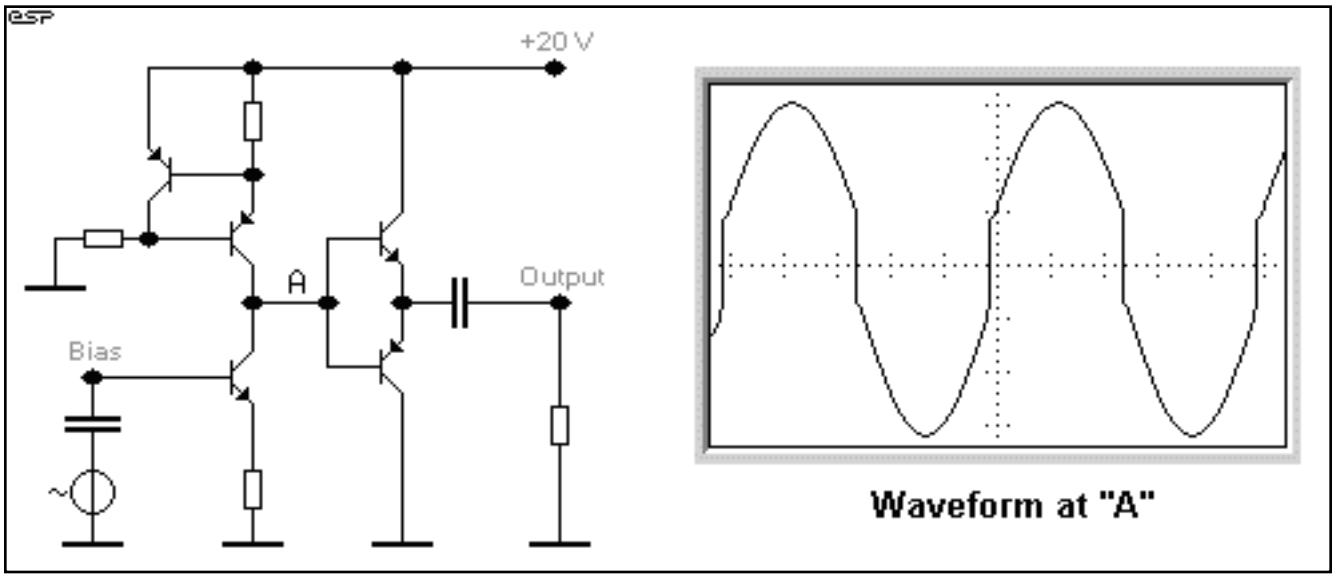


Figure 9b - Use Of A Current Source Driving an Unbiased Output Stage

Remember that this is with no feedback whatsoever - the addition of feedback will reduce the distortion even more - as will some biasing on the output stage - neither of these were included to show the effect more clearly. Note that the settings on the oscilloscope were not changed between these simulations, to ensure that the full effect was not obscured (or accentuated) for each example.

The inescapable conclusion is that the high impedance drive will greatly assist in overcoming non-linearities in the output stage, even those as nasty as a completely unbiased

complementary pair as shown. Naturally other non-linearities are also negated (to some degree) by these techniques, which are pretty much universal in modern amplifier designs.

None of the above should be taken to mean that we no longer need negative feedback - the idea is to make the amplifier as linear as possible before feedback is added. This makes a good amp better - feedback will never make a bad amp good !

Current Mirrors

The next item is current mirrors, another little understood circuit. These are extremely useful in amplifier design, and in this section I will show where they can be used, and the benefits that can be obtained.

The basic current mirror is shown in Figure 10, and it can be seen that whatever current is injected into the left side is mirrored, and the right hand side is a constant current source (sink) reflecting the injected current. Should the input current at change, so will the output current, but it will remain constant, regardless of the actual voltage (provided it remains within the supply limits of course).

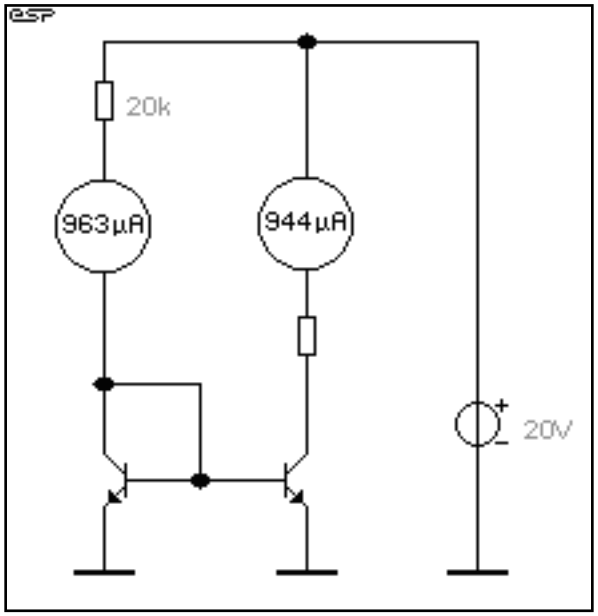


Figure 10 - A Basic Current Mirror

The problem with this circuit is that the current in the two halves is different - the mirrored current is too low, differing by 19 uA (it is actually 20 uA, but the simulator's accuracy is not great enough to show this). If we check, we will find that the emitter currents of both transistors are identical, so the 20 uA that "disappeared" is the base current that must be supplied to each transistor (10 uA each).

Adding emitter resistors does absolutely nothing to alleviate this, but is useful if the transistors

are not matched. Even then the resistors do not really do a lot of good, unless the voltage developed across them is significant (at least 100 mV, and preferably more) but it helps a little bit.

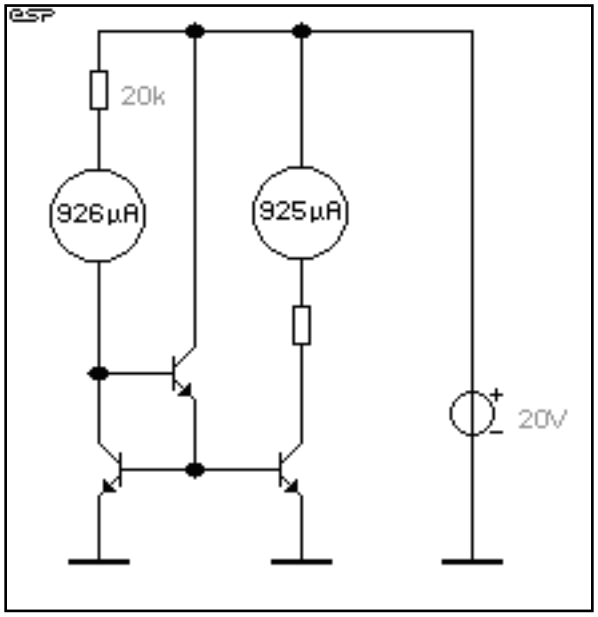


Figure 11 - Buffered Current Mirror

A better solution is to use a buffer as shown in Figure 11. This removes the base current component of the error, and makes the current mirror matching a lot better. This simple addition has reduced the error dramatically, and now all we have to do is find a use for the circuit.

Differential Pair Amplifier

Consider the differential pair (aka long tailed pair or LTP). Most of the time, we are losing half the gain of the circuit, since the output is taken from only one collector as shown in Figure 12. This configuration also suffers from linearity problems, unless the output is current only - as is the case when driving the base of a transistor (this is shown below).

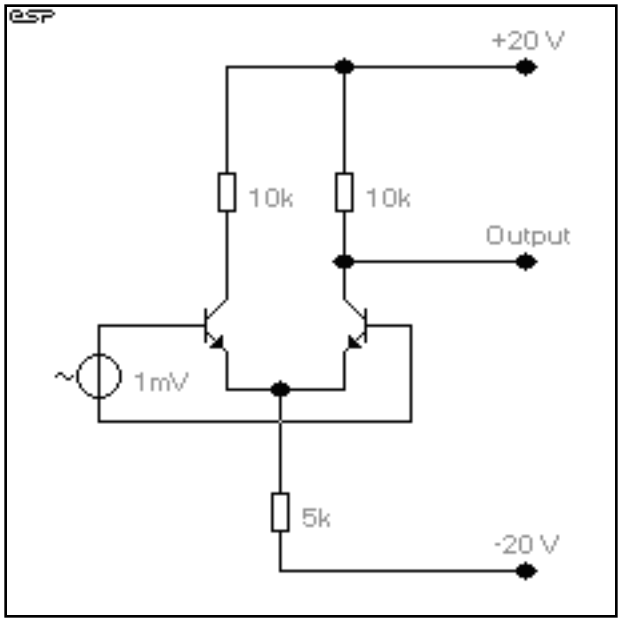


Figure 12 - The Long Tailed Pair As A Voltage Amplifier

The circuit as shown (without the essential biasing components, which were omitted for clarity) has a voltage gain of 315 (again using transistors with a Hfe of 100), and is quite linear at low output voltages. The linearity will suffer badly as the level increases, and even with the +/-20 volt supplies used is not satisfactory as a voltage amp for outputs greater than about 1.5 Volts RMS (this is at an input voltage of 5 mV).

Using a current source / sink in the "tail" is very common in amplifier circuits, and this variant is shown below. It is commonly (but entirely mistakenly) assumed that this increases the gain, but the real purpose is to improve the common mode capability of the circuit. Common mode signals are those that are applied to both inputs in the same polarity, and are generally required to be rejected. Using a simple resistive tail severely limits the common mode voltage that can be accommodated before severe distortion occurs, and indeed renders the common mode rejection of the circuit almost useless.

In the example above, the common mode rejection is only 1.8 dB but with a current sink tail the rejection is so good that the simulator cannot measure it. In reality this will rarely be the case, but in most amplifier circuits common mode signals (of the undesirable kind) are not an issue if the input stage is properly designed.

If a current mirror is used as the load, gain is increased by a very useful amount, and Figure 13 shows the arrangement used. The stage gain is now over 2,000 and the use of a current sink as the tail has only a marginal effect (provided that the current is maintained at 4 mA). This circuit is shown in Figure 13, and although useful in certain applications, it is not suitable to drive the output stage of a power amplifier.

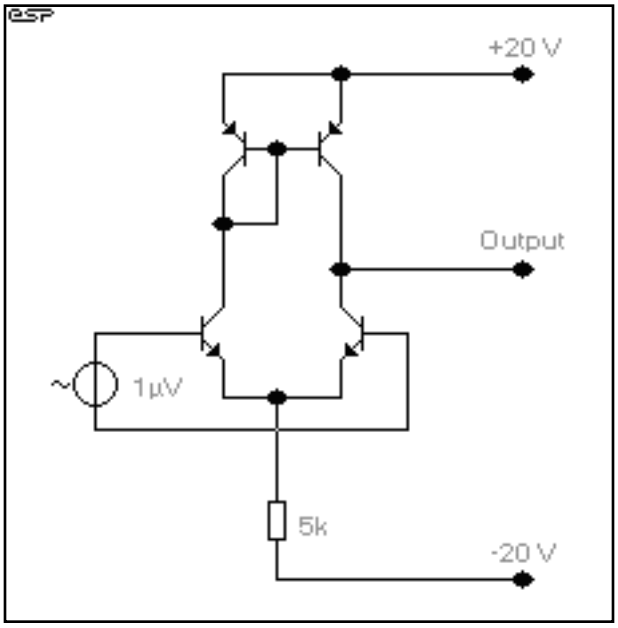


Figure 13 - Using A Current Mirror The LTP Load

If we really wanted to get silly (and I have seen it done), we can put all the bits together in one place, and finish up with an input stage + Class-A driver, with a total gain of a little over 400,000 (based on 100k feedback loading). This will increase dramatically when the output is buffered, but because of the extremely high impedance, suffers when loaded even with 100k. The buffered current mirror does not help the gain, but reduces output offset. The complete circuit is shown in Figure 14, and is a useful example of the techniques discussed in this article.

As we discovered earlier, extremely high output impedance is not a bad thing for this type of application, because the gain can be kept to a manageable level, but still have the necessary reserve to overcome output stage non-linearity before the application of global feedback.

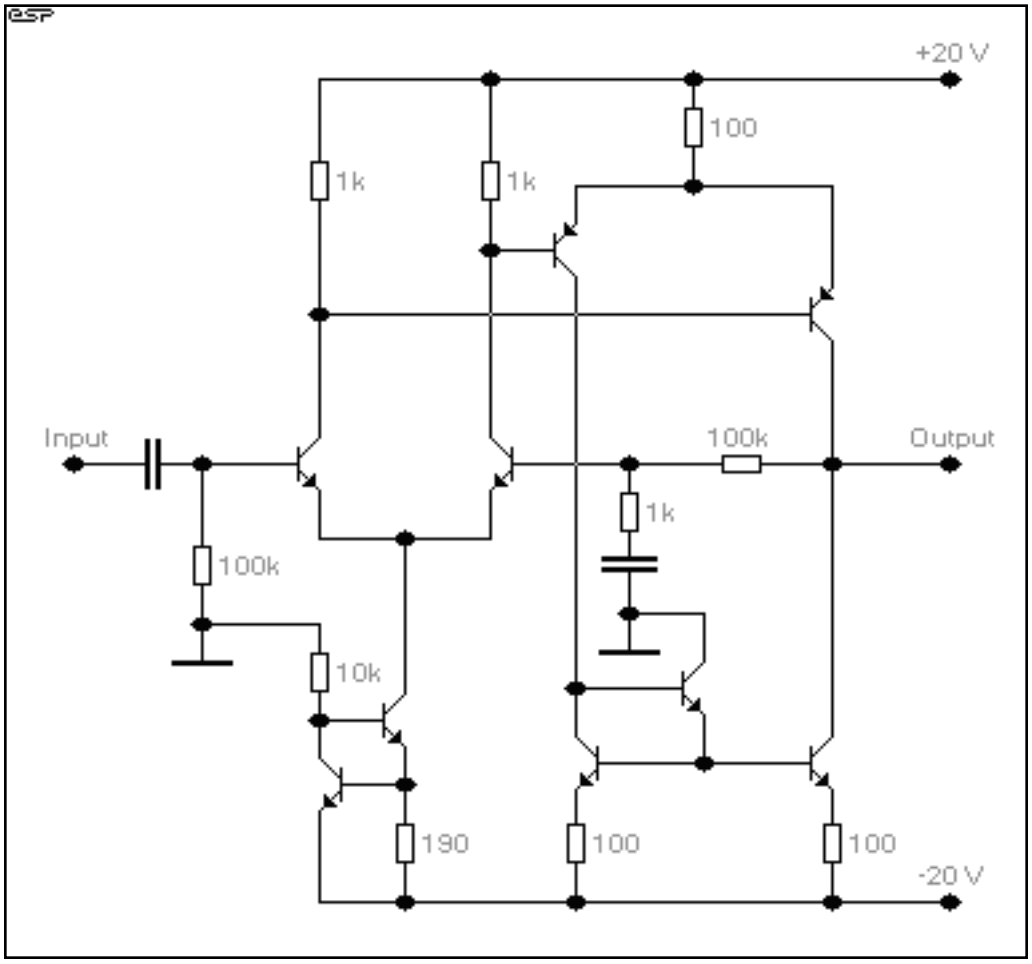


Figure 14 - Combination LTPs and Current Mirror

This circuit uses an input LTP that drives a secondary LTP as the Class-A amplification stage. The load for this second LTP is a current mirror, and this arrangement has excellent linearity. I suspect that it could be a cow to stabilise in a real amplifier circuit, and quite frankly, I do not see any reason to go this far. One amplifier I have seen using this arrangement was a MOSFET type, and given the inherent non-linearities of MOSFETs the additional gain is useful. Initial measurements on the MOSFET amp show that stability is not as good as I would like to see (there are traces of oscillation at some output levels), but overall stability seems to be acceptable. I expect that the wide bandwith of the MOSFET output devices might make this arrangement a little more tractable than would be the case with bipolar transistors.

Conclusion

At this stage, I suspect that I have either cleared up some degree of confusion, or created even more. There are so may choices, and so many reasons to use (or not use) any one of them. This article is meant only to describe and explain - you can see where my own preferences lie from the circuits published in the projects pages. I happen to be a strong supporter of the KIS principle (keep it simple), and in my own designs have not found it necessary to make the circuit any more complex than is needed to make it work and provide

good distortion figures. Listening tests bear this out, and IMHO I do not believe that the difference between a highly complex amp and a simple one (provided that both are well designed) is audible under any listening condition.

This being the case, I cannot think of any good reason to go for the complex solution, especially since it will almost certainly be much harder to stabilise (I happen to think that this is extremely important), and will use more parts. Granted that the additional cost is minimal in the greater scheme of things, but sound for sound, simple and reliable has to win out.

As a final point, with the LTP configuration and current mirrors, it is often beneficial to match the transistors, so they have gains that are approximately equal, and emitter-base voltages that are as closely matched as possible. It is also useful with both of these configurations and two transistor current sources to thermally bond the two transistors to prevent mismatch due to thermal effects.

Remember that the emitter-base voltage of a transistor falls at a rate of 2mV per degree Celcius, so even a small temperature difference will change the DC offset performance quite dramatically.

For more information on the design of amplifier stages, please refer to the Amplifier Design page.

ESP Home Page

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Revision History: 28 Jan 2000 - Corrected some errors and typos./ Page Created - 15 Aug 1999

Elliott Sound Products **The Design of Heatsinks**

Rod Elliott, ESP

Page Last Updated 17 Sep 2003

Contents

1 - Introduction
2 - Types of Heatsinks
3 - Sample Calculations
4 - Measuring Heatsink Thermal Resistance
5 - Transistor Case Styles
6 - Junction to Case Thermal Resistance
7 - Case to Heatsink Thermal Resistance
8 - Mica
9 - Thermal Compounds
10 - Thermal Inertia
11 - Minimising Thermal Resistance

- 11.1 - Mounting The Transistors
- 11.2 - Further Reductions of Thermal Resistance
- 11.3 - Some Additional Thoughts

12 - Introduced Thermal Resistance
13- Flat-Pack Transistors
14 - Where to Mount Transistors
15 - Heatsink Surface Area
16 - Altitude
17 - Fin Density
18 - Water Cooling

19 - Heat Pipes
20 - The Bottom Line
21 - Other Heatsink Information
22 - References

Main Index

Articles Index

---

1 - Introduction

This article is essential reading for anyone who uses heatsinks in audio, and the same principles apply for power control, RF and CPU cooling, as well as many other applications in electronics. There are many misunderstandings and misconceptions about transistor mounting. What are the essential requirements? Which things make the biggest difference? These questions may not have occurred to many readers, but they are very important to ensure the long term reliability of your projects.

Please ... read all of this article - not just the bit you think you need. You would be surprised how much I learned while researching and writing this material, and I have had a lot of additional information supplied by readers (you know who you are, and thank you for your input). The material here is the result of a lot of hard work, research, personal experience and (what should be) common knowledge.

The design of heatsinks, or to be more exact for most builders, the selection of a suitable heatsink, is not difficult once the basics have been mastered. Terms such as 'thermal resistance' and 'degrees C / Watt' are a little daunting for the uninitiated, and the purpose of this article is to explain how thermal transfer works, from the transistor die until it is finally dispersed into the atmosphere.

This sort of information seems to be unusually difficult to obtain, either on the web or elsewhere, and I have had great difficulty in getting thermal resistance data from anywhere - although as you shall see as you read on, I have managed it. Not that this article is the "last word" on heatsinks - far from it, but it is a collection of some of the most useful information I have been able to gather thus far.

Note that in this article, the 'transistor' can be a bipolar, MOSFET, TRIAC, CPU or any other semiconductor device (including diodes or ICs of all kinds) that is mounted in a case which must then be mounted on a heatsink of some description.

The reliability and longevity of any semiconductor device is (roughly) inversely proportional to the square of the junction temperature change. Thus halving the junction temperature will result in approximately 4 times the expected life of the component. The converse is also true! A worthwhile increase in reliability and component life can be achieved by a relatively small reduction in operating temperature, since these parameters increase exponentially as temperature is reduced.

The whole process of removing heat from a transistor's active area (the die or 'junction') involves many separate thermal transfers, and we shall examine each in turn. In order to see a meaningful result, we need a heat input value and an ambient air temperature. From these two basic elements, the entire heat flow can be established.

Towards this end, an electrical analogue may be drawn, showing the thermal generator as a current source, thermal resistance as a resistor and the thermal inertia (or transient capacity) of the various materials as a capacitor. The transient capacity is the ability of any material to absorb a quantity of heat for a short time, after which its temperature will rise in the same way as the voltage rises in a capacitor if the supply current is maintained.

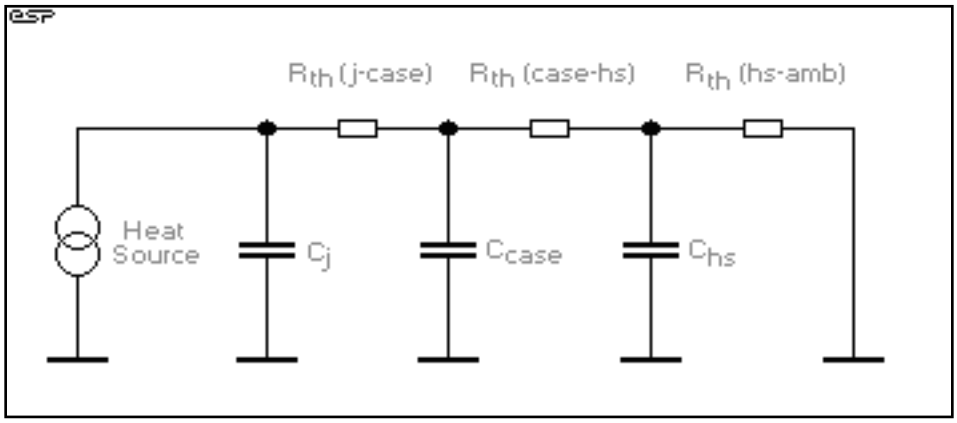


Figure 1 - Heat Flow From Generator To Ambient

From Figure 1, we can see the thermal resistance (Rth) from junction to case, the thermal inertia of the junction (this is very small), the thermal inertia of the case itself (again, not large), the thermal resistance from the case to the heatsink, thermal inertia of the heatsink (this might be very large) and finally the thermal resistance from the heatsink to ambient.

The small thermal inertia values should be ignored, as they will heat up very rapidly, and

the heatsink itself will only absorb so much heat. A state of equilibrium is required where the heat input equals the heat output, but without the junction temperature reaching a dangerous level - even momentarily.

---

2 - Types of Heatsinks

Heat sinks can be classified in terms of manufacturing methods and their final form shapes. The most common types of air-cooled heat sinks include:

- **Stampings** Copper or aluminium sheet metals are stamped into desired shapes. They are used in traditional air cooling of electronic components and offer a low cost solution to low density thermal problems. They are suitable for high volume production, because advanced tooling with high speed stamping lowers costs. Additional labour-saving options, such as clips and dry interface materials, can be factory applied to help to reduce the board assembly costs.
- **Extrusions** These are by far the most common for power amplifiers, and allow the formation of elaborate two-dimensional shapes capable of dissipating large heat loads. They may be cut, machined, and options added. Cross-cutting will produce omni-directional, rectangular pin fin heat sinks, and incorporating serrated fins improves the performance by approximately 10 to 20%. Extrusion limits, such as the fin height-to-gap ratio and fin thickness, usually dictate the flexibility in design options. Typical fin height-to-gap aspect ratio of up to 6:1 and a minimum fin thickness of about 1.3mm are attainable with standard extrusion processes.
- **Bonded / Fabricated Fins** Most air cooled heat sinks are convection limited, and the overall thermal performance of an air cooled heat sink can often be improved significantly if more surface area can be exposed to the air stream. Such high performance heat sinks utilise thermally conductive aluminium-filled epoxy to bond planar fins onto a grooved extrusion base plate. This process allows for a much greater fin height-to-gap aspect ratio of 20 to 40, greatly increasing the cooling capacity without increasing volume requirements.
- **Castings** Sand, lost core and die casting processes are available with or without vacuum assistance, in aluminium or (less frequently for audio applications) copper/ bronze. This technology is used in high density pin fin heat sinks which provide maximum performance when using forced air cooling.
- **Folded Fins** Corrugated sheet metal - either aluminium or copper, increases surface area and, hence the volumetric performance. The heat sink is then attached to either a base plate or directly to the heating surface via epoxying or brazing. It is not suitable for high profile heat sinks, but it allows high performance

heat sinks to be fabricated for specific applications.

---

3 - Sample Calculations

Let us assume that the heat generated in the transistor die is 50W, such as might be the case with a Class-A power amplifier operating from a +/-25V supply, and drawing 2A quiescent current. Only half the total supply voltage of 50V is across the transistor in the quiescent or no signal state when the output is at zero volts, so each of the transistors in the output stage will have a voltage of 25V at 2A, or 50W dissipation.

An ambient temperature of 25°C is a good starting point, and provides a safety margin for most domestic systems, although as we shall see later on, it is important to design for worst case if reliability is to be maintained.

The primary aim of the heatsink designer is to ensure that the total thermal resistance is kept to the minimum possible value, and the entire design process looks at thermal resistance as the primary item to be calculated. Only after this has been determined can the actual temperature of the transistor junction be predicted.

To some extent, this article was prompted by a reader of my pages, who complained that transistors in an amp he built ran very hot (mounted on a bracket), while others seemed to run fairly cool. So I got to thinking, didn't I?

How to actually mount transistors does not seem to have achieved much coverage on the web, but I have seen some absolute drivel spouted by PC types talking about heatsinking Pentium processors - the same principles apply, but the amount of heat and thermal dynamics are a little different.

Throughout this article I will often refer to "aluminium", which in this context means aluminium alloy. Pure aluminium is rarely used, since it is too soft and easily distorted. It is also quite difficult to drill and tap for mounting screws, because it tends to clog the flutes on the drill bit, and will even snap the drill if great care is not taken.

When drilling aluminium (or any of its alloys), the use of methylated spirit (denatured alcohol) or WD40 (or other similar 'water displacement' spray) as a lubricant is recommended. This trick is not well known, but is very effective.

Be warned that 'metho' (as it is commonly known in Australia) is highly flammable, and burns with an almost invisible - but very hot - flame. Extreme care is required, and use only enough to keep the drill bit wet - the last thing we need is one of my readers burning down the house while drilling a heatsink. Needless to say, I can accept no responsibility for any accident caused by the use of this technique.

On the positive side, metho evaporates quickly and leaves no residue. This also means that spontaneous combustion is far less likely than with any oily lubricant. You should *never* just drop oily rags in the dustbin, as it is quite possible for them to burst into flame for no apparent reason.

---

4 - Measuring Heatsink Thermal Resistance

The most accurate way to determine the thermal resistance of an unknown heatsink is to measure it. The exercise is not trivial though, since you will require a large metal clad resistor having a good flat bottom surface (or you can use transistors), a contact thermometer (a conventional alcohol in glass thermometer cannot be used), and a suitable low voltage, high current power supply. If you have a large number of heatsinks to test it may be worthwhile to build a dedicated test unit, however this is unlikely for most home constructors.

It is important that the heatsink under test is set up as closely as possible to the way it will be used. There is no point testing a sink just lying on the workbench (for example), as the results will be way off. If a heavy chassis is planned, then attach the heatsink to the chassis or a reasonable facsimile thereof. Ensure that the heating system is in the best possible thermal contact with the heatsink. Thermal compound is essential, and do not use any insulators.

The test is based on knowing the voltage and current you apply to the heatsink heating system (resistor(s) or transistors), and being able to accurately measure the ambient and heatsink temperatures. First, apply a relatively low power to the heater system of your choice, and wait for the heatsink temperature to stabilise - this could take an hour or more. If the heatsink is too hot or too cold the results will be inaccurate, so slowly (in steps) increase power until the heatsink is at approximately the maximum temperature you feel is reasonable (typically around 50-60°C).

Measure the ambient temperature and the heatsink temperature, preferably using the

same thermometer. A contact thermometer is essential for the heatsink (again, use thermal compound). Determine the temperature difference (temperature rise) between ambient and heatsink.

Next, determine the power applied to your heating system. Thermal resistance may now be established with some very simple maths ...

You will use the following terms -

Tr - Temperature rise
Ta - Ambient temperature (example 22°C)
Th - Heatsink temperature (example 54°C)
Vh - Voltage to heater (example 12V)
Ih - Current to heater (example 3.5A)
Ph - Power applied to heatsink
Rth - Thermal resistance (in °C/W) so ...

Tr = Th - Ta= 54 - 22 = 32°C
Ph = Vh * Ih= 12 * 3.5 = 42W
Rth = Tr / Ph= 32 / 42 = 0.76°C/W

This is as accurate as you need, and as good as you'll get in real life. To get accurate results is time-consuming, and is usually not necessary once you get a reasonable feel for heatsinks. Once you do this a couple of times you will be able to 'guesstimate' the approximate power handling capacity of a heatsink just by looking at it and checking the manufacturer's data. Bear in mind that few heatsink manufacturers supply the all important temperature rise information, so their figures can be off by 25% in either direction.

Thanks to "Scorpion" on the ESP forum for the idea

---

5 - Transistor Case Styles

There is a vast number of different case styles available. I shall only deal with the most common (as shown below), but the information is equally valid for other case styles. In many cases, an educated guess will be needed to determine case to heatsink thermal resistance if the case used by your device is wildly different from those quoted. Generally, this is based on surface area and the evenness (or otherwise) of pressure distribution directly beneath the internal silicon die. This is a wildly variable quantity, and has been known to cause may a constructor grief, not realising the importance of evenly distributed pressure.

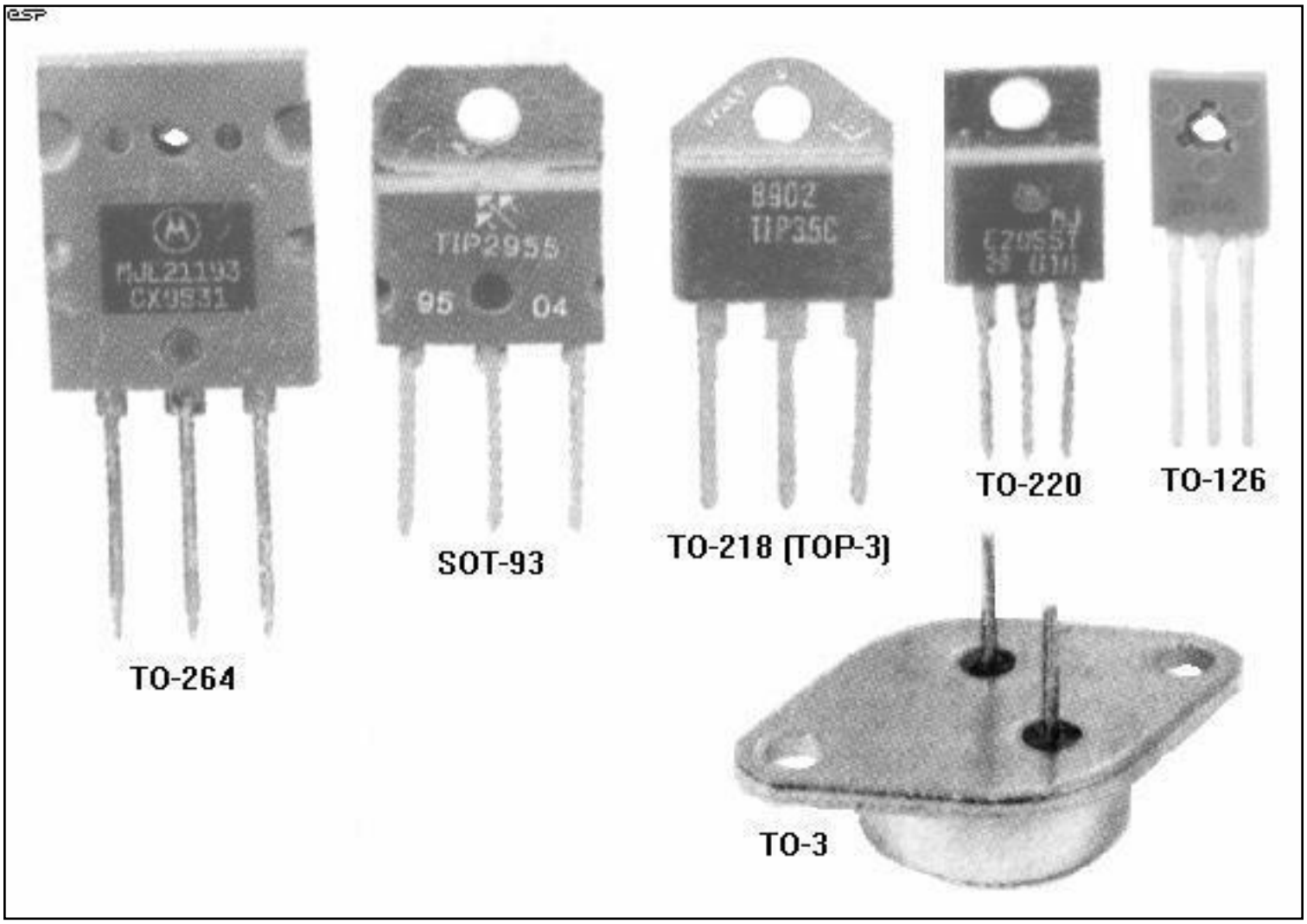


Figure 2 - Common Case Styles

This is only a sample of those currently available (including ICs of various types), but is representative of the most common for transistors, MOSFETs and the like. In most cases, the die can be assumed to be roughly in the middle of the plastic moulding for all plastic encapsulated devices. In the TO-3 package, it is located in the geometric centre of the package (more or less).

---

6 - Junction to Case Thermal Resistance

Although this is always an important consideration, there is nothing the amplifier designer can do to reduce this for any given transistor. Manufacturer's data will sometimes quote this figure, but it is more common to refer to the device case temperature - this is easier to deal with, since the data sheets have already taken die to case thermal resistance into consideration when the temperature derating graph was produced.

Derating is commonly applied to semiconductor devices, and typically the maximum power dissipation claimed for most devices will be at or below 25 degrees C. At any temperature above this, less power is available from the transistor with increasing temperature, until at some figure (typically around 150°) the permissible power dissipation is zero.

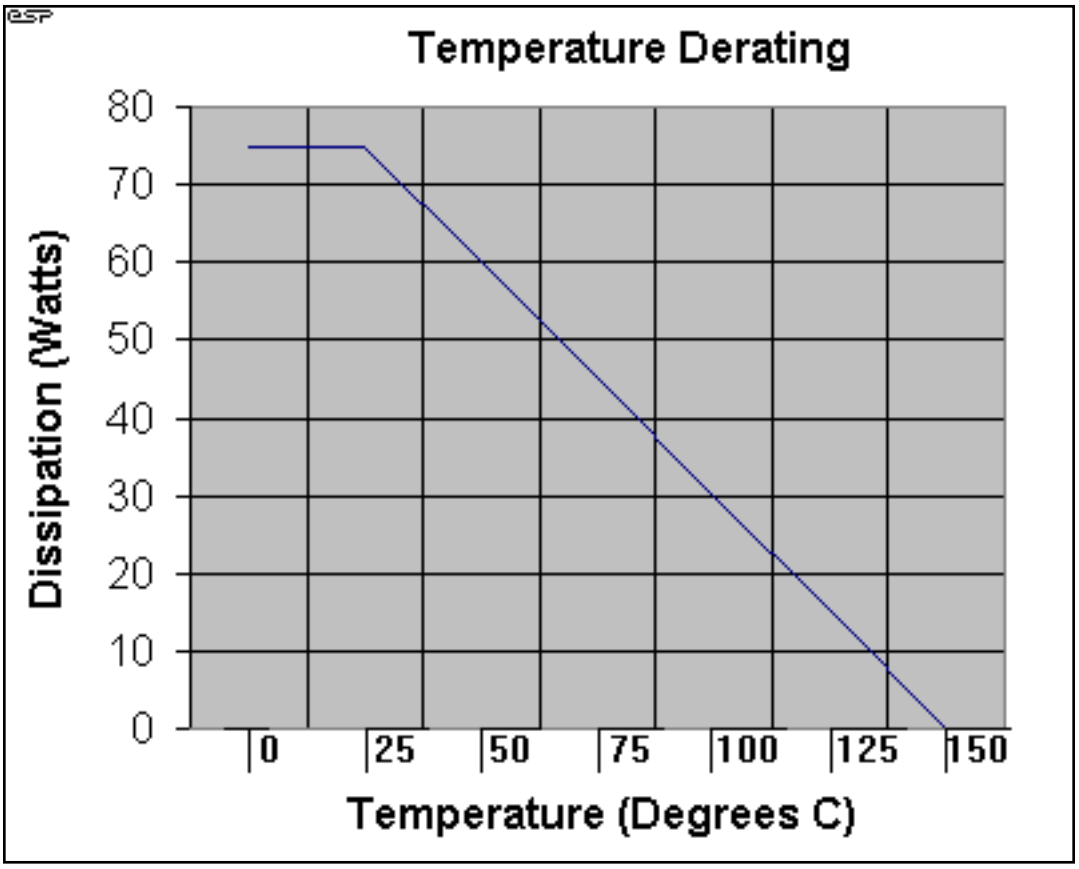


Figure 3 - Case Temperature Derating For MJE3055

Figure 3 shows the case temperature derating curve for an MJE3055 (TO-220) power transistor as an example. As can be seen, at 25 degrees and below, the device is rated at 75W dissipation, but at 100 degrees, 30W is the maximum permissible. At 150 degrees, no power may be dissipated at all.

This device has a 1.67°C thermal resistance from junction (die) to case (Motorola specification sheet for the MJE2955/3055), so the actual junction temperature will be somewhat higher than the case temperature until 150 degrees, at which no power may be dissipated, so case and junction will be at the same temperature.

Our first calculation shall be to see what the junction temperature will be at 75W device dissipation, with the case at 25 degrees (Celsius):

Thermal resistance = 1.67
Power = 75W
Therefore junction temperature rise = 1.67 x 75 = 125 degrees

If the junction is 125 degrees above ambient (25 degrees) then the total is 125 + 25 = 150°. This is a seriously high temperature, and if you were to do the same calculation for all the temperatures on the scale above you will quickly see that the junction temperature is 150° at maximum dissipation for any given case temperature. The thermal derating curve simply limits the allowed power dissipation to ensure that the junction never operates at above 150 degrees.

The information above is obtainable from all device manufacturers, and is essential reading to ensure that transistors are not damaged by excess temperature. The maximum quoted temperature should never be exceeded, as the device will have a considerably lower life expectancy if overheated. Instantaneous failure is not uncommon if a device at an already elevated temperature is called upon to suddenly do some work.

Using our theoretical amplifier above, we know that the dissipation will be 50W per transistor (or a total of 100W), and from the derating curve we see that for this power, the maximum case temperature is 65° C. The goal of this exercise is to determine the size of heatsink needed to ensure that the transistors ratings are not exceeded. We can also see if it is possible to operate them at a lower temperature, thus prolonging their operating life.

The junction to case thermal resistance varies widely, but will rarely be less than about 0.5 degrees C/W, depending (of course) on the manufacturer, the case style, and the type of device. A few examples:

- IRF540 MOSFET - 1.0°C/W
- MJL21193/4 BJT - 0.7°C/W
- MJE3055/2955 BJT - 1.67°C/W (our demonstration model)

Also remember that in use, the base-emitter voltage of transistors falls at 2mV / degree C, and leakage current increases exponentially (leakage current doubles for every 8 to 10 degree C increase in temperature). These are two very good reasons to try to keep the operating temperature as low as possible.

---

7 - Case to Heatsink Thermal Resistance

This is where the whole process often falls down, since semiconductor devices are nearly always electrically insulated from the heatsink. This means that some material (typically mica or Kapton - the latter is sadly hard to get now for some reason) must be used between the case of the device and the heatsink surface. This will invariably increase the thermal resistance - there is no known material which is a perfect electrical insulator and thermal conductor. Sad but true.

There are actually a number of alternatives for electrically insulating the transistor from the heatsink while still allowing heat transfer. Two were mentioned above, but there are others too. The basic list:

Material	Thermal	Electrical	Thermal Resistance	Other Properties
mica	Good	Excellent	~ 0.75 - 1.0	Fragile
Kapton	Good	Excellent	~ 0.9 - 1.5	Very robust
aluminium oxide	Excellent	Fair	~ 0.4	Fragile - easily damaged
beryllia (beryllium oxide)	Excellent	Excellent	~ 0.25	Toxic
Sil-Pads	Fair +	Excellent	~ 1.0 - 1.5	Convenient

Table 1 - Thermal Resistance of Various Mounting Methods (TO-220 Case)

The above table shows the approximate thermal resistance for various materials, and (except for the Sil-Pads) assumes that a good quality thermal compound has been used. Despite the claims made by some manufacturers, most thermal compounds are much of a muchness, since their primary purpose is to exclude air from the mating surfaces, and provide a modicum of heat transfer themselves. Air is a most excellent thermal insulator, and even the tiny air gaps that exist between the transistor case, insulating washer and the heatsink will increase thermal resistance dramatically.

Mica is probably the most commonly used insulator, but it has limitations. The worst of these is quality control, which results in mica washers being whatever thickness they happen to be (i.e. unpredictable). With infinite care and patience it is possible to shave mica down to a thickness of perhaps 0.05mm (0.002"), but few of have the time to fiddle about at this level. A micrometer is essential if you really want to do this. See the section on mica (below) for more information on this topic.

Kapton is an excellent material in this respect. Predictable thickness and very tough (try ripping one of these guys apart), but generally not quite as good as mica thermally. For reasons which remain entirely obscure, Kapton transistor washers do not seem to be as

readily available as they once were, which I find a real shame since they have always been a favourite of mine. However, Kapton tape can be made into very acceptable washers with a pair of scissors and a hole punch.

Aluminium oxide (anodised aluminium or 'anodized aluminum' if that makes you feel better) is an excellent electrical insulator, and provides very good heat transfer, but is fragile and easily damaged. Even a small scratch will generally allow an electrical short circuit, and obtaining a consistently thick (relatively speaking) layer of oxide is difficult in the extreme in mass produced heatsink extrusions. Anodised aluminium washers are available from some electronics suppliers, but are generally expensive and you will probably find that they are not available for all transistor case styles.

Beryllium oxide was quite popular a few years ago, until someone discovered that it is toxic. Consequently, such mounting washers have vanished (at least as far as the DIY fraternity is concerned). A pity, since most people I know have never eaten a transistor washer in their lives, but governments do so like to protect us from ourselves. In fact, it is inhalation of the beryllium oxide dust that causes problems (respiratory disease) rather than ingestion, and it turns out that they are still available, as are many other parts made from this unique and versatile ceramic. Because it *is* toxic, special precautions are needed during manufacture, and this makes it rather expensive.

> Thanks to Robert K. for providing me with some additional info on beryllium oxide and its continued use.

Sil-Pads are only comparatively recent additions to the mounting hardware options, and are very convenient to use. The down side is that they are not as good (thermally) as mica or Kapton, but they require no silicone grease or other thermal compound. There has been some comment made about the re-usability of Sil-Pads, with the claim being made (I do not remember by whom) that the deformations caused by prolonged heat and pressure will cause increased thermal resistance if they are re-used.

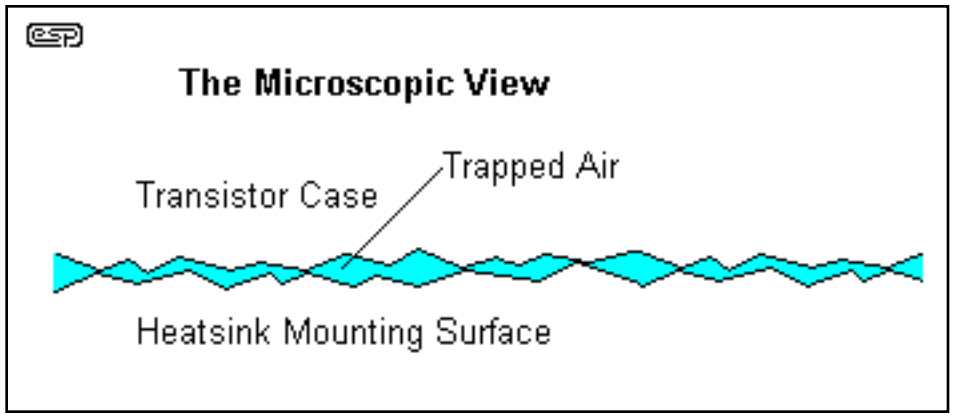
I have tested and verified this, and it makes perfect sense, since the thickness of the pad will not be even after it has been used, due to uneven pressure and mating surfaces (at a microscopic level). Even if the new device is located in exactly the same place on the pad, thermal resistance will be impaired. I suggest that Sil-Pads be used where maximum thermal conductivity is relatively unimportant, as their thermal resistance is generally not as low as claimed. The figure shown in Table 2 for the TO-3 and TOP-3 case styles is (IMHO) highly optimistic, and is unlikely to be achieved in practice.

Package	Direct Contact (Dry)	Direct Contact + Thermal Compound	Mica Insulator + Thermal Compound	Sil-Pad™
TO-3	0.5 - 0.7	0.3 - 0.5	0.4 - 0.6	0.4
TOP-3 (est)	0.8 - 1.1	0.5 - 0.7	0.6 - 0.9	0.65
TO-126	1.5 - 2.0	0.9 - 1.2	1.2 - 1.7	NA
TO-220	1.1 - 1.3	0.9 - 1.1	1.0 - 1.6	1.5

Table 2 - Comparison of Case Types

The information in Table 2 shows the variations expected with various transistor cases and mounting techniques. It should come as no surprise that these figures are different from those in Table 1, since this info is from a different source. I prefer to err on the side of caution when it comes to thermal resistance, because it is always better to have the devices run a little bit too cool than a lot too hot. I would suggest that Table 1 is more realistic (if not all that accurate).

The TOP-3 (TO-3 Plastic) package will have a case to heatsink thermal resistance (estimated) at about 1/2 way between TO-3 and TO-220 packages. I could find no information on this package for the above table, and the estimate is mine.



The importance of a thermal compound cannot be overstated. The drawing shows the surface of the transistor and heatsink (or mica washer) at a microscopic level. OK, so I just drew it, but this is exactly what you would see under a powerful microscope. The blue area represents air, and the surfaces only touch in a few places. This is why the thermal resistance is so much higher without the use of any thermal compound.

In some cases the situation is worse - I wonder how many readers have come across (especially) TO-220 transistors that look as if they had been machined with a chainsaw. A flat file, or fine wet/dry abrasive paper and water (no, it won't hurt the transistor) should be used to make sure that the surface is as flat and smooth as you can make it.

Never hold the transistor in a vice and file it! Hold the file still, and gently slide the

transistor on the file (or wet/dry paper on a sheet of glass) until it is smooth. There is no point having it nice and smooth if it is rounded - this *will* happen if you rigidly mount the transistor and hand-hold the file.

When you are finished, you should be able to lay the transistor on the heatsink and see no light between the two surfaces. This little bit of extra effort may mean the difference between the success or failure of your project - at least in the long term.

---

8 - Mica

What is Mica? Mica is transparent, and can be split into a very thin film along its cleavage. Electrically, it has the unique combination of high dielectric strength, endurance, uniform dielectric constant, low dielectric loss (or high Q), and extremely good insulating properties. Mica is moisture-proof and has low heat conductivity. It is infusible and may be exposed to high temperatures (in excess of 700°C) without any noticeable effect. If you want more info - try a web search :-)

So much for the descriptions, but why would mica get a section all of its own? Simply because the quality control is virtually non-existent (from what I have seen lately). The overall shape is fine, but the thickness is generally too great - commonly by a huge margin. I have used mica washers that were so thick that I was able to separate them with a (very) sharp knife, and obtained four washers for the price of one (plus some scrap from the splitting operation).

Splitting a mica washer is not hard, but requires a steady hand, a good magnifier, and the sharpest knife you can find (or a new razor blade). The thickness should ideally be in the order of 0.025 to 0.05 mm for normal use, but you will probably find that up to 0.1 mm is quite acceptable. It is possible to make it much thinner than this, which decreases thermal resistance but makes the washer fragile and easily damaged.

The electrical characteristics of mica are such that even the thinnest possible washer will be more than adequate for typical amplifier voltages. The dielectric strength allows mica to withstand 1,000 - 1,500 volts per mil (1/1000 inch), or about 0.025 mm (25 micrometres) of thickness without puncturing or arcing.

This being the case, I am unsure why commercially available transistor washers are anything up to 0.25mm thick (and no, I am not kidding - I have even found some thicker than that!). This is capable of withstanding over 10kV in theory, let alone a measly 200V or so, and will introduce considerable thermal resistance. If one had a steady enough

hand, just one of these would yield 10 washers of acceptable thickness (thinness?), with each still capable of at least 1kV insulation breakdown.

I am even tempted to make a mica splitter, that can be adjusted to a suitable thickness. One of these days ...

---

9 - Thermal Compounds

One of the most common mistakes made by hobby electronics enthusiasts (and quite a few professionals too), is to assume that if a little thermal compound is good, a lot must be better. Absolutely not so! The amount of thermal compound should be exactly that amount which ensures that an air-free join is made between the mating surfaces. If too much is applied it will cause an increase in thermal resistance, since it is not really that good at conducting heat. Generally speaking, any electrical insulator is also a thermal insulator, so the thinner the final composite insulation - including thermal "grease" - the better.

Having said that, one must ensure that the electrical insulation is sufficient for the applied voltage or disaster will surely follow - usually in a spectacular fashion - especially if high voltages or currents are available.

A quick note on the application of thermal compounds is in order. This is the method I generally use, and it works very well once you have a copious supply of workshop rags to clean the mess from your fingers.

Apply a small quantity of the thermal compound to one finger, then gently rub with your thumb to create an approximately even coating on finger and thumb. Now rub the thermal compound onto a washer, held between thumb and finger, ensuring that the coating is just thick enough to be opaque, but thin (and even) enough to ensure that the contact will be absolute on both surfaces (transistor and heatsink). Test your skill until with moderate pressure, you can leave a perfectly formed outline of the transistor, with no blobs or gaps, on the heatsink surface.

---

10 - Thermal Inertia

Since any heatsink has some mass, it will also posses thermal inertia, which is to say that it takes time for the body of heatsink material to heat up. Naturally, the larger the heatsink (and the more aluminium it has in it), the longer it will take to heat. This can

easily lead one to believe that the heatsink is large enough for the job, but only by running an amp for an extended period will the truth reveal itself.

Thermal inertia is a good thing, because it allows the heatsink to absorb transient power "surges", and will dissipate the heat again during low power operation. Because of the dynamics of music, this tends to work well, but for professional use (where heavy compression is often used), the average power into the heatsink can be very high for prolonged periods, so either massive bulk material must be used, or a combination of reasonable bulk and lots of surface area. Fan cooling is almost mandatory for professional use at high continuous power levels.

Even if one were to obtain an infinitely large block of aluminium, if a bracket or other mounting arrangement directly underneath the heat source (the transistors) is not thick enough, it will have significant thermal resistance, and the transistors may just overheat anyway, so we do need to look at all the resistances in the thermal circuit, not just the heatsink itself.

It is not uncommon to have transistors operating at well in excess of their ratings, but a casual "finger" test of temperature on the surface of the heatsink will appear to indicate that all is well.

---

11 - Minimising Thermal Resistance

The key to obtaining maximum power from any transistor is minimising the thermal resistance from junction to free air, and as can be seen from the above information, one of the worst offenders is the transistor itself.

11.1 - Mounting The Transistors

The most common way to mount the transistors is to use metal thread screws, either with nuts on the back of the heatsink, or with a thread tapped into the heatsink itself. In production environments, spring clips are often used, and even rivets are sometimes used for mounting. The latter is a *very* bad idea, as the following tale describes ...

> Transistors, especially the TO-220 are sensitive to mechanical shocks. The T0-220 for example should never be attached to its heatsink by means of a rivet, either conventional or pop rivet. I was building replacement voltage regulators for a large automotive manufacturer, where the original device was riveted to a metal base with a mica insulator in between. Our version of this worked okay for a while until one day the manufacturing setup guy over adjusted the rivet force

> and the was some deformation of the tab. The parts kept passing the 100% electrical test and the manufacturing group kept making them over my objections. This too worked okay for awhile until we started getting field failures. When we investigated, we found that the power Darlingtons inside were suffering "earthquake" damage due primarily to the shock of the rivet machine. Often it was the outer edges of the device that would crack, leaving the device functioning but affecting V_{ce} Sat. With the higher saturation voltages (not measured on our tester of the time) the devices overheated and failed.
>
> It took 7 tractor trailer loads to haul all of the inventory back from our warehouse. We had to rework over 100K voltage regulators.
>
> Having seen the problem once, you would be amazed at how often I have seen it since. One guy was using a pop rivet gun. There was no damage to the tab but the shock from the pop of the rivet sent its message through the tab to the silicon. I saw another guy who was cutting the tab off the device in order to package it in the available space. I saw a piece of home brew equipment where the tab was being held by a nut and bolt that had been over tightened distorting the tab. Your comment about not holding the device in a vice and not filing on it with a hand file basically speak to the same issue: Silicon is thin and brittle. Do not alter the package or subject it to shock

My thanks to Mike for that story. I have seen transistors riveted to heatsinks as well, but that was many years ago, and I don't know if it caused any problems. The main issue is that riveting is at best unpredictable, and also creates considerable shock when the mandrel breaks (for pop rivets at least).

The key to mounting is pressure - it must be just right. Too much, and you distort the device case and possibly damage the thread or cause protrusions in the soft aluminium heatsink. Too little, and the thermal resistance will be too high. High reliability applications (Mil-Spec) will demand that a torque driver is used, to ensure that every device is tightened to exactly the right pressure. I don't suggest for an instant that this is needed in your project (torque drivers tend to be rather expensive), but it is essential that you develop a feel for tightening the screws to get consistent and reliable results.

Once the transistor mounting technique is perfected, we can generally assume that the thermal resistance will be about 1° C/W. This may be bettered, but again, a safety margin is always useful. At a continuous dissipation of 50W, this means a temperature rise of 50° C due to mounting and insulation requirements. Since we already worked out from the derating graph that the absolute maximum case temperature is 65° for 50W dissipation, we are faced with a problem - the maximum heatsink temperature is only 15° C. If this were to be the (somewhat chilly) ambient temperature, a 0° C/W heatsink is needed!

Perhaps I have been looking in the wrong places, but I have not been able to find a heatsink which can manage 0° C/W, and if the ambient temperature were to rise to (say) 20° , the heatsink now needs a negative thermal resistance - this is called a heat pump, and the most common example is an air conditioner or refrigerator. This approach, although quite feasible, can become somewhat expensive. Peltier devices (hot/cold junction semiconductors) can be used, but tend to be expensive, and require a fairly heavy current. In order for the cold junction to work at a suitably low temperature, the maximum temperature of the hot junction must be kept as low as possible - this requires (only one guess!) ... a heatsink.

For the above example, because there must always be a thermal gradient (temperature difference) to allow heat to flow from one device to another, this means that ambient temperature must be -10° C to be able to use a 0.5° C/W heatsink! We have just designed an impossible output stage, which cannot be made to work in real life without considerable additional cost.

11.2 - Further Reductions of Thermal Resistance

It is generally far cheaper to reduce the thermal resistance of the devices and their mountings than it is to use the largest heatsink you can possibly find, or go to all the bother of water cooling (which works extremely well, but is difficult to implement in the listening room), or using fans which are noisy and spoil the signal to noise ratio of your equipment. There is no point ensuring that the S/N ratio is 100dB or more, only to have a background noise of perhaps 55dB SPL created by cooling fans. Remember that even if the heatsink stays at ambient temperature (an infinite amount of air will be needed :-), from the above example, the absolute maximum ambient temperature is 15° C.

One technique used to reduce thermal resistance is simple - use two (or more) transistors in parallel in place of a single device. Although the thermal resistances for each of the transistors remain the same, the resultant thermal resistances for a "parallel pair" are effectively halved. This is due to the fact that each transistor is only dissipating half the total power.

Heatsink calculations for the parallel pair are best carried out by considering each transistor individually. In this example, each transistor of a parallel pair has a power dissipation of 25W.

From Figure 3, the maximum case temperature must not exceed 105° C. With a case to heatsink thermal resistance of 1.0° C/W, there will be a 25° C temperature difference

between the two. The maximum heatsink temperature is therefore 80° C (at this temperature, protection would need to be provided to prevent accidental contact, but this *must not* affect airflow across the heatsink). If we assume an ambient temperature of 30° C, each transistor requires a heatsink with thermal rating of 2° C/W. The parallel pair will require a heatsink of twice this size, i.e. 1.0° C/W. The calculation in the preceding paragraph does not allow for any safety margin (other than that built-in to the case to heatsink thermal resistance assumptions) and it would be better to design using a higher ambient temperature. For an ambient temperature of 40° C, the heatsink for the parallel pair would need to have a rating of 0.8° C/W and for an ambient of 50° C, 0.6° C/W. If we are going to mount all the transistors for a single amp on the one heatsink (which is the most common approach), then the heatsink must have a rating of 0.5° C/W (assuming a conventional output stage) for an ambient temperature of 30° C and no safety margin. The ratings at 40° C and 50° C would be 0.4° C/W and 0.3° C/W respectively. Similar calculations can be carried out for any number of paralleled transistors.

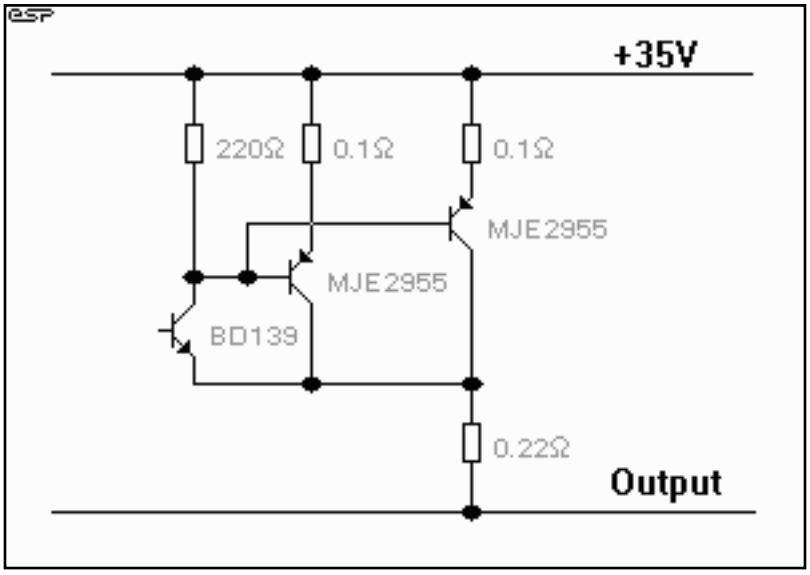


Figure 4 - Connecting Transistors in Parallel

The 0.1 Ohm resistors shown in Figure 4 ensure that each transistor carries the same (or at least roughly equal) current. Without them, the transistor with the higher gain (or the lower emitter-base voltage V_{be}) will take the majority of the current. This will cause it to get hotter than the 'lesser' transistor, which in turn increases its gain further and lowers V_{be} even further, which means it will get hotter, and so on. Note that the use of 0.1 ohm resistors is restricted to transistors with fairly closely matched V_{be}. Higher values are recommended where matching is impractical. This is all standard stuff in amplifier design. The 0.22 ohm resistor is used to stabilise bias current and introduce local feedback. These must not be omitted.

There are other ways in which thermal resistance can be reduced. For example, we are assuming that the transistors will all be on the same heatsink, which will probably be earthed to the chassis. If the NPN and PNP transistors were to be mounted on separate heatsinks, and were not earthed, but connected to the power supply (or output - depending on the topology of the output stage), we can reduce the case to heatsink thermal resistance even more.

This approach is generally considered too much of a nuisance though, because of the risk with high supply voltages, and the difficulty of mounting the heatsinks using insulating bushes, nylon screws etc. Also, it will be necessary to provide some form of shield, such as perforated steel mesh, to protect the heatsinks from becoming shorted to each other or the chassis should some object be dropped. Even the end of a lead swinging about could destroy the amplifier - not a happy thought.

11.3 - Additional Thoughts On Reducing Thermal Resistance

Some other methods also exist for reducing the thermal resistance, with one of the simplest being to use a higher power transistor, or even a simple change of case type. Use a TO-3 case or large footprint plastic package instead of a standard TOP-3 for example. A few real examples ...

TO3 package	**Watts**	**Max Temp.**	
2N3055	115W	200° C	1.5° C/W
MJ802	200W	200° C	0.875° C/W

TO3P package	**Watts**	**Max Temp.**	
TIP3055	90W	150° C	1.39° C/W (standard footprint 16x20)
2SA5358	150W	150° C	0.83° C/W (standard footprint 16x20)
MJL21193	200W	150° C	0.7° C/W (large footprint 20.5x26) (Actually TO-264 case)

The figures quoted above are typical, and will vary (sometimes significantly) depending on the manufacturer and the fabrication method used.

As you can see, simply selecting a higher power device in the same case style, the thermal resistance is lowered, since the manufacturer must make allowance for the junction to case thermal resistance to allow the higher power. In many instances, this will

probably be "automatic", since the silicon die must be larger to allow more power, and thus has greater surface area and better heat transfer.

(Thanks to Geoff Moss - my defacto editor for this information)

---

12 - Introduced Thermal Resistances

It is not uncommon for amplifier designs to use a bracket of some sort to connect the transistors to the heatsink. Where this is done, it is imperative that the minimum additional thermal resistance possible is created. I have seen designs where a simple aluminium bracket is simply bolted onto the heatsink, and everyone hopes that this will be enough. Generally, it is not. Transistors must be mounted as close as possible to the heatsink side of the bracket, since the aluminium is not a perfect conductor, and the smaller the distance the heat has to travel the better.

Any bracket used must be of the thickest material possible, and be mated to the heatsink with the greatest care. Otherwise, one is simply adding an unknown (possibly quite large) thermal resistance between the transistor's junction and the ambient air. I have heard, and seen in magazine readers' letters, complaints from amplifier builders that the transistors keep blowing for "no reason". Other more experienced (or careful) constructors have no such problems with the same design, and it is quite rare that the designers of magazine construction projects will suggest that the bracket to heatsink mounting could well be the problem.

If brackets must be used, ensure that the mating surface with the heatsink is completely flat, and free of burrs or other anomalies which will prevent the "perfect" contact. There should be no less than 1 screw (or aluminium rivet) for each square 25mm of surface area, and screws should be tightened from the centre outwards, much as one tightens the cylinder head bolts on a car engine. A thin film of thermal compound is a must, and the mating should be tested with firm pressure then inspection to ensure that the contact surface is even. If it is not, then fine valve grinding paste (available from automotive parts stores) can be used to ensure that the surfaces are matched to each other.

Simply place a small quantity on the bracket, and distribute it with your finger. Rub the two components together (bracket and heatsink) in small circular patterns, until inspection shows that the two surfaces are evenly polished. It will not hurt one little bit to use a metal polish to shine the surfaces when you are satisfied that they mate properly - the smoother the surfaces, the better the thermal contact, since surface irregularities are minimised. Thermal compound should be used sparingly - once the surfaces are mated,

the smallest amount only should provide coverage of the entire surface.

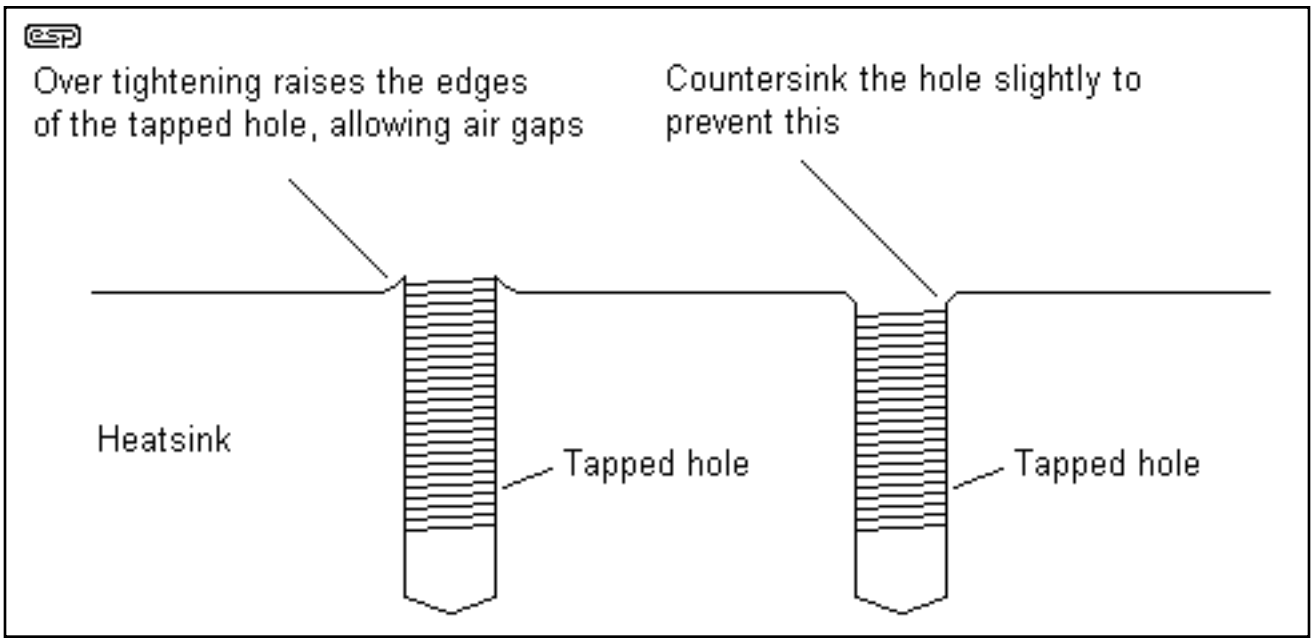


Figure 4a - Raised Edges Due to Aluminium "Stretching"

A short word about the mounting holes and screws is also called for. When a screw is tightened into a soft metal such as aluminium, it will tend to distort (or stretch) where the screw exerts pressure. On the heatsink, this will cause the metal to distort outwards, preventing the proper mating of the surfaces. After all your hard work ensuring that the surfaces were as good as you could get them, this will instantly negate your efforts. *Small* recesses in the underside of the bracket's screw holes will allow the distortion to occur, but will not lift the bracket off the heatsink. Likewise, a *very slight* recess in the surface of the heatsink will prevent the distortion from projecting above the surface - I tend to use both methods at once just to be sure. It is vitally important that any recess or countersinking is only large enough to prevent metal "stretch" - overly large recesses will do more harm than good, either by reducing the available surface area, and/ or allowing the transistor flange to be distorted.

Do **not** be tempted to tighten the screws as much as you can. They need to be tight enough so that there is good contact, but over tightening will distort the metal and will generally make the thermal resistance higher. The use of at least two washers is recommended, as this will help to distribute the pressure more evenly, and a spring washer is essential to stop the screws from loosening. If you can obtain them, 'cup' or compression washers are highly recommended, as they will maintain a consistent pressure, despite any movement ('flow') of the heatsink material. This is especially important with extruded aluminium heatsinks (i.e. most of those available) for long term thermal stability. Cast heatsinks are a completely different alloy, and distortion does not seem to be an issue.

If you use a mounting bracket, and can feel any difference in temperature between the main heatsink and the bracket, there is thermal resistance present, and more work might be needed to ensure that it is reduced to the minimum.

Remember too, that aluminium is a soft metal, and will tend to 'flow' when a consistent pressure is exerted. Over time, this may lead to the thermal resistance increasing, possibly to the point where the transistors overheat and are destroyed. For this reason, excessive tightening is never recommended, and the better the contact between bracket and heatsink, the less pressure is needed to ensure thermal resistance is kept to the minimum. Aluminium pop rivets can also be used, since they have the same rate of expansion as the heatsink and bracket (which are also aluminium), and will exert a consistent pressure which is not excessive. Care must be taken to ensure that the rivets are long enough to penetrate the heatsink, or the holding power is reduced to slightly less than useless.

Note that if you are using rivets, the bracket must be attached to the heatsink *before* the transistors are mounted. The shock from the riveter may cause damage to the transistor die.

---

13 - Flat-Pack Transistors

The TO-220, TOP-3 and other flat plastic packages have some real problems with thermal resistance, and their mounting arrangement does nothing to alleviate this - in fact the reverse is true. The problem is that the transistor junction is physically located more or less in the middle of the plastic encapsulated section, and the mounting is by a single tab, separated from the die by perhaps 10mm or so. This does not sound like very much, but the thickness of the tab is such that a significant heat difference can be measured between the two points.

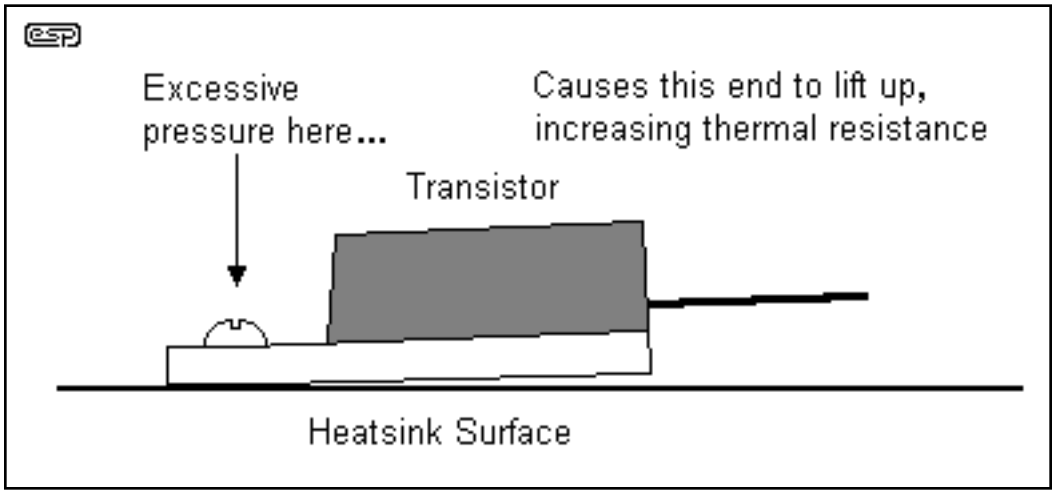


Figure 4b - One Result of Excessive Pressure

When the device is mounted, the pressure of the mounting screw is concentrated in one spot, and may even cause the remainder of the transistor case to lift up from the heatsink slightly. When maximum heat transfer is required (which is most of the time), a very worthwhile improvement can be obtained by clamping the transistor to the heatsink with a section of flat aluminium bar (or even better, a channel or u-section). This should be thick enough to not bend when the screws are tightened, and also alleviates the requirement for insulating bushes and the like, since the normal mounting screws are not needed (the insulating washer is naturally still needed). Because of the pressure exerted directly above the transistor junction, thermal resistance is reduced by a worthwhile margin, but care must be taken to ensure that the screws holding the pressure bar are tightened evenly to ensure that all devices are firmly clamped. This method has the additional advantage that any deformation of the heatsink surface due to the pressure of the screw thread will be away from the device itself, enhancing long-term reliability.

A smear of thermal compound on the top of each transistor allows the pressure bar to become a part of the heatsink - not to the degree that a smaller sink can be used, but any reduction in thermal resistance is worthwhile, however slight.

Excessive pressure is a definite no-no, as you might be strong enough to crack the transistor case if you are not careful. This is one for those with some skill with machines and the like, since we are attempting to provide the absolute maximum allowable pressure on the cases, without the risk of damage - mechanical engineering stuff, basically.

The other advantage of this technique is that it allows the use of larger screws than would normally be possible, since they no longer have to be able to fit through the little hole in the insulating bush. This means more pressure and less thermal resistance, but

only if the screws are tightened to a uniform torque so all devices have the same pressure exerted upon their cases.

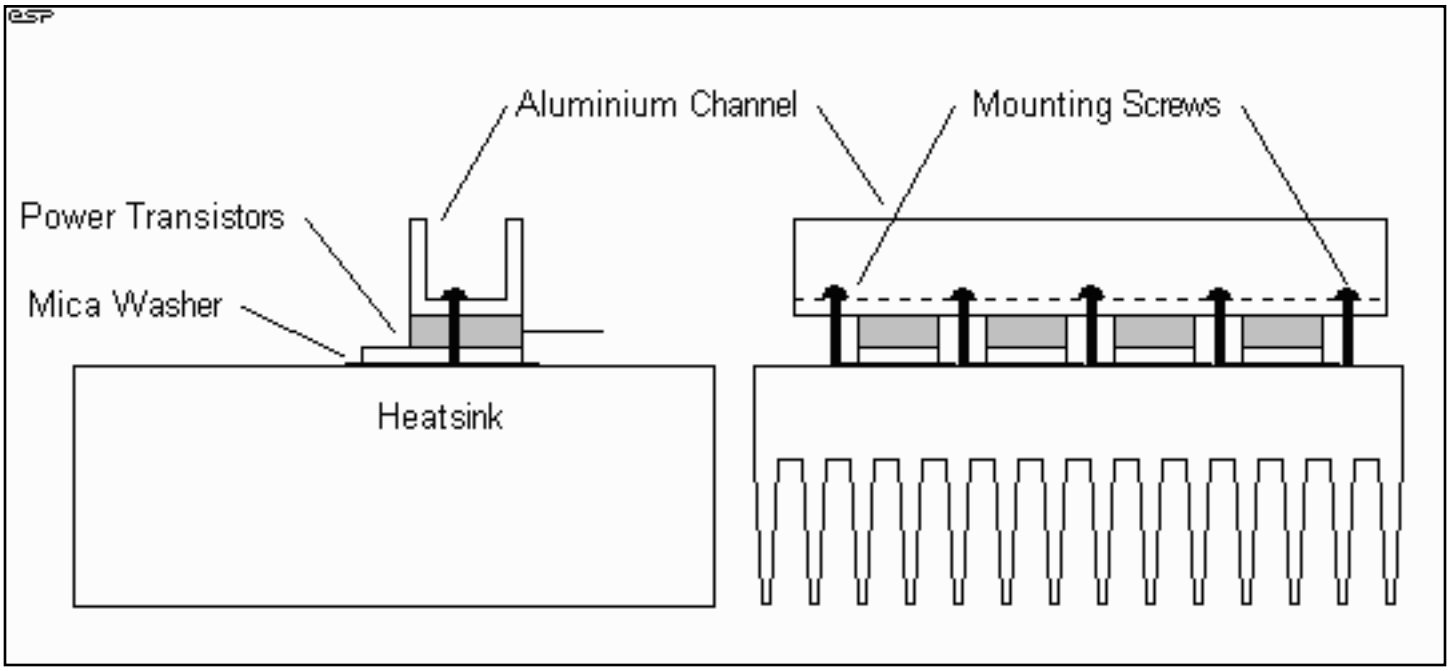


Figure 5 - Suggested Method For Mounting Flat-Packs

Figure 5 shows the general idea, with four transistors mounted on the heatsink. On the left is a side view, with an 'end-on' view of the transistors on the right. This method requires that there is some free area on the PCB, since the transistor leads will normally be mounted directly into the board. The aluminium channel section can then perform another very useful function - supporting the circuit board, so excessive vibration or flexing does not fracture the transistor leads.

In addition, some amps need a heatsink for the driver transistors, and - there it is! I am surprised that this method is not more widely used in production amps (but then again, I only thought of it a couple of years ago, so maybe they just haven't thought of it yet). If this is the case, just remember where you saw it first .

If you are a little wary, thinking that the mica washers might slip during assembly, use nylon screws to hold the transistors down loosely during assembly (don't tighten them too much, or they just snap off). These will hold everything in place while you go about the business of carefully tightening down the channel section screws. Remember that they should be all fairly even, and it is best to start from the centre, then work out towards each end to prevent any possibility of the channel section developing a buckle (however slight). If available, a torque wrench is very useful for tightening the screws (no, not the one you use for the car's cylinder head bolts :-)

14 - Where To Mount Transistors

Since heat rises by convection, one might think that mounting the transistors on the bottom of the heatsink might improve thermal performance. This is incorrect, and will actually cause the devices to run hotter. Aluminium is an excellent conductor of heat, and the idea is to ensure that the transistors are mounted in such a way that each has sufficient clearance from the others, and they are close to the geometric centre of the sink.

This method ensures that the heat from each transistor has a similar mass of aluminium to disperse its heat into, ensuring that all devices have a similar and equal chance of dispersing the heat generated.

Convection (as we normally think of it) does not apply in a solid, since it requires molecular movement on a grand scale. In a solid, the molecules merely jiggle about more, but stay in the same place.

15 - Heatsink Surface Area

There is only one thing on a heatsink that actually gets rid of heat to the surrounding air - surface area. The higher the surface area, the more heat will be disposed of. However, this is only the start, since fluid dynamics (in this case the fluid is air) also plays an important part. Make sure that you read the sections below on fin density and altitude effect - these are very important considerations in the overall design.

Heat is lost to the air by two mechanisms, and both should be maximised for best performance:

- Conduction / Convection
- Radiation

Conduction (and / or convection) requires that there is a continuous stream of air flowing past the fins of the heatsink, which means that the fins should be vertical if at all possible. Horizontally oriented fins will lose a vast amount of thermal transfer, since the air cannot flow through to the body of the heatsink. A fan solves this problem immediately, since air being blown onto the face of the heatsink has considerable turbulence, ensuring that there is always plenty of cool air at the surface of the heatsink.

Simple convection is not as effective (even for the same rate of flow of air), because of the "laminar" flow of air (where the air at the surface of the heatsink moves slower than that further away). This effect can be easily seen on a windy day. If you stay close to a wall or other large area (lying on the ground works too), it will be noticed that it is less windy than out in the open. Exactly the same thing happens with heatsinks (but on a somewhat reduced scale). Creating turbulence is an excellent way to defeat this process, but this requires fans, and fans are noisy.

Radiation requires that the surface has the maximum emissivity of heat, and this means that its colour is important. Shiny gold anodised heatsinks might look great (if you like that sort of thing), but are hopeless at radiating heat. It's no accident that the radiator in a car, or the condenser on the back of a refrigerator is matte black - not chrome plated and shiny. Matte black heatsinks are the best for radiation, and will have a significantly better thermal resistance than any other. Use of paint is to be avoided however, unless kept thin and even, as a thick layer of paint acts as an insulator, reducing the ability of all those unwanted therms to get out into the air.

This is one of the reasons that aluminium is so popular as a heatsink material - it can be anodised, and black dye is then introduced into the porous layer of aluminium oxide. This is far thinner than any coat of paint, and is very effective. Copper is actually a far better conductor of heat, but cannot be anodised, and its colour is such that it is a naturally terrible radiator. Such is life. Oxidised copper (kinda, sorta like anodising) is passably effective, but rarely used due to its cost. Table 3 shows the "emissivity" of various surfaces. This is a measure of their ability to emit infra-red radiation (heat), and a figure of 1 is as good as it gets for a passive heatsink (i.e. no heat pumps or the like). This list is not exhaustive, but is a fair indicator for the most common surface treatments.

Surface	**Emissivity**
Polished aluminium	0.05
Polished copper	0.07
Rolled sheet steel	0.66
Oxidised copper	0.70
Black anodised aluminium	0.70 - 0.90
Black air-drying enamel	0.85 - 0.91
Dark varnish	0.89 - 0.93
Black oil paint	0.92 - 0.96

Table 3 - Emissivity of Various Surface Treatments

Emissivity refers to radiation, which is only a relatively minor (typically 25% or less) but still important means of dissipating the heat. Most heat is conducted to the air at the surface boundary, and although black oil paint has excellent emissivity, it will also insulate the fins from the air. Overall, as discussed above, about the best treatment is black anodising, but matte black automotive type lacquer is also very good (IMO) - provided it is applied as *thinly as possible*.

Most high performance heatsinks have the surface of the fins ribbed, which further increases the surface area, so even a relatively small heatsink will have a radiating surface far greater that its moderate size would indicate, after all the individual fins, and the ribbing on the fins, is taken into account. Heavy application of black paint (for example) could fill the ribs, and reduce performance dramatically!

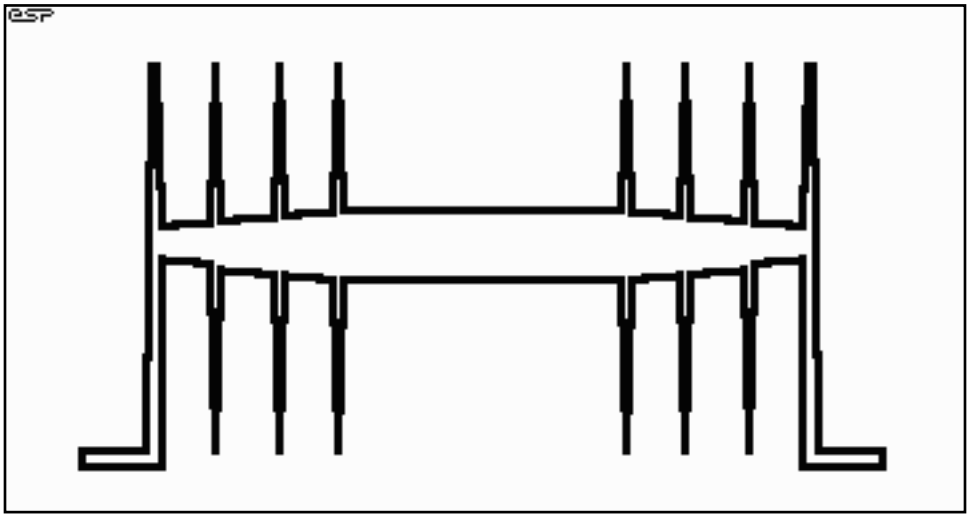


Figure 6 - A Typical Heatsink Section

The simple heatsink in Figure 6, (100 x 100 x 50mm high), has a base surface area of 200 square cm, and when the fins are added, this increases to 1000 square cm. Note the way the base tapers away from the centre - this saves metal, but also provides maximum thermal inertia where the transistors mount, and matches the heat carrying capacity at the extremities to the expected heat flow. In reality, such a heatsink would have more fins, and if they were ribbed, the resultant final unit could have a total surface area of perhaps 1500 square cm. This is the equivalent of a single plate over 270mm square, but will have far better performance because the thermal resistances are kept to a minimum by reduced distances. (Remember that both sides of the metal are in contact with the air, so the total surface area is double that expected.)

Note that when dealing with heatsink temperature, it is almost always referred to as temperature rise - the number of degrees above ambient that the heatsink will reach. If ambient temperature is lower than the "standard" 25°C, so too will be the heatsink temperature (and vice versa, of course). Temperature rise is a constant in the equations - this is where we must be careful when the ambient temperature is high. A car in the sun can have an internal temperature of 50°C quite easily, so a temperature rise of 25° will see the heatsink at 75°C.

The thermal resistance of a heatsink is determined by :-

- Its size
- The number of fins
- The total surface area - a ribbed fin design will be smaller than one using flat fins for the same thermal resistance
- Thermal resistance within the aluminium itself - the actual alloy used, its thickness, etc
- The colour - matte black is best, and other colours are all worse
- The orientation of the fins - vertical is best, any horizontal position is worse
- Availability of unrestricted air flow
- Its temperature!

According to the 'quick and dirty' formula on Harry's Homebrew Homepage (see link below), the thermal resistance is approximately equal to

$$\text{Thermal Resistance} = 50 / \sqrt{A}$$ Where A is the total surface area in cm^2

Using this formula on the above heatsink (without ribbed fins), gives a thermal resistance of about 1.58° C/W or 1.29° C/W with the ribbed fins. This is not an insignificant difference.

The same heatsink used in my heatsink calculator (see Downloads page) gave a thermal resistance of 1.12° C/W, assuming flat fins. I used a fin height of 50mm, depth of 100mm, and told it the heatsink has 8 fins. This assumed an ambient temperature of 25° C, and a maximum heatsink temperature of 40° C. This seems more or less realistic compared to manufacturer's data for similar heatsinks.

Manufacturers' ratings are not usually very specific, but we can safely assume that a given thermal resistance is the best obtainable, and probably assumes that the heatsink is in free air (no nearby casings, optimal fin orientation, etc). In most cases, a heatsink

has to be used in a real-life situation, which means that you might need to add anything from 10% to 50% to the quoted figure, so a 1° C/W unit could end up as anything from 1.1° to 2° C/W, depending upon its mounting and other impediments to airflow.

Where possible, make sure that the heatsink is firmly attached to the case, which can make a reasonably thick aluminium chassis (for example) a significant part of the heatsink. This can improve performance markedly if good thermal bonding is used between case and sink. I do not recommend that you include the case in any calculations unless it is large compared to the heatsink, but if you do, do *not* count the inside of the case as part of the surface area. This is sealed (or semi-sealed), and the air will not have the free movement needed to remove the heat from the inner surfaces.

The last point is likely to cause some confusion - after all, thermal resistance is a constant based on the other factors, right? Wrong.

Let's imagine a heatsink with a heat load of 10W. After careful measurement, we determine that the temperature rise is 10° C (i.e. 10° above ambient), so the heatsink has a thermal resistance of 1° C/W. We will almost certainly decide that a 10° rise is probably a bit restrained, and in fact we can allow a temperature rise of 20° C, and we may decide to allow for a maximum 25° ambient (not much use in outback Australia during summer, but it will do for this exercise).

Allowing an additional 10° brings the thermal resistance down to 0.89° C/W, and if we decided that even a 35° rise were acceptable (80° C heatsink temperature referred to a 25° ambient), the thermal resistance drops further, to 0.73° C/W.

This apparently anomalous behaviour is actually all completely normal, and the heatsink is actually obeying the Second Law of Thermodynamics perfectly (feel free to look that up if you don't know it already).

In fact, the term 'thermal runaway' - where a transistor gets hotter so conducts more current, so gets hotter, so ... - is a misnomer in terms of physics. At some point, the system will stabilise, as it must. Unfortunately, this will probably involve temperatures in excess of what the transistor junction can tolerate, so it (or they) will be destroyed.

The heatsink discussed above has a thermal resistance of only about 0.45° C/W if a temperature rise of 200° C were acceptable (which it absolutely is *not*).

This is no different from any other situation where there is a difference in potential, pressure, or anything else. The greater the difference, the greater the flow rate. Now you

know that the thermal resistance is *not* a constant - it is inversely proportional to the temperature difference between heatsink and ambient (this is a gross simplification, but will suffice for the time being). If heatsink and ambient temperatures are the same, the heatsink has an infinite thermal resistance!

---

16 - Altitude

Since the density of air varies with altitude, so does the efficiency of a heatsink. I wonder how many consumer equipment makers take this into consideration? As can be seen from the table below, the effects are not insignificant.

Altitude (Metres)	Altitude (Feet)	Derating Factor
0 (sea level)	0	1.00
1,000	3,000	0.95
1,500	5,000	0.90
2,000	7,000	0.86
3,000	10,000	0.80
3,500	12,000	0.75

Altitude Derating Factors

The altitude effect should be considered in all cases, as can be seen. While the air temperature of an indoor environment is normally controlled and is not affected by the altitude change, the indoor air pressure does change with the altitude. Since many electronic systems are installed at an elevated altitude, it is necessary to derate the heat sink performance mainly due to the lower air density caused by the lower air pressure at higher altitude. The table shows the performance derating factors for typical heat sinks at high altitudes. For example, in order to determine the actual thermal performance of a heat sink at altitudes other than sea level, the thermal resistance values read off from the performance graphs should be divided by the derating factor before the values are compared with the required thermal resistance.

Example: A 1° C/W heatsink would become 1.16° C/W at an altitude of 2,000 metres, or 1.25° C/W at 3,000 metres.

---

17 - Fin Density

Although it would seem reasonable to assume that the more fins per unit area the better, this is not always the case. Closely spaced fins will not be able to dispose of the heat well in a standard convection (i.e. not fan forced) heatsink. This is partly because of the airflow that will actually be able to establish itself within a confined space, and partly because the fins will tend to radiate much of the heat to adjoining fins, which helps to stabilise the temperature, but does little to dispose of the heat.

The maximum distance between fins is dependent on the depth or height of the fins - deep finned heatsinks will need more space between adjacent fins than a shallow design unless fan cooling is used. As shown in the following table, the minimum spacing is determined by fin depth and airflow.

Fin length (millimetres)	**75**	**150**	**225**	**300**
Airflow (metres / sec)	**Fin Spacing (mm)**			
Natural convection	6.5	7.5	10	13
1.0	4.0	5.0	6.0	7.0
2.5	2.5	3.3	4.0	5.0
5.0	2.0	2.5	3.0	3.5

Fin spacing (in mm) versus flow and fin length

The average performance of a typical heat sink is linearly proportional to the width of a heat sink in the direction perpendicular to the airflow, and approximately proportional to the square root of the fin length in the direction parallel to the flow. For example, an increase in the width of a heat sink by a factor of two would increase the heat dissipation capability by a factor of two, whereas doubling the depth or height will only increase the heat dissipation capability by a factor of 1.4. Therefore, if the choice is available, it is beneficial to increase the width of a heat sink rather than the length of the heat sink fins. Also, the effect of radiation heat transfer is very important in natural convection, as it can be responsible of up to 25% of the total heat dissipation. Unless the heatsink is facing a hotter surface nearby, it is imperative to have the heat sink surfaces painted or anodised to enhance radiation.

The reason for the difference is simply that as the height is increased, the air at the top of the heatsink is hotter than that entering at the bottom. If the fin depth is increased, there is more mutual radiation between fins, and as the spacing is reduced, mutual radiation increases again. Airflow is also restricted because of the smaller physical area

for air to pass, since more of the available space is occupied by the heatsink itself.

---

18 - Water Cooling

I was going to stay away from this completely, but it is worth a small paragraph at least. Water just happens to be the best heat removal medium known, with a specific heat of 1, it requires more energy to raise a gram of water by one degree C than any other material.

If extremely high power is the goal (and a bit of plumbing is Ok), a water cooled heatsink is the ideal. Provided the thermal resistance from junction to heatsink is minimised, a minuscule heatsink with only a moderate water flow will remove prodigious amounts of heat. Although uncommon for audio amplifiers, water cooling has been used for many years for cooling high power radio transmitters and the like.

However, it must be admitted that few audiophiles will go to the trouble of installing special plumbing to cool their amplifiers - it would be cheaper to put them in another room and use fan cooling, and a lot more convenient.

Having said that, the use of water cooling is gaining in popularity, and quite a few commercial H_20 cooling systems are available for microprocessors. As component density increases in the ICs, it becomes harder to get the heat out efficiently, so expect to see that side of the market expand in the next few years. While it must be admitted that water cooled heatsinks are the best possible choice for amplifiers, this is not an area that can be expected to grow, since it is just too expensive and unwieldy to implement. Finding a silent running pump is another hurdle, and although the plumbing is not at all difficult (low pressure systems are all that's needed), it still has considerable nuisance value - the cables in a typical system are bad enough, let alone having a plumbing system that must be 100% watertight. Of course, then there is the radiator, pump and fan to be considered - they have to live somewhere, and can be expected to have a SAF (Spouse Acceptance Factor) of perhaps -20dB ("I don't know what you're planning to do with *that* horrible looking thing, but if it comes in here, I'm moving out!" - a fairly typical response, I would suggest).

Still, I may do a water cooled heatsink project at some time, just for the fun of it ;-)

---

19 - Heat Pipes

Heat pipes are not common, and will not be discussed at any length in this article. They are expensive, and are most commonly used where a hot component must be cooled in a confined space. The heat pipe is actually two heatsinks, joined by a pipe containing a refrigerant. Natural convection is used to move the refrigerant from the hot area (the component to be cooled) and the main heatsink, which may be mounted some distance away.

For those who are interested in this technology, there are quite a few articles on the web that discuss heat pipes, and I suggest a web search to locate them.

There are some newer versions of the same basic principle now available. These make use of "micro-channels" to dramatically increase the surface area at the hot end, and allow the use of more environmentally friendly coolants. However, it is to be expected that these will be expensive, and outside the budget of most home constructors.

---

20 - The Bottom Line

As can be seen from the above, for a given heatsink, the most critical part of the whole thermal resistance equation is actually the transistor itself. If the thermal resistance of the active devices can be reduced, the heatsink requirements are lessened, resulting in a pleasing increase in reliability, and an equally pleasing reduction in the cost of the necessary heatsink.

The thermal resistance from transistor case to heatsink is also very important, and proper mounting technique can result in a significant improvement over the "just slop some grease on and bolt it down" approach.

If these two thermal resistances can be reduced enough, the heatsink itself can be quite a bit smaller than would otherwise be the case for the same (or even lower) transistor operating temperatures.

Always remember that manufacturers' data on heatsinks is under ideal conditions, and is usually measured at a temperature of up to 50 to 75°C above ambient - although this is rarely stated.

---

21 - Other Heatsink Information

I have included a heatsink calculator spreadsheet, which appears to work quite well. Go

to the Downloads page to get a copy, or you can just click here for a download.

A very useful link for those who are put off by the cost of large heatsinks:

HeatSinks Very good drawings, showing the complete process for building a heatsink. All measurements were in imperial - i.e. inches, but now include metric.

Harry's Homebrew Homepage has some good info on estimating the thermal resistance of a heatsink (such as the one you make yourself), and for determining how big a heatsink needs to be for a given thermal resistance.

A most excellent document has found it's way into my clutches! It is an application note from International Rectifier (irf.com) and I have linked it here for your convenience. Although it refers to a specific case style, the information is generic enough to ensure that you have a better all-round understanding of the subject.

Another excellent document that describes mounting methods and techniques can be found at the On Semiconductor website. This describes the subject in considerable detail. This document was only recently pointed out to me, and was not used as a reference for this article.

Please Note: With all articles of this nature, there will be some information that appears to be in conflict with other data you will see, either in this article, or other application notes or publications. This is perfectly normal, as no two engineers will ever be in complete agreement, but the basic data correlates very well, with only a few minor points of difference.

---

22 - References:

1. General Electric Transistor Manual
2. National Semiconductor Voltage Regulator Handbook
3. Jaycar Electronics Engineering Catalogue
4. Farnell Components Catalogue
5. International Rectifier Application Note AN-1000
6. Aavid Thermal Technologies, Inc. Laconia, New Hampshire

Main Index

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999 - 2006. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Revision History: 23 May 02 - corrected link to Harry's Homebrew page./ 18 Feb - fin density and altitude effect, plus other small additions./ 12 Feb 2001- added info on mica "splitting"./ 07 Aug 2000 - Added an-1000.pdf./ 08 Apr-Added info on heatsink calc spreadsheet./ 05 Feb-added "additional thoughts" (Geoff Moss contribution)./ 28 Jan 2000-Corrected several errors, added links for additional info and more info on emissivity etc. 16 Aug 1999-Added new Figure 1 and explanation of electrical analogue./ 16 Jun 1999-Initial publication

Elliott Sound Products **Noise In Audio Amplifiers**

Copyright © 1999 - Rod Elliott (ESP)

Last Updated 04 April 2006

Articles Index

Main Index

Introduction

It is the purpose of this article to give the reader an introduction to understanding noise in electronic circuits, why it happens, and how to read noise specifications. The latter are not usually explained in a way that makes sense to the uninitiated, so it is hoped that this article will assist those trying to make some sense of it all.

Noise has enormous nuisance value with sensitive (i.e. high gain) circuits, but the information provided by most IC and transistor makers does not always make the choice of the most suitable device easy. Certainly, there is copious information, but explanations of what it means and how to apply it are few and far between.

This short article will hopefully clear up some of the confusion. By nature, it is rather more technical than I generally prefer, but this is unavoidable for a passably thorough understanding of the subject.

In this context, noise refers only to circuit noise, and not hum, buzz or other extraneous outside influences. These are usually the result of bad (or misguided) earthing practices, or signal wiring running close to magnetic field or harmonic generating items such as transformers and bridge rectifiers. Also to blame can be radio frequency interference, which will often cause problems if adequate (and appropriate) precautionary measures are ignored.

Noise Figure

Noise is inherent in all electronic circuitry, and comes in three basic flavours:

- Thermal noise
- Shot noise
- Flicker (l/f) noise

Thermal noise is "white" in character (has a constant energy per unit bandwidth) and is generated by the thermal agitation of electrons in a conductor. The typical sound is hiss, hopefully at a low level so that it does not intrude on the programme material. It is calculated using Nyquist's relation:

$V^2R = 4k * T * R * f$

Where ...

V^2R = mean square noise voltage
k = Boltzmann constant (1.38×10^{-23})
T = Absolute temperature (Kelvin)
R = Resistance in ohms
f = Noise bandwidth in Hertz

Without even performing any calculations, we can see that the noise from a resistor is proportional to its resistance and temperature. Operating resistors at elevated temperatures in input stages is obviously undesirable, as are high resistance values. This also applies to any other resistive device, such as the voice coil of a dynamic (magnetic) microphone, the coils of a guitar pickup, or a vinyl disc pickup cartridge.

Shot noise is also white, and is inherent in semiconductor devices. It is primarily caused by the changes in molecular energy levels as a semiconductor device conducts.

Flicker noise is a low frequency effect, and as such is not so much of a problem with audio circuits. It becomes worse as frequency is reduced, and this can be seen in some data sheets. At the lower frequency extremes, the noise level increases.

Explanations

Before we continue, there are a couple of terms that need explanation.

Firstly, the term 'dBv' refers to decibels relative to 1V RMS, and 'dBu' means decibels

relative to 775mV. This is also known as dBm, and relates to the old convention of 1mW into a 600 Ohm load. This was common in telephony (and still is in some cases), but is of little relevance to audio applications. However, we are stuck with it. 0dBv is equivalent to +2.2dBu.

Secondly, noise is commonly referred to the input of an amplifier circuit. This allows the instant calculation of output noise by simply subtracting the dB figures. So an amplifier with an 'Equivalent Input Noise' (EIN) of -120dBu having a gain of 40dB will have an output noise of -80dBu (120 - 40). This is the equivalent of 80dB Signal to Noise ratio (S/N) relative to 0dBu. Many equipment manufacturers will state S/N relative to maximum output, thereby gaining perhaps 10dB better figures. This is actually meaningless, since no-one will operate equipment at the maximum level, and the average will be considerably less.

Thirdly, it is commonly accepted that the minimum theoretical input noise (EIN) for any amplifier is -129dBu. Although not explicitly stated, this implies that the input will be terminated with a resistance. Typically, a 200 ohm source resistor will give this figure at 25°C. Sometimes, a short circuit is used instead, and this gives better apparent noise performance. A short circuit is actually meaningless though, since no real-world signal sources have zero impedance. Some do come close though, so the amplifier should always be terminated with an impedance that matches (as closely as possible) the output impedance of the signal source. This should be stated in any specification.

This means that a perfect (as noiseless as it is possible to be) amplifier with a gain of 40dB and a 600 ohm source impedance will have an output noise level of -89dBu, and if the gain were to be increased to 60dB, then output noise will be -69dBu.

It is the nature of noise that it does not add in the same way as two equal frequencies. Because of its random nature, two equal noise voltages will increase the output by only 3dB, not 6dB as might be expected. As a result, we can be reasonably sure that it is the input noise of the most sensitive section of a preamp that will set the final limit to the signal to noise ratio of the entire unit.

The way the noise figure of an opamp is commonly described is something else that needs a little explanation, since it is hardly specified in terms that most constructors will be able to relate to. The data sheet telling you that the "noise figure is 5nV/√Hz" is not very friendly. To get this into something we can understand, first we need to take the 'square root of Hz' and make some sense of it. The audio bandwidth is taken as 20Hz to 20kHz, so the square root of this is ...

√20,000 = 141 (it is not worth the effort of subtracting the 20Hz, so 141 is close enough)

With a noise figure of 5nV / √Hz, the equivalent input noise (EIN) is therefore ...

5nV x 141 = 707nV

If we assume a typical gain of a sensitive microphone stage (for example) as 100 (40dB) and an output level of 1V (0dBv), this means that the output noise equals the input noise, multiplied by gain. Signal to noise ratio can then be calculated ...

707nV x 100 = 70.7uV (EIN = -120.8dBu)
Signal to noise (dB) = 20 x log (1V / 70.7uV) = 20 x log (14144) = 83dB

We can also calculate this using dB alone.

EIN = -120.8dBu
Gain = 40dB
S/N = 120.8 - 40 = 80.8 (ref 0dBu), or 80.8 + 2.2 = 83dB (ref 0dBv)

For low level preamps (such as microphone or moving coil phono pre-amplifiers), it is common to specify the EIN only, allowing the user to calculate the noise for any gain setting, since it changes as the gain is varied. The same amplifier as above with unity gain will have a theoretical signal to noise ratio of 123dB (relative to 1V). All of this assumes that the passive components (especially resistors) do not contribute any noise. This is false, as any device operating at a temperature above 0K (zero Kelvin, absolute zero, or about -273° Celsius) generates noise, however the contributions of passive components are relatively small with quality devices provided resistance is kept as low as possible, and voltages minimised.

Remember that a 'perfect' amplifier (contributing noise at the theoretical minimum possible), will have an equivalent input noise of -129dBu .This means that with a gain of 60dB, the best possible signal to noise ratio will be 69dB relative to 775mV (or 71.2 ref 0dBv).

As an experiment, I built a three opamp precision microphone preamp using 1458 opamps (equivalent to a dual uA741). These have a noise input figure of about 4uV - this translates to about 30 to 35nV / √Hz, or nearly 20dB worse than the NE5534A. With a gain of 46dB (200), the circuit managed a signal to noise ratio of 65dB, referred to 0dBv (1 Volt RMS). The apparently better than expected S/N ratio is because the

bandwidth was so limited because of the opamps I used for this test.

I measured a S/N ratio of better than 80dB (about 82dB) again at a gain of 46dB using LM833 opamps (dual version of the NE5534). When I say that I measured this, it was with extreme difficulty. Because of the low noise, my test instruments were at their limits, so I had to guess a bit. The theoretical 'best possible' at this gain is 85.2dB referred to 0dBv, or -83dB ref. 0dBu.

Search carefully for devices with low noise for sensitive circuitry, and make sure they also have the bandwidth needed to achieve high gains. LM833 or NE5532 dual opamps are an excellent choice for low noise, but they also have wide bandwidth and can be troublesome to keep stable. Do not be tempted to use lesser devices, since their bandwidth is too limited - the 1458 was 3dB down at only 8kHz, and died rapidly after that.

In some cases, it will be found that better noise performance can be obtained using discrete opamps - built using individual components. A common technique for low noise is to select transistors based on their noise data, which will indicate the optimum collector current for a given source impedance.

Then, by using multiple devices in parallel, the noise is reduced further. Two transistors in parallel will have a noise level 3dB better than a single device. Using four will reduce this by another 3dB, and eight will give a further 3dB reduction. This is the theory behind it, but of course it will never be as good as ideal theory might indicate. It is generally considered (based on the many such designs I have seen) that between 2 and a maximum of six devices in parallel will achieve the best overall compromise. Project 25 shows a couple of designs using this method, and has some descriptive text explaining the two (very different) techniques. Project 66 gives the circuit diagram for a microphone preamplifier that uses a discrete front end to obtain low noise. No devices are paralleled as such, although the two sections appear in parallel to the following opamp.

Other Stuff

For all resistors in low noise input circuits, you absolutely, positively, must use 1% tolerance resistors, which will be metal film for lowest noise.

Many noise tests are performed using A-weighting, which introduces a filter prior to measurement. The theory of this is that it compensates for the ear's natural rolloff at low and high frequencies, and makes the measurement 'meaningful'. While the idea is quite sound in principle, I do not believe that this should be done, as not everyone is scrupulous about stating that this technique has been used, so results can be very

misleading. An A-weighting filter is described in the ESP Project Pages (Project 17), along with an extensive description of the theory behind this practice.

Noise In Digital Equipment

For each digital bit, the relative noise floor is lowered by 6dB. A 1 bit system is of little use, and it is necessary to go to a minimum of 8 bits before even ordinary speech is intelligible - not acceptable, but intelligible. This gives a noise floor of 48dB - about what you would expect from the modern telephone system. Even there, speech is digitised at 16 bits (using an 8kHz sampling rate), and then compressed digitally to 8 bits.

(BTW, Worldwide, there are two different digital compression systems used - A-Law, used by all European countries, much of South America, Australia and New Zealand etc. u-Law (as in mu, the Greek letter) is used in the US, Canada and Japan. Prior to digital to analogue conversion, the signal is returned to the 16 bit format. This has nothing to do with noise, I just thought I'd mention it.)

Some digital answering machines use 8 bit digitisation, which explains why they sound so disgusting. Even though the noise floor is (barely) low enough, there is an insufficient number of discrete levels to faithfully reproduce speech. 8 bits only provides 256 discrete levels, and it has been generally accepted that anything less than 12 bits is unacceptable (4096 discrete levels). This is easily verified by recording something on your PC at the various available bit rates and making a comparison.

$$\text{Minimum Digital Noise} = 20 * \text{Log}_{10}(\text{number of discrete levels})$$

When professional digital recording systems were first introduced they were 16 bit. Although this gives a theoretical noise floor 96dB below the maximum level, in reality 90dB was more likely. If maximum level corresponds to +4dBu, this indicates a noise level for each digital channel output of -86dBu.

$$\text{Noise} = 20 * \text{Log}_{10}65535 = -96.33\text{dB}$$

In the sound recording industry, the relative differences between analogue tape and digital recording must be considered. In an analogue machine, it is the tape that clips, which it does in a 'soft' manner, introducing predominantly low order harmonics. These are relatively inaudible, provided the duration is kept short (1ms or so), such as on transients.

A digital system by comparison clips suddenly and with great clarity, and it is essential to leave sufficient headroom to prevent this. If 10dB of headroom is left below maximum level to allow for transients (I would suggest this as a workable minimum), then this implies that the noise level actually present at the output of the digital playback system is -80dBu, relative to nominal (average) playback level.

Many digital systems now have 20 bit or greater resolution, although generally only 20 bits is achieved in practice. This reduces the theoretical noise floor by a further 4 bits, or 24dB. Therefore the noise level at the output of such a machine should be -110dB. Allowing the same 10dB headroom rule as above, this gives a final output noise figure of about -100dB. It is possible in many cases that the associated analogue circuitry within the digital system will be worse than this figure, so the final noise figure is somewhat unpredictable.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 1999-2006. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Updated 04 Apr 2006 - Modified page layout, added explanatory text./ 29 Jan 2000 - Corrected typo./ Page created and copyright (c) 12 Dec 1999.

The Audio Pages

Elliott Sound Products **Opamp Alternatives**

Copyright © 2005 - Rod Elliott (ESP)
Page Created 10 April 2005

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Single Transistor Amplifiers
  - 2.1 How the Circuit is Designed
  - 2.2 Single Transistor Summary
- 3.0 Two Transistor Feedback Pair
  - 3.1 Two Transistor Feedback Pair Summary
- 4.0 Buffers
  - 4.1 Non-Inverting Buffers
  - 4.2 Inverting Buffers
  - 4.3 Buffers Summary
- 5.0 Tone Control
- 6.0 Other Solutions
- 7.0 Conclusion
- 7.0 References

1.0 - Introduction

In the field of audio, there are many people who, for one reason or another, dislike

opamps. In some cases this is almost a passionate hatred, despite the vast number of extraordinary opamps available. Some have been around for a long time, and although they replaced the circuits described in this article in 99% of all equipment, it is worth looking at the circuits that were used in the pre-opamp (but post valve/ tube) years.

Of these circuits, Philips was responsible for many of the two transistor feedback amplifiers that were commonplace in high end equipment of the day. There is no denying that the performance of some of these circuits is/ was very good indeed, but it is exceeded (in many respects) by even mediocre opamps.

This article cannot even hope to show all the variations, but the most common circuits are examined. This is an excellent way to learn more about transistors, how they work, and what you can do with them, and as such is recommended reading for anyone who is learning electronics or wants to brush up on what is now an 'ancient technology'.

For all examples, an output voltage of 2V RMS will be used - this is for no other reason than to be able to show the difference between the circuits and to give a reasonably good idea of their relative performance. For the same reason, a supply voltage of 18V was chosen for all circuits that follow (and including the audio stage in Figure 1). Likewise, the gain has been set to 10 (or as close as standard value resistors will allow). This is a voltage gain of 20dB. All circuits are loaded by 10k, again for repeatable results. This is also a reasonably representative load - perhaps a little on the low side, but chosen to give the most realistic result available from the SIMetrix [1] simulator (or any simulator for that matter).

2.0 - Single Transistor Amplifiers

The simplest of all transistor amplifiers, these were used commonly in the very early days of transistor applications. They are very limited even today, and were much more limited when only germanium transistors were available. Unlike a valve (vacuum tube), input impedance is relatively low - this caused many of the early designs some grief because so many of the sources were high impedance.

There are ways around anything, but noise performance was usually rather ordinary - in many cases worse than that obtained from valve circuits. However, the overall convenience of equipment that could be operated from a single (low voltage) battery was just too overwhelming, and the days of the valve were numbered when Sony released one of the first all transistor portable radios. This early radio was reliable, and had surprisingly good performance (germanium transistors, remember). Yes, I did have

one

The circuit in Figure 1 became the standard for single transistor amplification stages. There are other variants, but they either have poor thermal stability or poor tolerance for different transistor gains (or both).

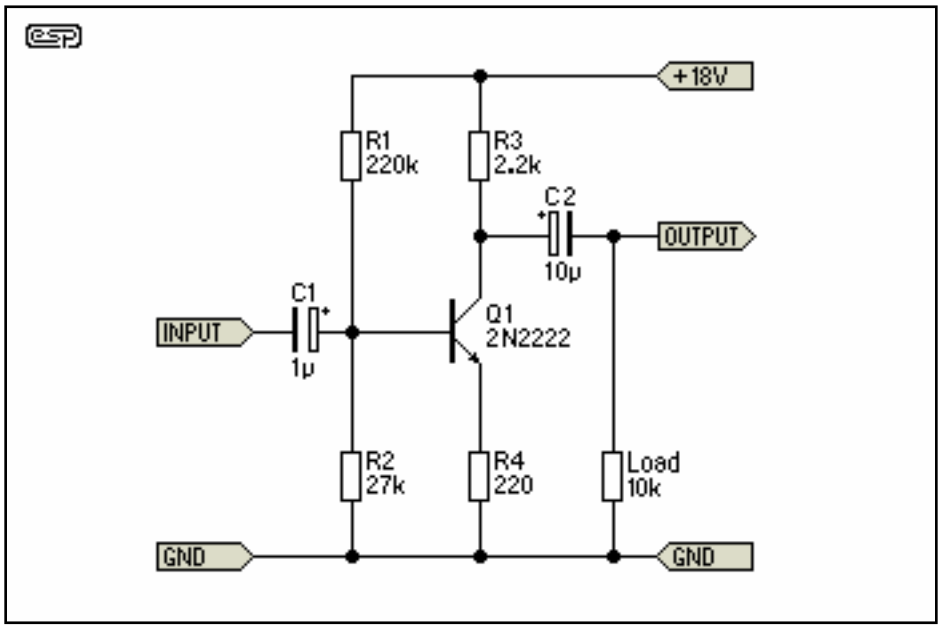


Figure 1 - Single Transistor Amplifier

The drawing above shows a capacitor coupled AF (audio frequency) stage. Although transformer coupling was also used on occasion this was usually restricted to RF circuits. There were some instances where transformer coupling was used for audio, but the demands for lower cost and better performance saw the demise of transformers in this application. Transformer coupled audio stages will not be covered.

As always, these 'simple' amplifiers are almost always biased so that the collector voltage is about ½ the supply voltage, so in the example above (AF only) the collector voltage is set to about 9V.

With the standard load of 10k, the Fig 1 amp has a total harmonic distortion of 0.218% - not too shabby, but several orders of magnitude worse than even an 'ordinary' opamp. The output FFT spectrum is shown in Figure 2 - low order harmonics predominate, and it is worth noting that 1uV referred to 2V peak output is -126dB. This is well below the noise floor that can be expected for this type of circuit.

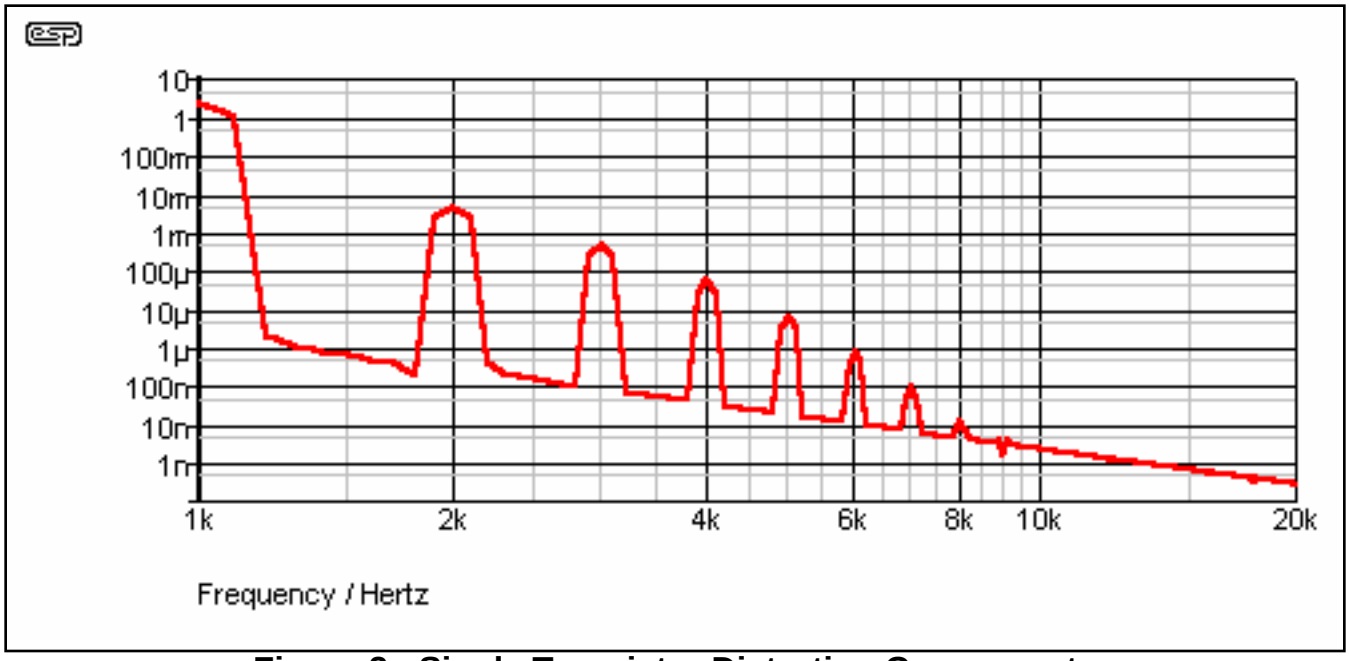


Figure 2 - Single Transistor Distortion Components

The basic amplifier has many disadvantages. Input impedance is rather low, and is only 16k for the circuit as shown. Output impedance is equal to the value of the collector resistance (2.2k in this case), and the local feedback provided by the emitter resistor does not affect this. Stage gain is (approximately) equal to the collector resistance divided by the emitter resistance. For the example shown the gain is actually only 7.95 - somewhat less than expected.

Two small traps for the unwary -

- in nearly all simple transistor amplifier designs, the effective collector load resistance is actually the value of the physical collector resistor in parallel with the external load impedance - in the case of the Fig 1 amp, this works out to about 1.8k when it is loaded with 10k as shown.
- The stage is inverting - a positive going input produces a negative going output (and vice versa, of course). Unless an inverted signal is acceptable, all such designs must be used in even numbers to provide a non-inverted output.

2.1 - How the Circuit is Designed

The design process for this amp is simple enough to explain (unlike most of the others shown below), and it gives a good insight into the way to approach analogue design. The key factor is the value of the collector resistance and the design voltage that will be across it under quiescent conditions. In this case, the resistor is 2.2k and it will have 9V across it, so ...

I = E / R = 9 / 2.2k = 4.09mA (4mA close enough).

The next step is to decide on the gain - in this case I chose a gain of 10. The gain is (in theory) determined by the ratio of the collector load (RL - resistor R3) to RE (the emitter resistor R4), so ...

Av = R3 / R4 = 2.2k / 220 = 10

As we saw though, the gain is less than this, so what went wrong? Two things - firstly (as noted above), the collector resistance (RL) is equal to the parallel combination of the actual resistance used, and the load resistor. Secondly, there is an intrinsic resistor inside the transistor - commonly referred to as re (literally 'little r e'), and its value is inversely proportional to the emitter current (IE) determined by

re = 26 / IE (mA) = 26 / 4 = 6.5

Recalculating the collector load resistance (the 10k load in parallel with the 2.2k collector resistor) ...

$RL = R3 * R_{load} / (R3 + R_{load}) = 1.8k$

Now we can recalculate the gain using the proper load resistance and the total emitter resistance ...

Av = RL / RE + re = 1.8k / 220 + 6.5 = 7.96

That is almost exactly what was measured, so the design process actually does work

Although the gain is less than the original target of 10, we shall accept that anyway for the purposes of this exercise. The next task is to determine the value of the bias resistors (R1 and R2). The first stage is to find out the lowest current gain for the selected transistor - let's assume that to be 200. The base current is equal to the collector current divided by the transistor gain. That works out to ...

$Ib = Ic / h_{FE} = 4 / 200 = 0.02mA$ (20uA)

The general rule is that the bias circuit should carry between 2 and 5 times the base current, so we are looking for somewhere between 40 to 100uA through the bias

network (this may go as high as 10 times if greater predictability is needed). Before we can start on that, we need to know the base voltage. Again, we need only Ohm's law and a simple addition to determine this. Knowing that the emitter current is 4mA (plus the base current of 0.04mA which can be ignored for high gain transistors), the emitter voltage will be ...

$$Ve = Re * Ie = 220 * 4mA = 880mV$$

The base voltage is therefore the sum of the emitter voltage and the base to emitter voltage (typically taken as between 600 and 700mV). Assuming 600mV, the base voltage should therefore be ...

$$Vb = Ve + Vbe = 880 + 600 = 1480mV\ (1.48V)$$

R1 can now be found. The voltage across it is the supply voltage less the base voltage, the current through it is worked out using Ohm's law (again). We shall start with a value that is 10 times the collector resistance (always a good starting value), so let's try 220k ...

$$I_{R1} = V_{R1} / R1 = (18 - 1.48) / 220k = 75uA$$

That's 3.75 times more than the base current, so should be fine. Now, we only need to work out the value for R2 ...

$$R2 = Vb / (I_{R1} - Ib) = 1.48 / (75 - 20) = 26.9k\ (27k)$$

And that's it. The exercise is rather tedious (much more so before the advent of the calculator), and has to be re-calculated each time you change the supply voltage or the stage gain. The impedances are fixed by the component values and there is little you can do about it. It is possible to bypass R4 with a large value cap to get more gain (which can be very high even for a single stage), but then distortion goes through the roof.

Output impedance is set by the value of the collector resistance (2.2k in this case), and input impedance is determined by the parallel combination of R1, R2 and Zb (the input impedance of the transistor's base). We need to know this so we know how much this circuit will load the source ...

$$Zb = (RE + re) * h_{FE} = (220 + 6.5) * 200 = 45.3k$$

The total input impedance is the parallel combination of all three impedances ...

$$Zin = 1 / (1 / R1 + 1 / R2 + 1 / Zb) = 1 / (1 /220k + 1 / 27k + 1 / 45.3k) = 15.7k$$

Since we measured (well, I did anyway) the input impedance at about 16k, it looks like everything has fallen into place nicely.

It is worth noting that 'little re' can be ignored if its value is less than one tenth that of RE (the external resistance). For what it's worth, when RE is bypassed by a cap re becomes significant. I used a value of 2,200uF, and the circuit gain becomes roughly 220, but distortion rises rapidly - try 8.25% at 2V RMS output! The vast majority of that distortion is the result of the modulation of re as the collector voltage varies - as must the current through the load resistance and hence the emitter. This is explained for interest value - feel free to examine this yourselves, you will learn a great deal in the process.

You may have noticed that the vast majority of the design process involved only Ohm's law. It is commonly (but mistakenly) believed that all sorts of intricate maths are needed to design basic circuits. This is not the case at all.

2.2 - Single Transistor Summary

For many years, the single transistor stage was the fundamental building block of 'solid state' audio, and it must be admitted that these circuits did not compare particularly favourably (in some areas) to valves. Although the gain was more predictable, distortion was about equal for low levels, but the valve circuit with its much higher voltage had much greater headroom as well.

The advantages of the circuit were that it consumed almost no power (by comparison to valve circuits), was not microphonic, and had an indefinite life. Added to this was the fact that it also occupied very little space (again compared to valves), and it used a single supply voltage (no heater) that was a nice safe low voltage.

Compare the angst of having to perform all those calculations (and we didn't even consider noise contour curves or anything even remotely esoteric), with the ease of an opamp. There is no comparison. During the reign of germanium transistors things were even worse, because the leakage of germanium devices is so high (compared to silicon). Although the distortion is at a tolerable level, intermodulation products are higher than is desirable - 1kHz + 9kHz at 200mV input voltage each generates

intermodulation at -53dB relative to the 1kHz signal level. This will almost certainly be audible.

There is no reason to use circuitry of this type for any reason any more, as there are no benefits whatsoever.

3.0 - Two Transistor Feedback Pair

The increasing demands of customers and the need for circuits whose characteristics were fixed by resistors rather than device characteristics led to the design of better circuits. Device dependency was an issue that plagued valve equipment designers - if you wanted a different gain you often had to use a different valve. The ability to break away from that dependency was almost a miracle, and the improved circuits gained rapid acceptance.

All of these amplifiers are pure Class-A (as is the single transistor version), but the addition of feedback allowed much higher performance than was ever achieved with *any* preceding (affordable) technology. It was these circuits that essentially brought hi-fi to the masses - people were no longer satisfied with the 'mantel radio' (so-called because it fit neatly on the mantelpiece over the fireplace), and wanted better reproduction for prices that were not out of reach.

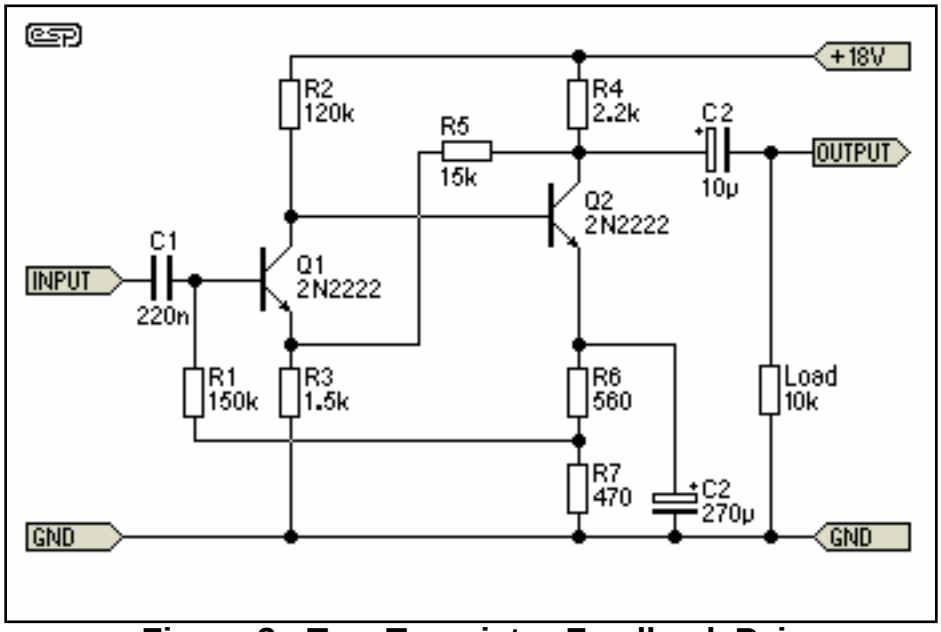


Figure 3 - Two Transistor Feedback Pair

Under exactly the same drive conditions, this circuit has a distortion of 0.065%. The spectrum is shown below, using the same axes as Fig 2. It is very apparent that it has

greatly reduced distortion. Although the harmonics do extend a little further than the first example, they are all well below the noise floor. This extension of the harmonics is obviously typical of feedback circuits, but when they are at such a low level it is of no consequence.

More interesting is intermodulation performance. Using the same criteria as in the first example, primary intermodulation products are at -63dB, 10dB lower than the simple amp above. Since intermod is by far more irritating and obtrusive than harmonic distortion, it is important to minimise this wherever possible.

It is worth pointing out at this stage that a simulated TL072 (not regarded as an audiophile opamp by any means) will give intermodulation figures under identical operating conditions that are **96dB** below the 1kHz or 9kHz levels. This is so much better than you will obtain with any of these circuits that it's just not worth considering them (IMHO). However, I have started this article, and will continue to the bitter end

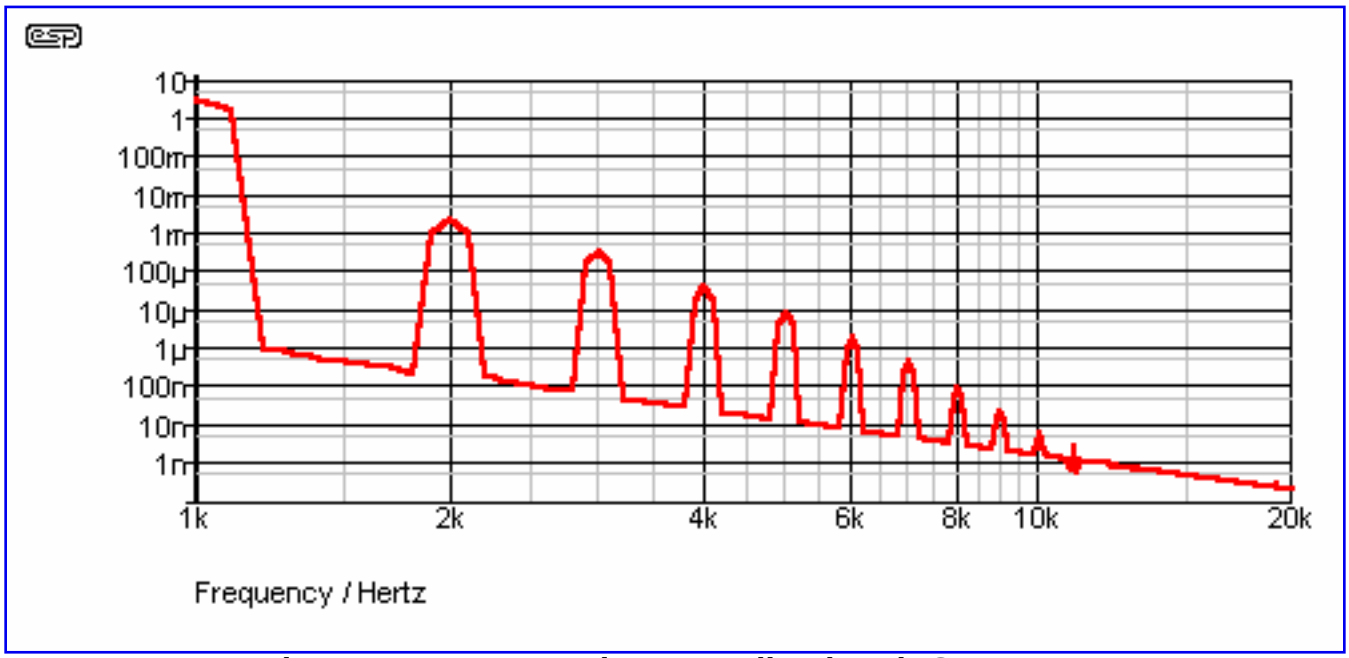


Figure 4 - Two Transistor Feedback Pair Spectrum

The distortion components are reduced, and to make comparison easier, click on the above image to open it in a new window. Now, you can go back to Figure 2 and compare the two. Not so readily apparent is the fact that input impedance is much higher and output impedance much lower than a simple amplifier stage.

These are important parameters for audio (and indeed for all amplification circuits). The input impedance of the Fig 3 amp is around 145k - not much less than the value of the bias resistor (R1). The input impedance of the base of Q1 is about 4.6Meg with the circuit shown, and this is almost entirely due to the feedback (applied via R5 to the

emitter of Q1). This feedback also defines the gain of the circuit, which is much closer to the design value (10 times, or 20dB). In theory, gain should be 11 - it can be worked out from ...

$$Av = (R5 + R3) / R3) = (15k + 1.5k) / 1.5k = 11$$

While it is easy to tweak the values to get the gain to exactly 10, in all cases in this examination we will be satisfied if it is within 3dB of the target. The circuit shown is a direct translation from the circuits presented in Reference 1 below.

Output impedance is harder to calculate than gain or input impedance, because you need to know the open loop gain of the amplifier. It was measured at about 82 ohms, a vast improvement, and this means that sensible load impedances will not change the gain of the amplifier. It is very important to understand that the output impedance of any amplifier does *not* mean that it can drive that impedance.

I measured the open loop gain of the simulated circuit at 1,220, so (in theory) the output impedance is equal to ...

$$Zout = R_{load} / Av_{open\ loop} = 1.8k / 2,220 = 81\ Ohms$$

Note that the effective load resistance is *always* equal to the collector resistance in parallel with the external load resistance. This applies for all transistor topologies, and is independent of feedback or anything else you may imagine.

Particularly with resistor loaded circuits, the peak current available is determined by the output resistor - for the Fig 3 circuit, this means that it cannot supply more than 4mA peak into any load, and it is important that this is not exceeded. In fact, the current demanded by the load (be it a following amplifier stage, a volume control or external equipment) should be no more than ½ the peak current available to maintain linearity. Ideally, it will always be much less than this. For the circuit shown, the minimum suggested load is around 10k, although it will tolerate less if the peak output voltage swing is restricted to no more than (say) ±3-4V.

It is impractical to try to describe the design process for this circuit - not because it is hard, but because it is just time-consuming. Anyone who is interested enough can use the principles used in 2.2 to analyse the circuit. Despite the additional complexity, the basic analysis methods are the same.

3.1 - Dual Transistor Feedback Pair Summary

This class of circuit pretty much sealed the fate of simple stages, and also entrenched the use of feedback as an indispensable tool to obtain improved performance with predictable results despite device parameter spread. Variations on these feedback amps were used for equalisers (especially RIAA phono and tape head preamps), microphone preamps, and a great many others.

Because of the tedious design process, Philips published a table of component values for different gains for the circuit of Fig. 3, although they probably could have saved themselves some trouble with a fairly simple design modification. By partially bypassing R3 (using a resistor and capacitor in series) the gain is easily modified, however, I have elected to reproduce the table below ...

Component	10dB Gain	20dB Gain	30dB Gain	40dB Gain
R1	150k	150k	150k	150k
R2	120k	120k	120k	120k
R3	4.7k	1.5k	1.5k	1k
R4	1.8k	2.2k	2.2k	2.2k
R5	12k	15k	56k	180k
R6	470R	560R	330R	680R
R7	1.2k	470R	270R	220R

Table 1 - Component Values for Different Gains (Figure 3)

It is recommended that a 10pF cap be wired in parallel with R5 for the 40dB gain version.

It is worth noting that there are two feedback loops in this design. One is DC only, and uses the voltage divider formed by R6 and R7. This biases the input transistor. The second feedback loop is via R5 and R3 - this is both AC and DC, so stabilises the DC operating conditions as well as the AC circuit gain.

All in all, this circuit was a breakthrough in the early days of transistor circuits, and I learned a great deal about the design process by analysing circuits such as this when I first became serious about electronics. That circuits such as this will never be the equal of one of today's opamps is readily apparent, but given the resurgence of people wanting to use discrete circuitry someone will hopefully find this information interesting. Even today (after so many years of electronics) it is still fascinating to see the ways that

were used to solve the problems that are inherent in all basic circuits.

4.0 - Buffers

Buffers were always a tricky issue before opamps, but the emitter follower is probably one of the best known single transistor circuit known. It is still used to this day in some cases, and is perfectly ok where extreme precision is not needed.

4.1 - Non-Inverting Buffers

The emitter follower is non-inverting, has very good linearity for a simple circuit, but does not have unity gain. Typically, the gain is quoted as being between about 0.90 to 0.99.

Figure 5 should be easily recognisable - this is a classic emitter follower. Gain is 0.98 and the circuit has very low distortion at 0.059%. The highest intermodulation product is at -65dB (using the same parameters as before). While not as good as an opamp, these figures are fairly respectable, and the circuit is very cheap and almost infinitely reliable. The distortion spectrum is not shown, but is almost identical to that in Figure 4.

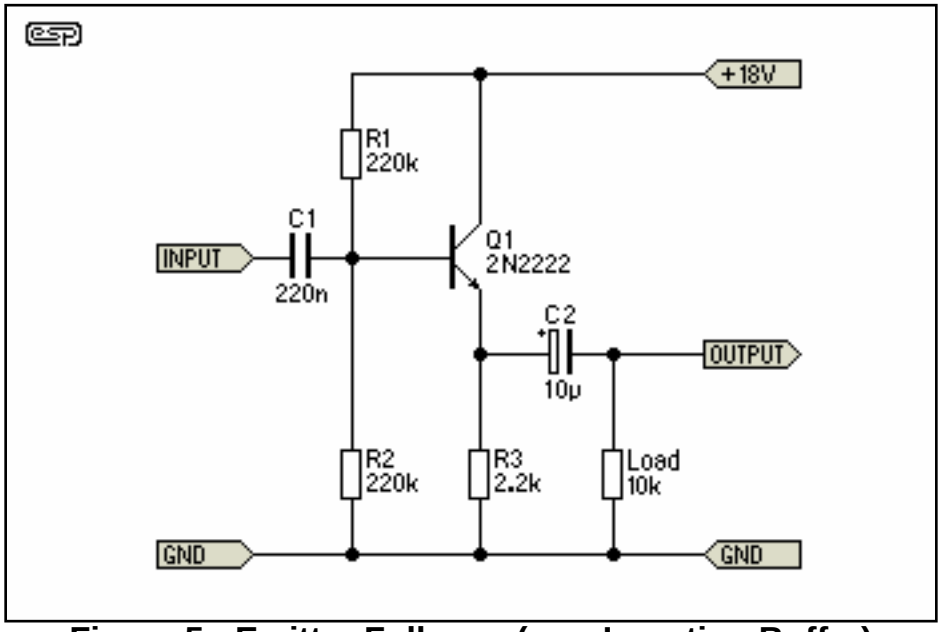


Figure 5 - Emitter Follower (non-Inverting Buffer)

The emitter follower is the easiest transistor circuit to design, requiring the absolute minimum of maths to get a working circuit. Indeed, in most cases you can simply grab a handful of resistors, a transistor and a couple of caps and be in business. It is important to make sure that the emitter resistor is capable of supplying the current needed by the

load, but this is usually as simple as making its value equal to around one fifth of the expected load impedance. For a load of 10k, that means a 2.2k resistor is fine.

Measurements of the Fig 5 circuit show that the input impedance at the base of Q1 is 390k. The calculation for this is simply (and roughly) ...

$$Zb = (Ze + re) * h_{FE} = 1.8k * 200 = 360k$$

Remember that re can be ignored if it is less than one tenth the value of the external load in parallel with the transistor's load resistance (now in the emitter circuit). Calculated and measured values are close enough. As before, total input impedance is

$$Zin = 1 / (1 / R1 + 1 / R2 + 1 / Zb) = 1 / (1 / 220k + 1 / 220k + 1 / 360k) = 84k$$

Output impedance is (approximately) equal to the total impedance seen by the base divided by the transistor's gain. This means that output impedance depends on the source impedance. This level of dependency affects all simple transistor circuits, where inputs are dependent on outputs or vice versa.

There is so little to the emitter follower that I have completely run out of things to say ... except that it must be remembered that an emitter follower operates with 100% local feedback, and the circuit will become unstable if a reactive load (such as a capacitor) is connected to the output. I always recommend that a series resistor be used at the output of any emitter follower that will drive a cable that is more than 100mm long.

Despite my claim that I've run out of things to say, there actually are many more things that could be said about emitter followers, but this article is intended to cover the basics - if every variation were to be discussed it would become a book.

4.0 - Inverting Buffers

Inverting buffers are harder with discrete transistors without excessive circuit complexity. This rather paradoxical state of affairs occurs because of the very nature of the transistor and how it works. While a simple amp is inverting, a simple buffer (emitter follower) is non-inverting. It is possible to create an inverting buffer using the two circuits in tandem, as shown in Figure 6.

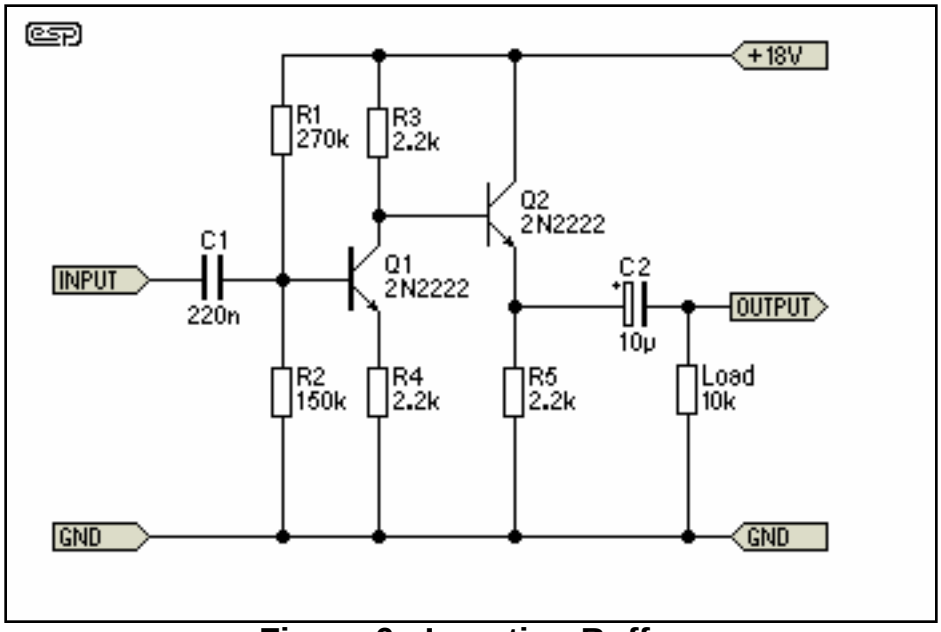


Figure 6 - Inverting Buffer

This circuit has many limitations, the main one being that voltage swing is rather limited. The absolute best that you can expect is about two thirds of the supply voltage (peak to peak), so for the circuit shown, about 12V P-P or about 4V RMS. Assume that you will actually get less than this - the circuit above will manage 3V RMS before clipping.

Because Q1 has the same resistor value for both collector and emitter, the voltages on each are essentially equal and opposite. The first stage therefore has unity gain, but inverted at the collector, while the emitter follower also has unity gain, but non-inverting.

4.3 - Buffers Summary

Simple transistor buffers are acceptable where one needs to provide a low output impedance with the minimum of fuss. The emitter follower is not good enough to use in accurate filters, as its gain of less than unity produces filters with a Q that is lower than expected, changing the filter characteristics.

The inverting buffer is a travesty - IMO it is next to useless. Given the limitations and complexity of the circuit, it is difficult to think of a single area where it might be used to advantage.

By comparison, even a cheap opamp will outperform either of these circuits in almost every respect, and with fewer components.

5.0 - Tone Control

The original Baxandall feedback tone control was based on a simple amplification stage. It was used with both valve and transistor circuits, and although it certainly works, it also has some bad habits when used this way. Again, the problem is mainly one of distortion - with only a limited amount of gain available, there is inevitably more distortion than is usually considered acceptable.

Figure 7 shows the basic arrangement, but it must be admitted that there were a great many alternatives that performed far better than the circuit shown here. Again, it is impossible to give all the variants (or even a subset) because there are so many. *All* of the discrete versions have worse distortion characteristics than a similar circuit made using an opamp of reasonable specification.

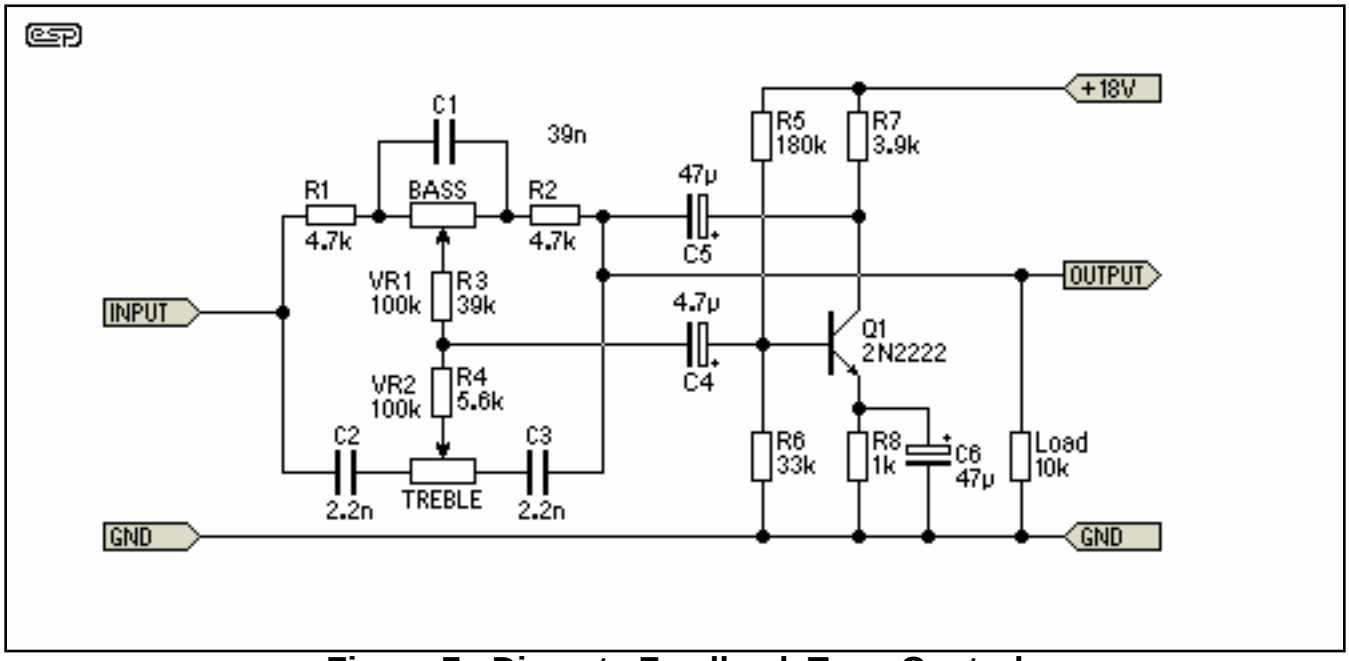


Figure 7 - Discrete Feedback Tone Control

Distortion is only around 0.03% with the controls centred, and it manages to stay reasonably constant at 1kHz regardless of setting. At either frequency extreme things change radically though. Intermodulation rises dramatically at maximum treble cut or boost, with the first intermodulation products at -32dB below the 9kHz tone at either extreme. It must be considered that if the system has to be used with maximum treble boost or cut, then intermodulation is probably the least of your problems.

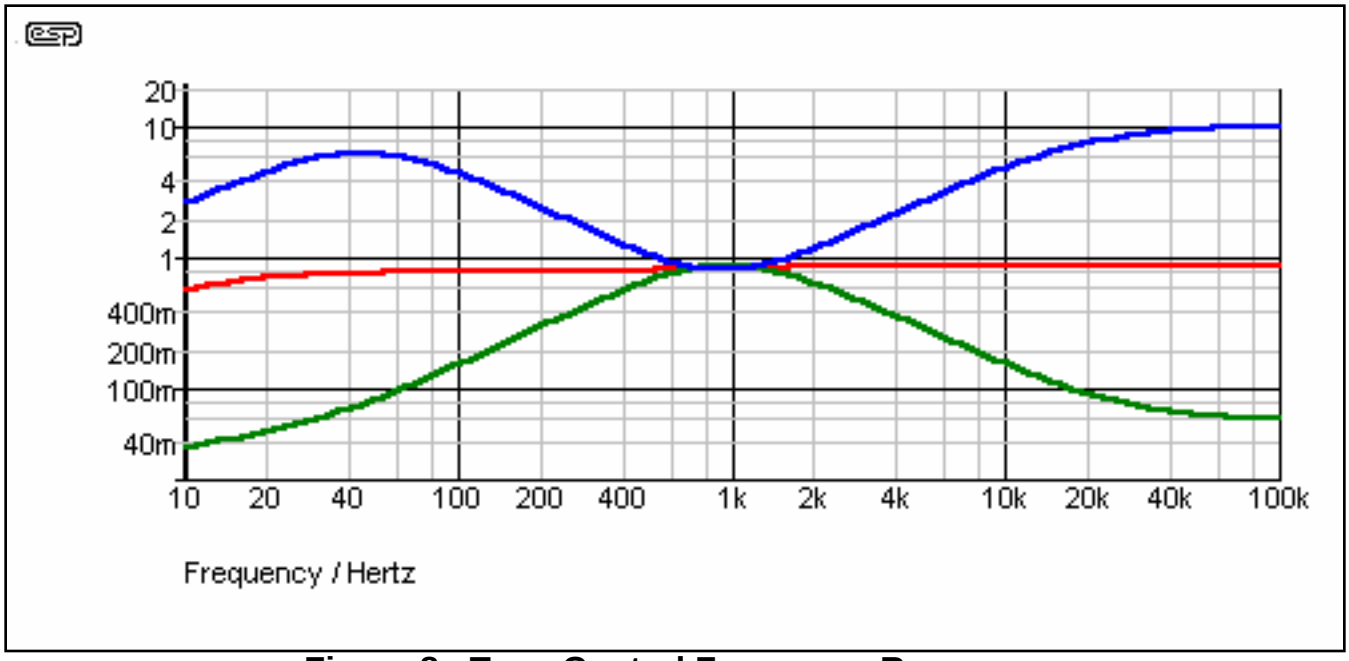


Figure 8 - Tone Control Frequency Response

The response is shown above. It is quite obvious that there is much more treble boost (especially) than you will ever need (or hear), with the boosted signal extending to 100kHz and well beyond. Bass is reasonably well behaved, but with the peak boost at 42Hz, it needs to be extended to work with a system with a subwoofer.

6.0 - Other Solutions

There is a lot to be said for discrete opamps - or at least circuits that have better performance than those looked at earlier in this article. With 0.024% THD and a very clean spectrum, the Project 37 amplifier is better than all of the alternatives shown previously. The main reason is the use of a current source (Q2) as the transistor load. Simply replacing Q2 with a 2.2k resistor increases THD to 0.25%.

To obtain the required gain (20dB), R5 had to be reduced to 1.5k - with 2.2k (theoretical gain of 11), the gain was well down from what it should have been. As we saw in previous examples, this is quite normal where the open loop gain is not high enough.

Although the circuit looks much more complex, it is only marginally worse in this respect than the circuit of Fig 3. The simple fact of the matter is that this level of performance is just not available from less complex circuits. As Einstein said, "All things should be as simple as possible, but no simpler". This is very true when it comes to circuits of this type, and in the author's opinion, far too many 'hi-fi' amplifiers are much simpler than they should be.

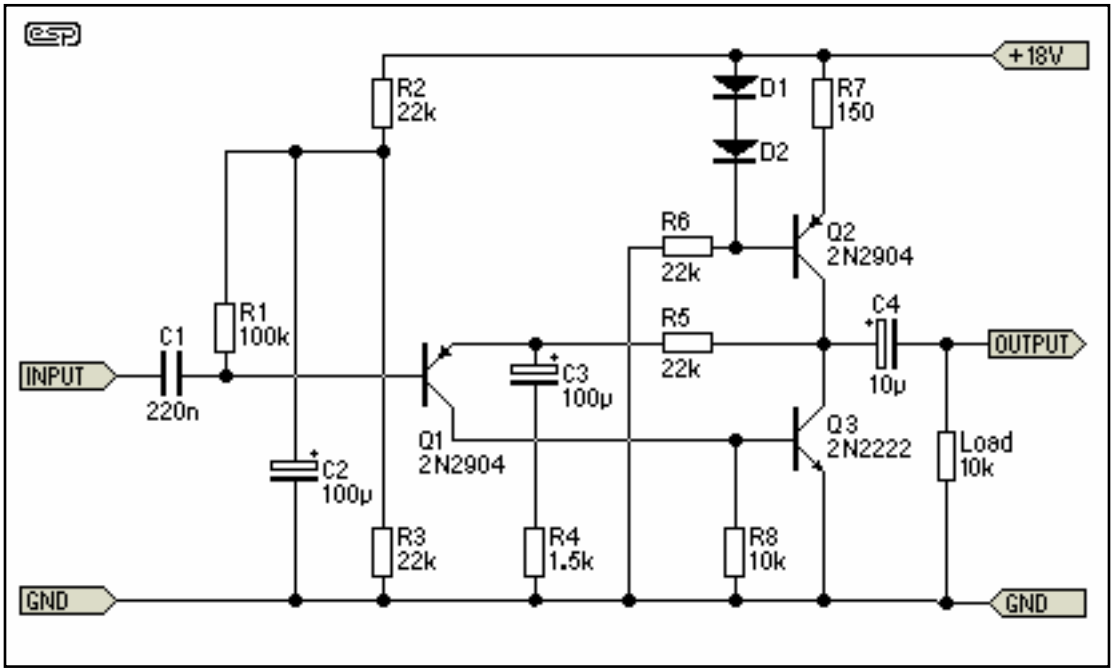


Figure 9 - Discrete Amp (as used in Project 37)

The spectrum is shown below. This shows that the harmonic structure is almost entirely second and third harmonic, and even the third is at -98dB. This is negligible, so it is not unreasonable to claim that distortion is almost entirely second harmonic. The intermodulation results are even better - the first intermodulation product is at -72dB, almost 10dB better than the Fig 3 circuit, and nearly 20dB better than the simple amp in Fig 1.

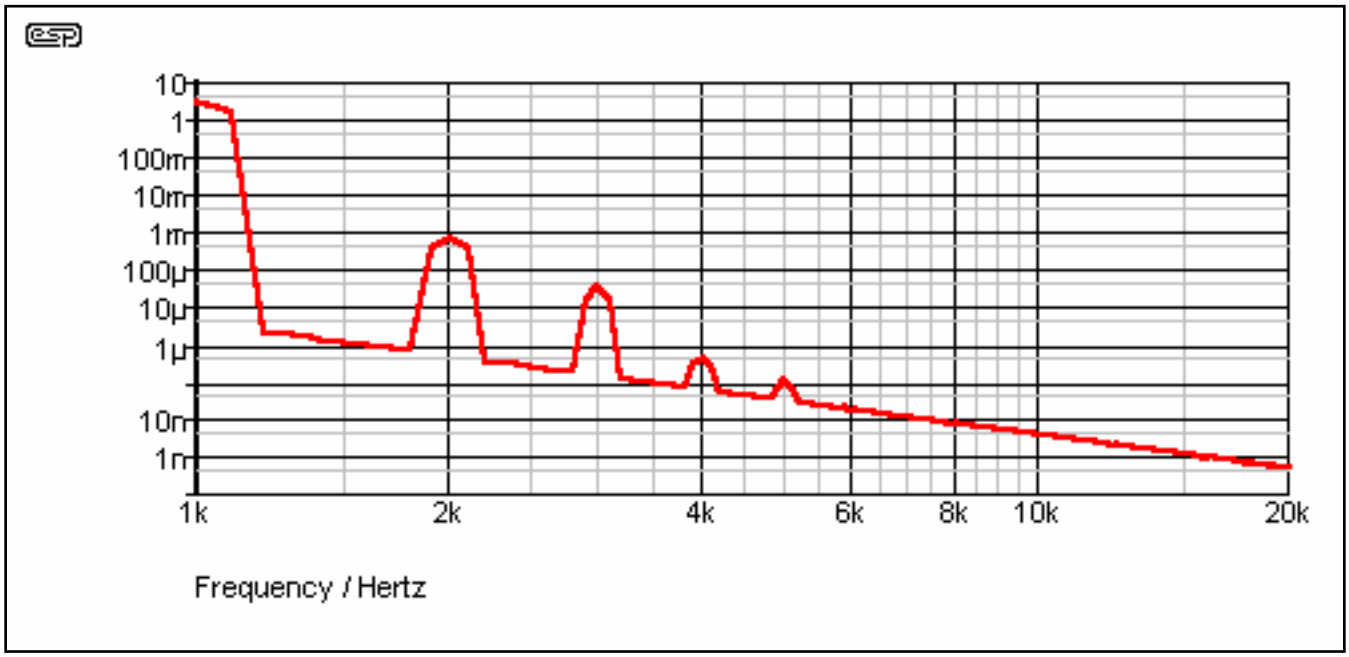


Figure 10 - Distortion Spectrum

The input impedance is about 96k - note the biasing arrangement, which is designed to minimise noise and maximise input impedance. C2 decouples any supply noise, and

bias current is applied via R1. Q1's base input impedance is greater than 2.8 Meg, so R1 could be increased if desired.

It is fairly obvious that the small increase in complexity has yielded a large benefit - much lower distortion being the major improvement. This is particularly important for intermodulation distortion, and this is markedly improved from any of the alternatives.

7.0 - Conclusion

From the details here, it is easy to see that it it not easy to obtain good performance from simple discrete circuits. While improved versions (such as the P37 preamp we looked at in section 6) do perform very well, in general the 'old' circuits do not.

They were certainly an enormous improvement in their day, but that day has passed. Modern opamps are so superior that there is no reason to look elsewhere, unless you wish to experiment and learn more about the design processes involved. While there is certainly some historical value in the circuits shown, they are no longer to the standard expected of true high fidelity.

Where the simulator shows that the first intermodulation product using a TL072 is 109dB below the 9kHz tone, none of the circuits shown come even close to this. The P37 amp is certainly better than any of the others, but it is still significantly worse then the opamp. It is probable that the intermod from P37 is inaudible (after all, many people have commented on just how good it sounds), and it is equally probable that the simulator over-estimated the TL072's quality - it *is* a simulator, after all. I ran a new intermodulation test using an LM324 opamp (hardly a reknowned device for audio), and SIMetrix said that intermod was at -68dB. From that, we can deduce that the models are reasonable so the simulated results will not be too far from reality.

There is no doubt that devices such as the OPA2134 or the venerable NE5534 (or the dual version, NE5532) will be much better than anything shown here, but will they *sound* as good, better or worse? This is something that can only be determined with a proper blind AB test. Logically, it probably should not be possible to hear the difference between either opamp, or the P37 circuit. As for the others ... I don't know, but if you ever find out, be sure to let me know.

References

1. SIMetrix SIMetrix circuit simulator, Catena Software (UK)
2. Audio Amplifier Systems Philips Application Book , MD Hull, Third Edition, June 1972
3. Project 37 Elliott Sound Products , Rod Elliott, November 1999

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 10 Apr 2005

Elliott Sound Products — Linear Power Supply Design

Rod Elliott - Copyright (c) 2001
Page Last Updated 03 Mar 2001

Main Index

Articles Index

Contents

- Introduction
- Power Amplifier Definitions
- Power Supply Requirements
- Transformers
  - Transformer Series Resistance
- VA Versus Watts
- Further Analysis
  - Voltages and Currents
  - Increasing Capacitance / Transformer Size
- Rectifier Types
  - Full Wave
- Temperature
- Capacitor Value
- Rectifier Diodes
- DC Sound
- Summary
- Fusing and Protection
- Disclaimer
- References

Introduction

Having searched the Web for reference material (and found very little!), this would appear to be the definitive article on the design of a 'simple' linear power supply for a power amplifier. Power supplies are needed for every type of amplifier we will ever use. I do not intend to deal with "esoteric" designs with interesting names, but the simple, linear power supply that is still the mainstay of audio.

These supplies should not create any problems for anyone, because they are so simple, right? *Wrong!* They appear simple, but there are many inter-related factors that should be considered before just embarking on your next masterpiece. The purpose of this article is to explain the terminology used, traps and pitfalls, and give some insight by way of a few practical examples.

Power Amplifier Definitions

Power supplies themselves require several definitions (these are discussed later in this article), but the requirements of the amplifier that is to be connected need to be understood before we start. This makes a very big difference to the way the supply performs.

Poor earthing practices, such as connecting components to the nearest available ground reference can (and do) also create problems, and these can introduce hum, or more usually a "buzz" into the signal circuits. This applies equally to Class-AB and Class-A amplifiers, but is usually more apparent with Class-A since the maximum current is drawn on a continuous basis.

Class-AB Amplifiers

I shall refer to the standard power amplifier as Class-AB - of all the amplifier types, these are the most common. Any amp that draws a quiescent current through the output devices is by definition, Class-AB. For true Class-B, there is no quiescent at all, and the output devices will conduct for exactly 180 degrees - this is rare.

Class-AB amps have a very widely varying current drain, which may be only 100mA or less with no signal, but rising to many amps when driven. The main problem is the revolting waveshape of the current on each supply lead, typically half-wave pulses, in sympathy with the program content.

These waveforms - this is current, not voltage - have sharply defined transitions, and as such will generate a magnetic field which varies with the current. Since a sharp transition equates to high order harmonics, care must be taken to ensure that voltages are not induced into the input stages of the amp from the supply lines. Because of the low inductance of the wiring of an amp, these problems are going to create distortion components which will tend to be worse at higher frequencies.

It is not only the amplifier which creates current pulses, but the rectifier/ filter capacitor combination as well. The power supply rectifier diodes usually conduct for only a short time during each AC cycle - this may be as little as 3 or 4 degrees at idle, but both the angle of conduction and the amplitude of the current pulse will increase as more power is drawn from the supply.

Class-A Amplifiers

The other common amplifier type is Class-A. These amps draw a large current on a continuous basis, and place a completely different loading on the supply. The current pulses are gone from the supply leads, but the rectifier and filter now must handle the maximum current on a continuous basis.

The continuous load creates a new set of constraints on the design of a power supply, and the use of a Class-A amp implies that the builder already wants the very lowest noise. Although the noise of the power supply DC output (hum/ ripple) will normally be low because of extensive filtering, regulation or a capacitance multiplier, the switching noise of the diodes in the rectifier can become more than a nuisance if proper care is not taken.

Power Supply Requirements

Somewhat surprisingly perhaps, the fundamental requirements of the final design are not greatly influenced by the different loading presented by the two amp types described above. The continuous rating of a Class-A amp means that you must design the supply for a continuous (rather than transient) current, but since we are discussing properly designed, quality power supplies, the final result will often be quite similar.

When a power supply is used with an amplifier, the basic things we need to know before starting are as follows

- Power output and minimum impedance, or ...
- Peak / average current

Acceptable power supply ripple voltage

With only these three criteria, it is possible to design a suitable supply for almost any amplifier. I shall not be describing high current regulators or capacitance multipliers in this article - only the basic elements of the supply itself. These other devices are complete designs in themselves, and rely on the rectifier/ filter combination to provide them with DC of suitable voltage and current.

Transformers

The first component of the power supply is the transformer. Using magnetic coupling between windings, the transformer is used to isolate the amplifier (and the users) from the mains voltage, and to reduce (for solid state equipment at least) the voltage to something the amplifier can tolerate. The primary winding will be rated at 240, 220 or 120V AC depending on where you live, and the secondary will be a more user friendly (or less user hostile) voltage to suit the amplifier.

DC Output: The DC output is approximately equal to the secondary voltage multiplied by 1.414, but as we shall see, this is a rather simplistic calculation, and does not take the many variables into consideration. At light loading, this rule can be applied without fear, and it will be accurate enough for most applications. When an appreciable current is drawn, this simple approach falls flat on its face.

Mains variations: These occur in all situations, and the mains voltage at any point in time will usually be somewhat different from the nominal voltage quoted by the supplier. Any variation of 10% or less can be considered "normal", and greater variations are not at all uncommon. In nearly all cases, an amplifier is rated at a certain power output into a specified load impedance, and at the nominal mains voltage. For those who live close to a sub-station or pole transformer, expect the voltage (and power output) to be higher than quoted - the rest of us can expect a lower mains voltage and less power, especially during peak electricity usage times.

Losses: Since all transformers have losses, these cannot be ignored in the design phase. Magnetising loss (AKA iron loss) is the current that is required to maintain the design value of magnetic flux in the transformer core. There is nothing you can do to affect this loss, as it is dependent on the size of the core and the design criteria of the manufacturer. Large transformers will have a larger magnetising loss than small ones, but will be less affected by it due to the larger surface area which allows the transformer to remain cool at no load. Small transformers have a greater loss per VA than bigger

ones, and this is one of the reasons that small transformers run quite warm even when unloaded.

The iron losses are greatest at no-load and fall as more current is drawn from the transformer. Copper losses are caused by the resistance of the winding, and are negligible at no load, and rise with increasing output current. There is a fine balance between iron and copper losses during transformer design. A relatively high iron loss means that copper losses will be reduced (thus improving regulation), but if too high, the transformer will overheat with no load. A full description of the magnetising current and its effect on regulation is outside the scope of this article, and since there is little you can do about it, it shall be discussed no further.

Mains noise: Noise can easily get through a transformer, both in transverse and common modes. Transverse noise is any noise or waveform distortion that is effectively superimposed on the incoming AC waveform, and this is coupled through the transformer along with the wanted signal - the mains.

Common mode noise is any noise signal that is common to both the active (hot) and neutral mains leads. This is not coupled through the transformer magnetically, but capacitively. The higher the capacitance between primary and secondary windings, the more common mode noise will get through to the amplifier. The much loved toroidal transformer is much worse than conventional "EI" (Ee-Eye) lamination transformers in this respect because of the large inter-winding capacitance. An electrostatic shield will help, but these are uncommon in mass produced toroidal transformers. The conventional transformer is usually better, and by using side-by-side windings instead of the more common (and cheaper) concentric windings, common mode noise can be reduced by an order of magnitude.

Input mains filters can remove either form of high frequency noise component to some degree, and large spikes can be removed using Metal Oxide Varistors (MOVs) that effectively short circuit the noise pulse, reducing it to a level that is (hopefully) inaudible. Contrary to the beliefs of some, there is no panacea for noise, and it is best attacked in the equipment, rather than the now popular (but mainly misconceived) notion that an expensive mains lead will cure all.

Regulation: The "regulation" of typical power transformers seems to be rather poor when some measurements are taken on a complete power supply. When specified, regulation is based upon a resistive load over the full cycle, but when used in a capacitor input filter (99.9% of all amplifier power supplies), the quoted and measured figures will never match.

Since the applied AC spends so much of its time at a voltage lower than that of the capacitor, there is no diode conduction. During the brief periods when the diode conducts, the transformer has to replace all energy drained from the capacitor in the intervening 7 to 8ms or so (assuming 50Hz and full wave rectification).

Consider a power supply as shown in Figure 1. This is a completely conventional full-wave capacitor input filter (it is shown as single polarity for convenience). The circuit is assumed to have a total effective series resistance of 1 Ohm - this includes transformer winding resistances (primary and secondary) and diode losses. The capacitor C1 has an initial value of 4,700uF.

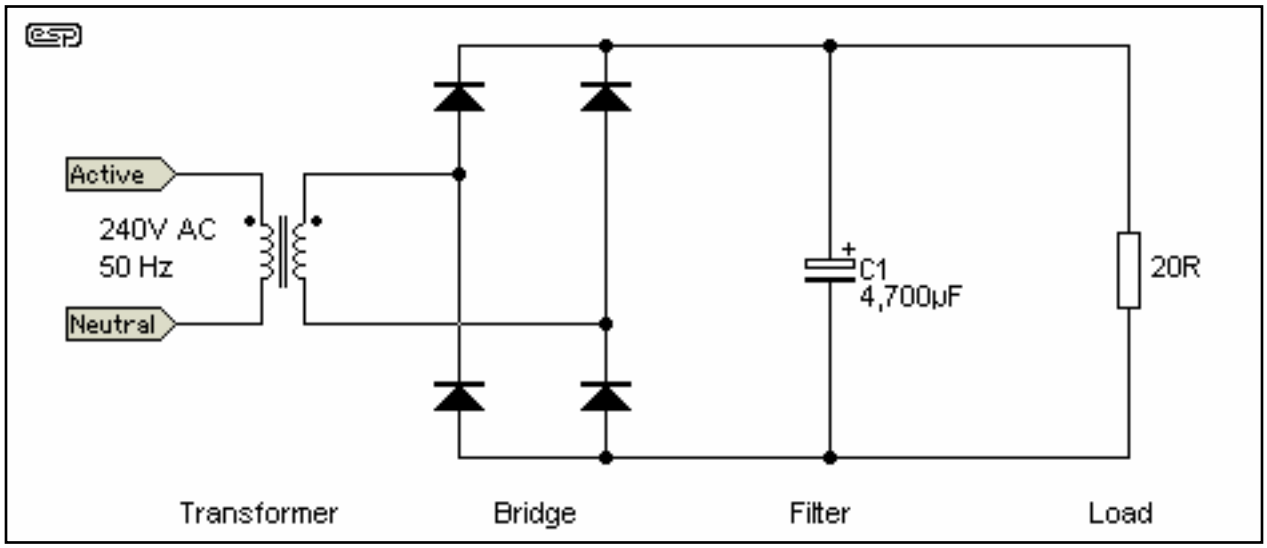


Figure 1 - Full Wave, Capacitor Input Filter Rectifier

The transformer is rated at 60VA and has a primary resistance of 15 Ohms, and a secondary resistance of 0.5 Ohms. This calculates to an internal "copper loss" of 0.75 Ohm.

With a 20 Ohm load as shown, and at an output current of about 1.7 amps (1.69A actually), diode conduction is about 2.8ms, and the peak value of the current flowing into the capacitor is 6A - 100 times per second (10ms interval). Diode conduction is therefore 28% of the cycle. (To check that these figures are valid, we can take the peak capacitor current and the duty cycle, and calculate the average ...

$I_a = (I_p / 100) * \text{duty cycle}$ so, $I_a = (6 / 100) * 28 = 1.68$

(close enough for our purposes)

Ripple across the load is 2.4V peak-peak (or about 740mV RMS), and is the traditional

sawtooth waveform. Average DC load voltage is 33.8V. The no-load voltage of this supply is 41V, so at a mere 1.7A load, the regulation is ...

$$\text{Reg (\%)} = ((V_n - V_l) / Vn) * 100 = 17.5\%$$

Where V_n is the no-load voltage, and V_l is the loaded voltage.

For this example, this works out to 17.5% which is terrible. By comparison, the actual transformer regulation would be in the order of 5% for a load current of 2A at 30V. Note that the RMS current in the secondary of the transformer will be just over 3 amps AC for a load current of 1.7A DC - this must be so, since otherwise we would be getting something for nothing - a practice frowned upon by physics and the taxman.

Output power is 33.8V * 1.7A = 57.46W, and input power is 240V * 0.25A = 60VA. The losses are caused by the diode voltage drop and winding resistance of the transformer.

Transformer Series Resistance

As described above, I assumed a total equivalent series resistance for the transformer of 0.75 Ohm, which is about typical for a 60VA transformer as used here. Larger transformers will have lower series resistance (and vice versa), and the equivalent may be calculated - this is easier than actually measuring it under load.

If the secondary resistance is (say) 0.5 Ohm for a 240V to 30V transformer, it will be found that the primary resistance is (or should be) in the order of 15 Ohms. The actual figure will vary from one transformer type to another (e.g. "conventional" EI (ee-eye) laminations versus toroidal).

The effective primary series resistance is calculated (approximately) by ...

$$R_e = R_p / (T_r)^2$$

Where R_e is equivalent primary resistance, R_p is measured primary resistance, and T_r is the turns ratio (in this case, 240 / 30 = 8

Therefore ...

$$R_e = 15 / 64 = .234 \text{ Ohm}$$

This value is now added to the secondary resistance to calculate total series resistance. Please don't bother to e-mail to tell me that these figures are not correct - this is intended as a rough approximation - calculating actual values for transformers is worthy of an article in itself (which I am not about to write! :-)

The sad truth is that accuracy here is completely unimportant, as there is also series resistance in the mains power wiring from the power generation plant all the way to the power transformer primary winding. This is going to vary from one outlet to another and from one house to the next. Although it can be measured, this is a completely pointless exercise since it will only be relevant for one household. Other factors are the actual supply voltage (nominal 115V, 220V, 240V, etc.) which varies widely from day to day and hour to hour. (For what its worth, the actual supply voltage was 233V when I measured the mains impedance at approximately 0.8 Ohms at my workbench - does this help?). We shall now do what everyone else does, and ignore it completely.

However - again for what its worth - your 100W / 8 Ohm amplifier will be reduced to just over 90W, simply by connecting a 2400W heater to an adjacent power outlet, based on 0.8 Ohms mains wiring impedance and a genuine 240V supply voltage (before connection of the heater). The situation is likely to be slightly worse in the US, because the much lower AC supply voltage means that all currents are doubled for the same power.

VA Versus Watts

An important distinction must be made between power (Watts) and VA. Power is a measure of work, and it is quite possible (common, actually) to have a situation where there is voltage and current, but little or no work. The product of voltage and current is Volts * Amps, or VA, and there is commonly a wide variance between VA and Watts.

Various loads (capacitive or inductive) will draw current from the output of a transformer, amplifier or the mains supply. If the load is purely inductive or capacitive, there is no power (work) at all, even though the current may be quite high. Fluorescent lighting fixtures are renowned for this, where the current can be several times what was expected based on the power rating of the tubes.

This phenomenon is called "power factor", and a power factor of 1 means that there are no power losses due to inductance or capacitance. Likewise, a power factor (PF) of 0 means that there is lots of voltage and current, but no power. In the case of fluorescent

lighting, power factor correction capacitors are used to try to maintain the PF at as close to unity as possible. If this were not done, the wiring to the fittings (especially in large commercial buildings) will overheat, and a much greater load than necessary is placed on the local power sub-station, and indeed on the entire power grid. Electricity supply companies worldwide have the same problems, and in most countries, there is legislation that determines the minimum acceptable power factor for any installation.

The switch mode power supplies used in computers have a very poor PF, and there are many new designs that improve this. These can be expected to become mandatory in the not too distant future, as a poor power factor makes electricity more expensive to supply, and therefore more expensive for the consumer.

Note that I do not propose to cover the topic of power factor in depth (in fact that was it!), but a basic understanding is useful, and will make some of the following information more sensible.

A power transformer does not care if work is being done at the output or not. It has internal resistance and inductive losses, and cares only about the output voltage and current. A power transformer can be overloaded and destroyed by a large capacitance directly across the output terminals. The capacitor does not even get warm, since it dissipates no power and does no work. The transformer "sees" only the load current, and heats up proportionally - if the VA rating is exceeded consistently, the transformer will eventually overheat and die. Equally, a transformer may be operated at 500% of its ratings for a short period, and as long as it has enough time to cool down between overloads, will be unaffected by the ordeal. Unfortunately, this otherwise useful characteristic is pointless in audio, since the voltage will fall too far with the load, and amplifier power output suffers badly. Having said this, most "mainstream" power amps will economise on the transformer, and rely on the duty cycle of typical programme material to provide an adequate supply voltage for normal music signals. The continuous (erroneously called "RMS") power will be lower, sometimes significantly.

The term "dynamic headroom" used to be used to describe the difference between continuous and peak output power. A large figure (2dB or more) indicates that the transformer is too small for the job, since the supply voltage collapses under a sustained load.

Because we are going to use the transformer in an unfriendly manner, with a rectifier and large capacitance as the load, the VA rating is much higher than the power rating of the amplifier may indicate. There are some basic rules of thumb for the most common rectifier types, and these are shown below.

Further Analysis

To properly see the effects of the losses and currents involved, a simpler circuit will be used from this point. This consists of a 25V RMS "ideal" generator, and the copper losses are simulated by a resistance. Since the full-wave bridge rectifier is a very common configuration, this is what shall be used for the detailed analyses that follow. There are variations and exceptions to everything, but simulations and real-life testing on these simple circuits are very close, so this is what shall be used.

A simple resistance is the load, and we shall see the vast differences in peak AC current, capacitor ripple current and output voltage as the various parameters are changed. A solid understanding of the behaviour of the transformer, rectifier and filter capacitor is essential if worthwhile power supplies are to be designed.

Voltages and Currents

Figure 2 shows the voltages and currents present in a typical supply. The waveforms will be examined shortly - for now we are interested in the average current and voltage in each section of the supply. The generator voltage is 25V RMS, and for this supply I have used a winding resistance of 0.75 Ohm - roughly equivalent to a 120VA transformer. The voltages and currents are all RMS - although in practice very few RMS meters can give an accurate reading of the spiky current waveform.

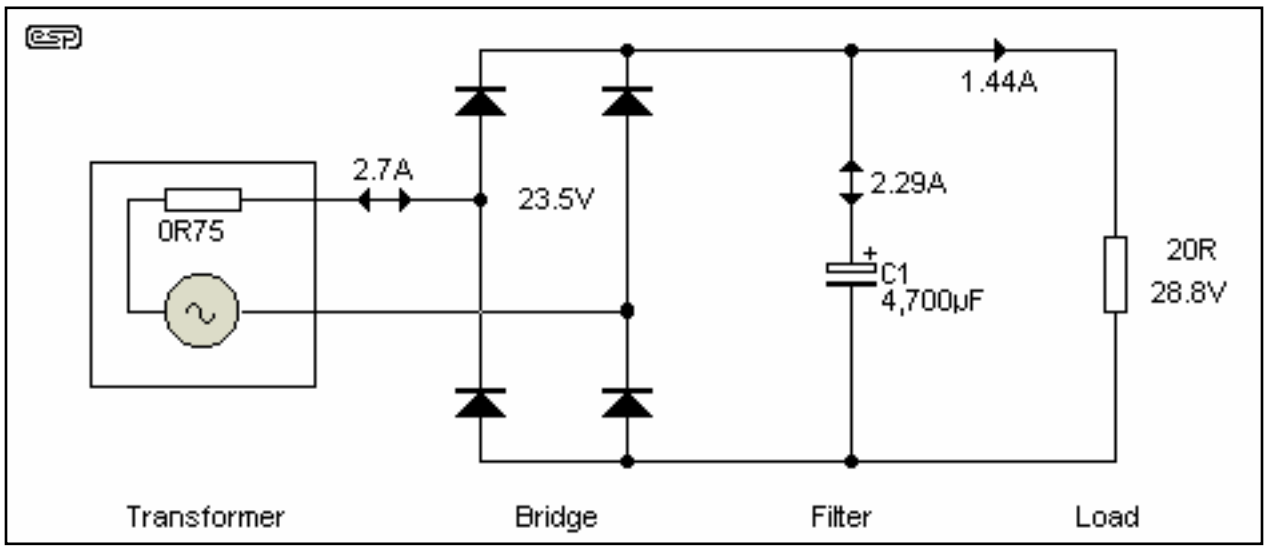


Figure 2 - Basic Bridge Rectifier - Voltages and Currents

Note the big difference between DC output current, capacitor ripple current and AC input

current. The important parameters are listed in the table below ...

Parameter	RMS	Peak
AC Voltage	23.6 V	31 V
AC Current	2.70 A	6.25 A
DC Voltage	28.8 V	-
DC Current	1.44 A	-
Ripple Current	2.29 A	4.8 A **
Ripple Voltage	650 mV	2 V (P-P)

Table 1 - Supply Voltages and Currents

** This figure is somewhat misleading, since there is both a charge and discharge cycle. During the discharge, there is a relatively constant current of -1.44A (the negative means the current is flowing out of the capacitor). During the charge period, the rectifier takes over the supply of current to the load, then re-charges the capacitor at the peak current shown.

Input power is 25V * 2.7A, or 67 VA, and output power is 41W, so in all, over 26W has been lost in the rectification and filtering process. About 2.6W is lost as copper losses in the transformer (dissipated in the 0.75 ohm resistor that simulates the winding resistances). Attempting to quantify each individual loss is a relatively pointless exercise, since the end result is to make a power supply that works - we can do nothing about the losses. In reality, many of the losses are very different from those calculated, since the RMS values are somewhat dubious (although quite OK for the purpose of this article).

About 0.9V is lost across each diode during conduction, but this will vary in practice, based on the current capacity of the rectifier diodes. Since this is a bridge rectifier, there are two diodes conducting at the +ve and -ve peaks of the waveform, so the total voltage loss is 1.8V - so the output DC should be around 32V. The measured values of 23.6V AC and 28.8V DC is a direct result of the measurement inaccuracy! Because current is drawn only at the peak of the AC waveform, the input to the rectifiers is not a sinewave. The voltage and current waveforms are shown below, and it can be seen that the voltage waveform has been "flattened" at the peaks. This is due to the high peak current drawn during this time, and no voltmeter will give the correct value - you must use an oscilloscope to be able to measure the peak-to-peak value of the waveform.

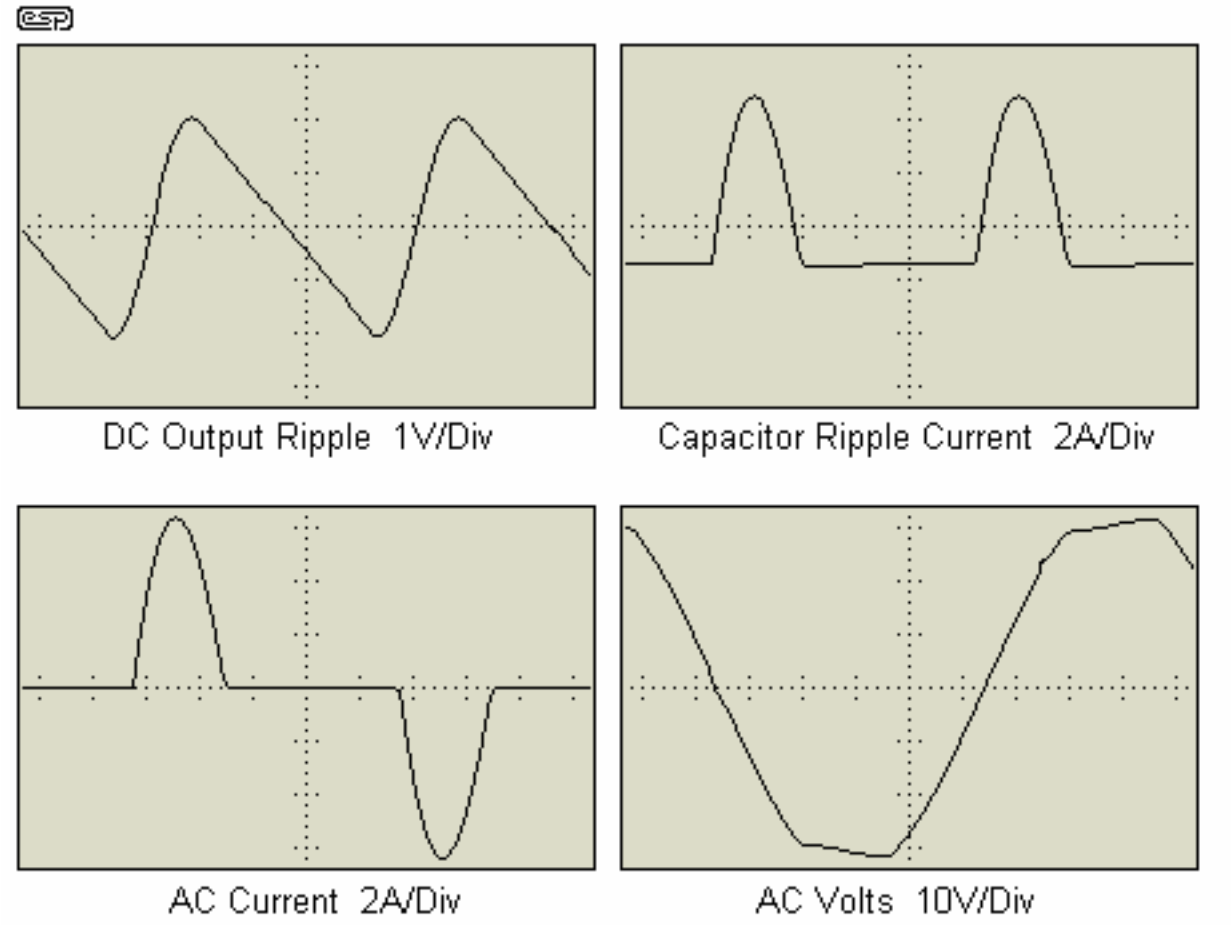


Figure 3 - Voltage and Current Waveforms

This reveals very different information from the voltage and current measurements taken before. Both are essential in understanding the rectification process. The peak AC input is only 31V, where we would normally expect 25 * 1.414 = 35V or (allowing for the voltage drop within the transformer) 33.4V. We have some missing voltage (31 - 1.8 is 29.2, not the 28.8V measured) if we simply use normal multimeters. Examination with the oscilloscope shows that the peak DC voltage is higher than the average value shown by the meter, and is 29.2V, so everything really does fall into place.

It must be pointed out that these waveforms were taken at different times, and are not exactly in phase. The positive going part of the output ripple voltage, the peaks of the AC current, the positive peaks in capacitor ripple current and the flattening of the AC input voltage all occur at exactly the same time.

Increasing Capacitance / Transformer Size

It is well known that bigger transformers have better efficiency than small ones, so it is a common practice to use a transformer that is over-rated for the application. This can

improve the effective regulation considerably, but also places greater stresses on the filter capacitor due to higher ripple current. This is quoted in manufacturer data for capacitors intended for use in power supplies, and must not be exceeded. Excessive ripple current will cause overheating and eventual failure of the capacitor.

Capacitor ripple current ratings can be ignored at your peril.

Large capacitors usually have a higher ripple current rating than small ones (both physical size and capacitance). It is useful to know that two 4,700uF caps will usually have a higher combined ripple current than a single 10,000uF cap, and will also show a lower ESR (equivalent series resistance). The combination will generally be cheaper as well - one of the very few instances where you really can get something for nothing. Using ten 1,000uF caps will generally give even better overall figures again, but the cost (in time and effort) of assembling them into a proper filter bank may not be felt worthwhile.

Above a specific value, as the capacitance is increased, the peak charging current will remain much the same for the same sized transformer, but the capacitor retains more of its charge between cycles. The switch-on current will be very much higher, and the surge will last longer as the capacitor charges. At capacitor values below optimum, the peak charge current will decrease somewhat, but there will be far greater output ripple.

There is no hard and fast rule for determining the optimum value for the filter cap, but in general I would suggest that the value should be at least that required to give a full load ripple voltage of less than 5V peak to peak. Based on this, my recommendation is that the absolute minimum value is 2,000uF per amp DC, so a 5A (continuous) power supply will have an absolute minimum of 10,000uF capacitance.

What is achieved by increasing the capacitance, is the ability of the capacitors to retain more of their charge between AC cycles. Since the current demands of a Class-AB amplifier vary so widely - with the majority of the time at very low average currents - the actual operating voltage will be closer to the no-load voltage.

With large capacitors, the momentary current peaks created by the program material will not be of sufficient duration to discharge the caps to the full load voltage levels, so there is more voltage available on a more or less consistent basis. This equates to more power for transient signals, and lower ripple voltages the rest of the time.

With a 4,700uF capacitor and a peak current of 5A (equivalent to the peak current of a 100W amp into 8 Ohms), the capacitor will lose voltage at the rate of 1V / ms between

"charges". As the capacitance is increased, this discharge rate naturally falls proportional to the capacitance. Doubling the capacitance halves the discharge rate and the ripple voltage for a given current, but increases the capacitor ripple current and the peak AC current - although the average value remains much the same. There are some small variations, but these are eventually accounted for if we analyse the waveforms critically - again, this is a relatively pointless exercise, and will not be undertaken.

What we will do, is see what happens in each individual case when ...

- The capacitance is increased
- The transformer is made bigger

Table 2 shows the currents and voltages with the same transformer used in Figure 2, but with a 10,000uF filter cap. There are insignificant increases in currents, and no worthwhile increase in the average DC output voltage. Output ripple is half that of the previous example. As can be seen, more capacitance will affect the DC ripple voltage but little else, and one may wonder if it is worth the effort (the answer is generally "yes", but it depends on the application).

Parameter	RMS	Peak
AC Voltage	23.6 V	31 V
AC Current	2.71 A	6.27 A
DC Voltage	28.8 V	-
DC Current	1.44 A	-
Ripple Current	2.29 A	4.8 A
Ripple Voltage	306 mV	1 V (P-P)

Table 2 - 120 VA Transformer / 10,000uF

Table 3 is the same data, with the original 4,700uF capacitor, but now with a transformer having 0.5 of the original total winding resistance (0.375 ohms) - this is equivalent to a transformer of about 4 times the 120VA rating used before (or about 500 VA).

Parameter	RMS	Peak
AC Voltage	24.2 V	32 V
AC Current	3.15 A	8.32 A
DC Voltage	30.3 V	-
DC Current	1.52 A	-

Ripple Current	2.76 A	6.8 A **
Ripple Voltage	731 mV	2.2 V (P-P)

Table 3 - 500 VA Transformer / 4,700uF

The increases in both average and peak currents are quite substantial, and the output voltage is higher by a small (but not especially useful) amount. The DC current is higher, only because the voltage is greater, and this is the sole reason for the increase in ripple voltage. The big test is to use the 10,000uF filter cap with this very much larger transformer, and see what increases occur.

Parameter	**RMS**	**Peak**
AC Voltage	24.2 V	32 V
AC Current	3.15 A	8.52 A
DC Voltage	30.4 V	-
DC Current	1.52 A	-
Ripple Current	2.79 A	7.0 A **
Ripple Voltage	345 mV	1.16 V (P-P)

Table 4 - 500VA Transformer / 10,000uF

In case anyone is wondering why I used a load resistance of 20 Ohms, this was to simulate one half of a 55W Class-AB amplifier operating at the onset of clipping into an 8 Ohm load, with a steady sinewave input. Any dynamic analysis is very difficult, and the results are not particularly meaningful unless the exact signal source is known, along with the specifics of the power amplifier that is connected to the supply.

> Actually, the current of 1.44A is slightly on the high side, since the RMS current will only be about 1.375A, which is 1/2 the speaker current (the other half of the speaker current is provided by the negative power supply).

In these calculations, I also made no allowance for the fact that nearly all transformers are rated for an output voltage at full current - this is invariably the voltage into a resistive load, and not a rectifier / filter combination. This means that the voltage will always be a little higher than specified at no load, and now you know why the DC is less than expected at full load.

Rectifier Types

So far we have looked at the full wave bridge rectifier, and this is but one of several different configurations. The most common rectifiers are ...

- Full Wave Bridge
- Full Wave
- Dual Full Wave (Full Wave Centre Tap)

There are others, but they are not commonly used in low voltage amplifiers. The dual full wave (using a bridge rectifier and a centre tapped transformer) is probably the most common of all, and some further analysis of this rectifier shall be covered next. I have chosen to ignore half wave rectifiers (in all their forms) and voltage doublers, since these are normally only used for low power applications where their operation is not critical. This is especially true of preamp supplies, which will almost always be followed by a regulator.

The full wave voltage doubler is still common in valve amplifier circuits and for preamp supplies.

Of the major types, the generally accepted voltage and current ratios are as follows ...

Rectifier Type	Filter Type *	RMS AC Input
Full Wave	Capacitor Input	DC x 1.2 **
Full Wave CT	Capacitor Input	DC x 1.2 **
Bridge	Capacitor Input	DC x 1.8

Table 5 - DC Output Vs. AC Input Current

* Choke input filters have not been included, since although they provide superior filtering and lower noise, they are very expensive to produce because of the sheer size of the inductance.
** This figure is for each winding of the transformer.

As can be seen from the circuits analysed so far, the figure for bridge rectifiers (AC current 1.8 times the DC current) is close enough to what was measured. It must be remembered that this is not an exact science, since there are so many variables to deal with.

Full Wave Rectifier

Unlike the bridge rectifier which uses 100% of the transformer winding at all times, the

full wave rectifier uses only 1/2 of the winding on each half cycle of the AC waveform. This leads to some additional losses, since the winding must be double the number of turns of that for a bridge rectifier. This usually means that the winding resistance is more than double that for a bridge rectifier.

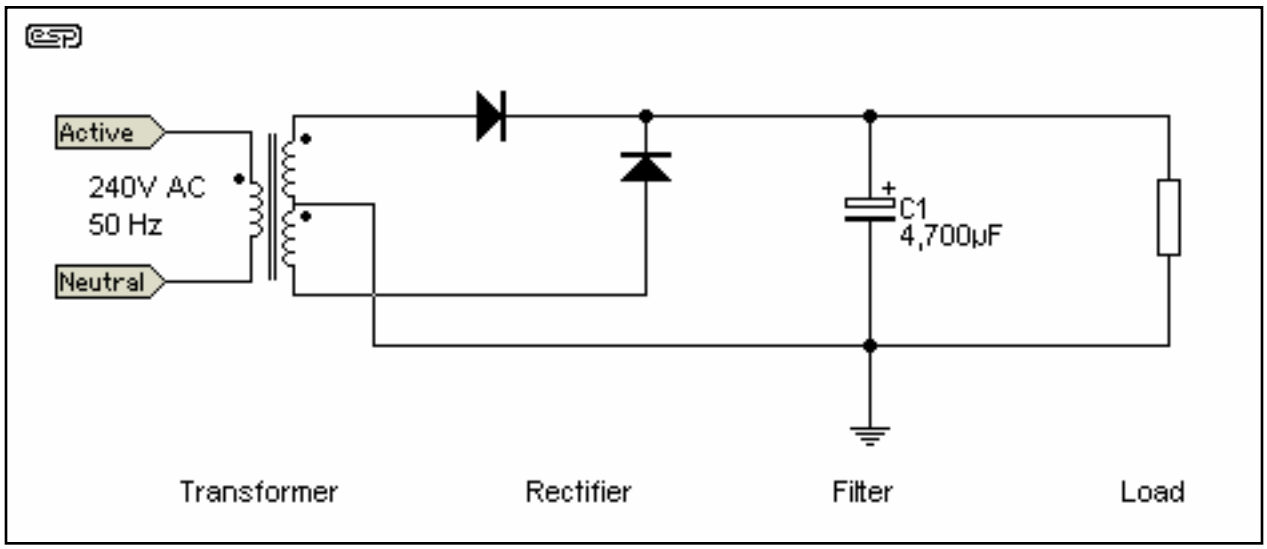


Figure 4 - Full Wave Rectifier

I have reverted to a transformer for this section, rather than the simulated version used before. This makes the drawings clearer, and the same depth of analysis will not be performed again this time. The capacitor ripple current is unaffected in principle, except that it will be slightly lower for a given VA rating since it is directly related to transformer winding resistance.

For the example in Figure 4, the AC current in each winding is 1.82A for a 1.47A DC load. This is quite close to the ratio of 1.2:1 shown in Table 5, and the difference is readily explained by measurement inaccuracy.

The biggest difference between a bridge rectifier and a full wave for a single output voltage is the transformer VA rating. Using the data from Table 5, and staying with a 25V (or 25-0-25 for the full wave) transformer and a nominal load of 1.5A ...

Bridge: Voltage = 25V, Current = 1.5 * 1.8 or 2.7A, so 25 * 2.7 = 67VA

Full wave: Voltage = 50V, Current = 1.5 * 1.2 or 1.8A, so 50 * 1.8 = 90VA

The relative inefficiency of the full wave is quite apparent, but this remains the most popular for power amps but in a modified form - the full wave centre-tapped (dual) power supply.

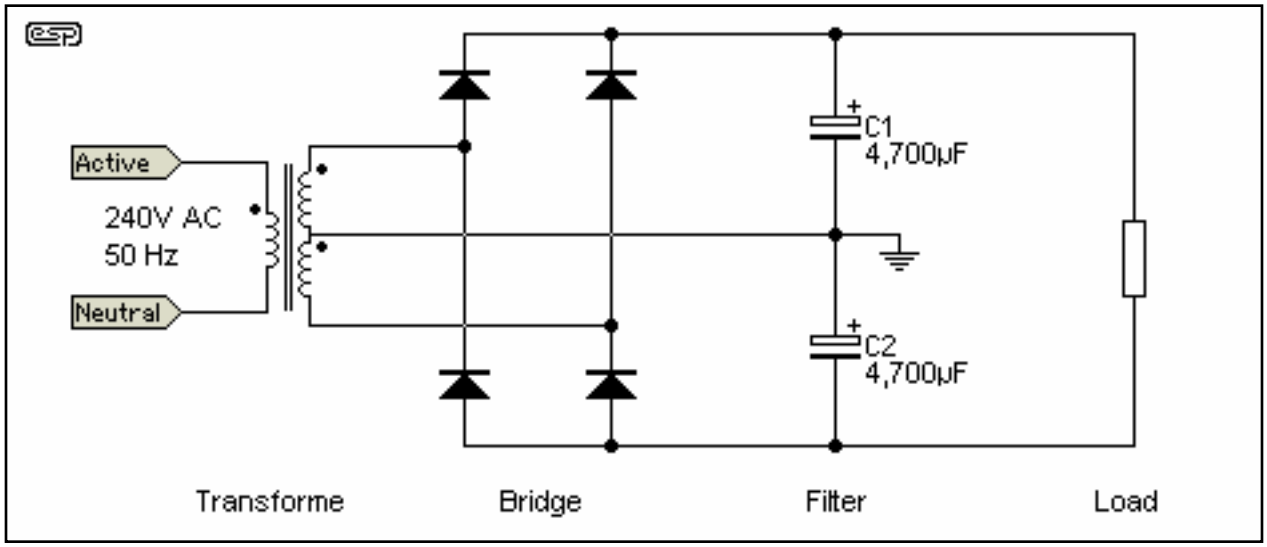


Figure 5 - Full Wave Centre Tapped Rectifier

This version now uses the entire winding all the time - each winding is used for both positive and negative supplies. Full winding utilisation means that the AC current is now the same as for a bridge rectifier, at 1.8 times the DC current, but only for a common mode load (i.e. between the supplies, rather than from one supply or the other to ground - this is identical to equal current from each supply to ground). Where the load is from only one supply or the other, the 1.2 rule applies, but power amplifiers will draw from both supplies (more or less) equally. The frequency of the signal waveform is mostly above the power supply input frequency, so the supply will effectively be loaded as common mode. All dual power supply designs must assume this load, or the result will be most unsatisfactory.

Temperature

In all power supply designs where the power is significant, the transformer temperature rise must be accounted for. Apart from the transformer radiating its heat to nearby components, any temperature rise will increase the copper losses, leading to reduced performance and even more heat. Use of a larger than required transformer will help considerably, but at the expense of capacitor ripple current.

Naturally, an increase in ripple current will cause the capacitors to become hotter, and as always, an increase in temperature causes increased losses and shortened component life expectancy.

For similar reasons, it is unwise to mount the filter capacitors anywhere near a significant heat source, such as large wirewound resistors, heatsinks or other heat generating components. Valve amplifiers are the natural enemy of electrolytic caps, due

to the often elevated temperatures within the chassis. Some manufacturers have resorted to mounting the filter caps in a separate metal enclosure on the outside of the chassis (and reasonably well away from output valves) in an attempt to keep temperatures down.

Capacitor Value

The required capacitance for a given load current and ripple voltage is determined (approximately) by the formula [1]...

$$C = (I_L / \Delta V) * k * 1,000 \text{ uF} \quad \text{... where}$$

I_L = Load current

ΔV = peak-peak ripple voltage
k = 6 for 120Hz or 7 for 100Hz ripple frequency

Since all my calculations above were done at 100Hz ripple current, this can be checked easily, so ...

I_L = 1.44, ripple = 2V p-p, therefore C = 5,040uF

It can safely be concluded that this formula is more than acceptable for our needs, the error being about 7%, which is far better than the tolerance of electrolytics anyway. The net result is that the required capacitance is about 3,500uF per amp, for a 2V peak to peak ripple (50Hz supply). The required capacitance will be less for 60Hz countries, at 3,000uF per amp - again for a 2V p-p ripple voltage.

I have seen it advocated that 100,000uF is the minimum that should be used with a powerful amp (say 200W / channel or so), but I find this difficult to justify. The law of diminishing returns comes into play quite quickly, and with both capacitance and transformer VA rating, this law becomes significant once you have doubled each of these values. With Class-AB amps, a ripple voltage of 2V P-P at full power will do nothing more than reduce the power by a few watts at the onset of clipping. Even reducing the capacitance further to (say) 1,500uF per amp will only reduce the continuous power output by a small amount. Normal music signals with their dynamic range will allow an amp with a relatively small capacitance to still provide the same maximum power for short transients.

As an example, a 100W/ 8 Ohm power amp will have a maximum output current of about 3.5A RMS into a resistive load. Since we know that a loudspeaker is not resistive, this figure should be doubled, to 7A. This is not an absolute rule - you can multiply by three if that makes you any happier. The supply current for each supply (+ve and -ve) will therefore have a peak value of about 10A (7 * 1.414), and the average will be 1/2 of the speaker current, or 3.5A. Based on the 3,500uF / amp figure above (assuming a 50Hz supply), 12,250uF per side is enough to ensure that ripple voltage never exceeds 2V P-P, but since this is a non-standard value, we would use 10,000uF.

In reality, it is probably going to be quite OK with 4,700uF, and the loss of power is negligible in real terms. As stated above, continuous power will be reduced, but normal music signals will not have transients of sufficient duration to discharge the filter caps appreciably. Note that this does not apply to amplifiers that are used for sub-woofers, nor for Class-A amps. These will place a greater load on the supply, and on a more consistent basis.

If we were to increase the capacitance to infinity (big cap!), the ripple voltage will be 0V, and we will get an extra volt (peak) from the amp before the onset of clipping with a continuous signal. This represents about 4W increase, which is completely insignificant - especially when we remember that the mains voltage could (will - at some stage) fall by up to 10%, which represents a power drop of around 20W - the 100W amp will only be capable of 80W with 10% low mains. Considering that an amplifier should not be operated at or near clipping anyway, the difference is inconsequential.

Use of a small (e.g. 1uF) polyester or polycarbonate capacitor across the DC output is a common practice. Electrolytics all exhibit a small inductance, and this causes their impedance to rise at higher frequencies. This is dependent on the physical size (specifically the length) of the cap - bigger caps usually have greater inductance. Again, the use of a paralleled bank of small (1,000uF) electros will be better in this respect than a single large can type, and will also be easier to mount and cheaper. I have never found the necessity to add a bypass cap to maintain amp stability, but it cannot hurt (some amplifiers *will* oscillate if the supply impedance is allowed to exceed some specific (low) impedance at high frequencies). Essentially it is a good idea, and in the greater scheme of things, is inexpensive. Should you choose not to include a film bypass, it is unlikely that anything 'bad' will happen - the impedance of the large electrolytic will remain much lower than that of the film cap at frequencies up to 1MHz or so.

Capacitor Ripple Current

The manufacturers' ripple current rating is the maximum continuous ripple if the quoted life expectancy of the capacitor is to be achieved (usually 2000 hours, but 12000 to 26000 hours for some manufacturers). The ripple current rating is for the capacitor operating in an ambient of 85°C (or higher for high temperature types)and the maximum ripple current can be increased by up to 2.5 times as the operating temperature is reduced (2.5 times at 30°C), though going above about 1.5 times is risky because the ESR increases as the capacitor ages and causes more heat for the same ripple current. [3]. Personally, I prefer not to exceed the quoted ripple current rating.

Capacitors in power supplies feeding Class-A amps should be operated well within their ripple current rating. In a Class-AB amp, the maximum ripple is at maximum output which only occurs occasionally (if at all!). Occasional excursions up to or even above the maximum ripple current will not significantly affect the life of the capacitor. In a Class-A, however, the ripple is at a maximum whenever the amp is switched on. If the ripple current is at the maximum for the capacitor, the life expectancy would be 2000 hours (for the normal types). This equates to a life of less than 2 years if the amp is used for 3 hours a day.

A formula for calculating ripple current would be very useful, but unfortunately (despite claims made in some articles I have read), it is almost entirely dependent on the series resistance provided by the incoming mains and the power transformer. Any formulae that do exist are only true for "sub-optimal" values of capacitance (in other words, the cap is too small to be useful). The summary (below) has some guidelines that may be useful, but be aware that these are guidelines only - the final outcome has so many variables that it is impossible to give an accurate prediction of capacitor ripple current.

Remember that large capacitor values will have a smaller surface area per unit capacitance than smaller ones, so the use of multiple small caps instead of a single large component can be beneficial. There is more surface area, the ESR will be lower, ripple current rating higher, and the combination will most often be cheaper as well. This is an "all win" situation - rarely achieved in any form of engineering. An example is worth showing - the following details are from an Australian electronics retailer's catalogue:

Value (uF)	Voltage	Size (mm) Dia x H	Surface Area (mm²)	Ripple Current (mA)	Price (AU $)
1,000	63	16 x 32	1659	1,400	1.95
2,200	50	16 x 35	1810	1,900	2.85
4,000	75	30 x 80	7634	4,600	14.50

8,000	80	35 x 76	8467	3,460	18.95
10,000	100	51 x 85	13779	8,100	32.95

Table 6 - Capacitor Comparison

For the sake of the exercise, we will assume that we need 8,000uF at a minimum of 50V, and with a ripple current rating of 7A - this is more than adequate for a 100W amp, and meets all the design criteria.

- A single 8,000uF 80V cap could be used, with a price of $18.95, ripple current of 3.46A and a surface area of 8467 mm² Overall, a simple solution but ripple current rating is 1/2 what I want, so it is excluded.
- Two 4,000uF 75V caps will cost $29, but have a ripple current of 9.2A and a surface area of 15,268 mm². Considerably more expensive, but very good performance.
- Four 2,200uF 50V caps cost $11.40, ripple current is 7.6A, and surface area is 7,240 mm². The most economical, but there is a minor performance deficiency by comparison with the previous and next options (but you do get some extra capacitance). This would be my choice for most systems, as it meets or exceeds all requirements.
- Eight 1,000uF 63V caps will cost $15.60. The ripple current is 11.2A, and surface area is 13,272 mm². For performance vs. price, there really is no contest. More effort is required to mount them, though.
- Naturally, you could also use the 10,000uF cap, but why would you do that?

I shall leave it to you to do your own comparisons, but most of the time you will get similar results. This is a very good way to reduce the size requirements as well - it is far easier to fit a number of small caps into an enclosure than a couple of large ones, and since there is an overall improvement in specs as well as a price advantage, it is an elegant solution.

It is worth pointing out that historically, filter capacitors are the number one cause of power supply failure. This is almost always because of the effects of temperature and ripple current, and close attention to this is very much worth your while.

Rectifier Diodes

One thing I strongly recommend is the use of 35A chassis mounted bridge rectifiers. Because of the size of the diode junctions, these exhibit a lower forward voltage drop

than smaller diodes, and they are much easier to keep cool, since they will be mounted to the chassis which acts as a heatsink. As always, lower temperatures mean longer life, and as was demonstrated above, the peak currents are quite high, so the use of a bigger than normal rectifier does no harm at all.

Even given the above, I have had to replace bridge rectifiers on a number of occasions - like any other component, they can (and do) fail. The bigger transformers increase the risk of failure, due to the enormous current that flows at power-on, since the capacitors are completely discharged and act as a momentary short circuit.

Diodes used in a FWCT (Full Wave Centre Tapped) or single Full Wave supply rectifier must be rated at a minimum of double the worst case peak AC voltage. So for example, a 25V RMS transformer will have a peak AC voltage of 35V when loaded, but may be as high as 40V unloaded, and double this is 80V. 100V Peak Inverse Voltage (PIV) diodes would be the minimum acceptable for this application.

For a single bridge rectifier, PIV needs to be greater than the peak AC voltage, since there are effectively two diodes in series.

In the case of a dual supply (using a 25-0-25V transformer), the worst case peak AC voltage is 80V, so the diodes must be rated at 200V PIV. The most common 35A chassis mounted bridge rectifiers are rated at 400V, and this is sufficient for all supplies commonly used for power amplifiers of any normal (i.e. < 500W into 8 ohms). Beyond this, the voltage rating is fine, but the current rating is inadequate, and a higher current bridge should be used.

There is currently a trend towards using fast recovery diodes in power supplies, since these supposedly "sound better". This is basically a bad idea, and there is absolutely no requirement for them. The purpose of a fast recovery (or any other fast diode) is to be able to switch off quickly when the voltage across the diode is reversed. All diodes will tend to remain in a conducting state for a brief period when they are suddenly reverse biased. This is extremely important for switched supplies, since they operate at high frequency, and have a squarewave input. Standard diodes will fail in seconds with the reverse current, since it causes a huge power loss in the diode.

At 50 or 60Hz, and with a sinewave input, the slowest diodes in the universe are still faster than they need to be. Indeed, Nelson Pass suggests that even the standard diodes should be slowed down with paralleled capacitors [2]. This can be an excellent idea, as it reduces the radiated harmonics from the diode switching. Typically, 100nF capacitors (optionally with a small series resistance) are wired in parallel with each diode

in the bridge, and this is quite common with some high end equipment and test gear where minimum noise is essential.

DC Sound

For reasons that I find completely obscure, there are some who claim that there are audible differences between power supply filter caps, diodes and mains leads. The net result of all the transforming, rectifying and filtering described above, is to produce DC. OK, it is not pure DC, in that it has some superimposed AC, as ripple voltage. Very few power amplifiers are so intolerant of ripple or other signals on their supply lines that a couple of volts will be audible. If this is the case, then a regulated supply, or at least a capacitance multiplier, is essential.

Very small amounts of noise that manage to get through the supply also should have no effect on an amp, and if the amp is so afflicted, the remedy is the same as for ripple. Remember that at all audio frequencies, the reactance of the capacitor is very low, and will act as a short circuit for any stray noise signals. A 10,000uF cap has a theoretical reactance (impedance) of 1.6 milliohms at 10kHz. This will never be achieved in practice - the wire has more resistance than that. Suffice to say that the impedance is pretty low, and it is extremely hard for any appreciable signal to get past the filter capacitors. The addition of one or more 1uF film type bypass caps ensures that impedance stays low even as the inductance of the electrolytics starts to have an influence.

The above assumes that the problem is noise, but this is rarely stated as the "improvement" - it is more likely to be bass "authority" or extension, or perhaps the "veil" is lifted from the highs. It can readily be demonstrated that as long as a power supply is well engineered in the first place, the "quality" of the DC output is unaffected by any of the "remedies". Ripple voltage will remain the same, amplifier power output (at all frequencies) will be unchanged, and mostly the signal to noise ratio will be unaffected. It is very easy to monitor the supply rails with an oscilloscope and a monitor loudspeaker (capacitively coupled of course), and from this it is possible to directly assess any difference should it exist.

In a similar vein, I have yet to hear from anyone who can give a plausible explanation or send me a sound file that demonstrates the difference between any two mains leads, all else being equal.

Summary

In conclusion, there are some rules of thumb that can be applied to save calculations and test measurements. These should not be considered as gospel - they are merely my suggestions on acceptable minimum requirements for a power supply.

Diode Ratings

- For most amplifiers or 50W or greater, use a 35A 400V rectifier bridge. Smaller amps can naturally use something with lower ratings. Current should be based on 10A per 100W for an 8 ohm load, or 20A per 100W for 4 ohms.
- Voltage rating must be a minimum of 200V for 100W into 8 ohms, or 100V for 100W for 4 ohms. This is not linear, so direct extrapolation is not advised. 400V diodes are adequate for amplifiers up to 2000W/ 8 ohms.

Capacitor Value

- Capacitance
  - Class-AB: The filter cap should be a minimum of 4,700uF per 100W into 8 ohms, and 10,000uF per 100W into 4 ohms. The actual value can be extrapolated from here. It is often possible to use less than the suggested capacitance, and the values shown may be halved without a significant loss of performance - this is up to the individual constructor.
  - Class-A: The cap needs to be selected on the requirements of the amplifier, but a minimum of 4,700uF for each 10W into 8 ohms is recommended.

- Voltage rating is as required by the maximum DC supply, and will usually be 50, 75 or 100V

- Ripple current ranges (from the examples shown here) from 3.3 times the load current up to 4.6 times load current, depending largely on the transformer size.
  - Class-AB: Because the amplifier will rarely draw maximum current for prolonged periods, a ripple current rating equal to double the peak output current will usually be enough. Thus a 100W/ 8 ohm amp having a peak output current of 3.5A will usually be OK with a ripple current rating of 7A. (Note: guitar amplifiers are specifically excluded from this! - I suggest double the above value as a minimum.)
  - Class-A: For the worst case possibility, I suggest that the ripple current rating should be 5 times the load current for these amplifiers. A 20W/ 8 ohm Class-A amp will draw a continuous 2.5A (typical), so ripple current rating for the capacitors needs to be 12.5A.

Transformer

- The voltage is determined by the power that you need from the amplifier. Calculate power from the formula ...

  $P = Va^2 / R$ Where P is power in Watts, Va is RMS speaker voltage, and R is speaker nominal impedance

- The supply voltage (allowing only for basic losses) is calculated as follows

  $V_{RMS} = Va * 1.1$ Where V_{RMS} is the transformer secondary voltage (for each supply rail)

- **VA Rating - Class-AB**

  The absolute minimum VA rating suggested is equal to the amplifier power. A 50W amp therefore needs a 50VA transformer, or 100VA for stereo 50W amps. Larger transformers (up to double the amp power rating) will provide a "stiffer" power supply, and this may be beneficial. For continuous operation at full power (never actually needed in audio), the transformer should have a VA rating of up to 4 times the amplifier power.

  It is suggested by some transformer manufacturers (and no doubt gleefully adhered to by many amplifier makers) that the VA rating needs only to be 0.7 of the maximum amplifier power. While this will work well enough in most cases, you will not have a "stiff" power supply - I would suggest "soggy" as an appropriate term - the voltage will waffle about like a limp biscuit.

- **VA Rating - Class-A**

  The minimum VA rating suggested is 4 to 5 times the amplifier power. A 20W Class-A amp therefore needs an 80-100VA transformer, or 160-200VA for stereo 20W Class-A amps. Larger transformers will again provide a "stiffer" power supply, and this will be beneficial. Nelson Pass suggests the ratio should be 7.5 times the amp power [2].

Fusing and Protection

Since the power supply is connected to the mains, it is necessary to protect the building wiring and the equipment from any failure that may occur. To this end, fuses are the most common form of protection, and if properly sized will generally prevent catastrophic

damage should a component fail.

Toroidal transformers have a very high "inrush" current at power-on, and slow-blow fuses are highly recommended to prevent nuisance blowing. In the case of any toroid over 500VA, a slow-start circuit is very useful to ensure that the initial currents are limited to a safe value. An example of such a circuit is presented on the Project Pages, and represents excellent insurance against surge damage to rectifiers and capacitors.

Calculating the correct value for a mains fuse is not easy, since there are many variables, but a few basic rules may help.
First, determine the maximum operating current of the system, based on continuous maximum power. The calculations done previously will help ...

The mains current is determined by the turns ratio of the transformer, calculated by

> Tr = Vpri / Vsec Where Tr is turns ratio, Vpri is primary (mains) voltage, and Vsec is secondary voltage

A 240V to 25-0-25V (i.e. a 50V secondary! - the whole secondary winding must be used in the calculations) transformer has a turns ratio of 4.8:1, the same transformer with a 120V secondary has a turns ratio of 2.4:1 - this can be calculated for any transformer. The primary current is calculated by ...

> Ipri = Isec / Tr Where Ipri is primary current, Isec is secondary current

A supply designed for our hypothetical 100W/ 8 ohm amplifier will have a secondary current of about 6.3A at full power, so primary current (for 240V) is about 1.3A. A 2 Amp fuse would seem OK for this, but if a 500VA transformer was used (for example), this is barely enough to handle the maximum transformer primary current. A 3A slow-blow fuse should be used, and this should survive the inrush current (120V countries will need a 6A slow-blow fuse).

Thermal protection (often by way of a once-only thermal fuse) is often included in transformers. If the "one-time" thermal fuse has been used, should the transformer overheat it must be discarded, since the fuse is buried inside the windings and cannot be replaced. All the more reason to ensure that the transformer is properly protected at the outset.

Note that primary fuse or circuit breaker protection does not protect the amplifier against overload or shorted speaker leads. If this happens, or should the amplifier fail, the primary fuse offers no protection against catastrophic failure and possibly fire. For this reason, secondary DC fuses should always be used - no exceptions.

Multi-tapped primaries (e.g. 120, 220, 240V) create additional problems with fusing, and often a compromise value will be used. The transformer protection is then not as good as it could be, but will generally still provide protection against shorted diodes or filter caps.

Additionally, it may be an advantage to fit Metal Oxide Varistors (MOVs) to the mains - between the active and neutral leads. These will absorb any spikes on the mains, and help to prevent clicks and pops coming through the amplifier.

Disclaimer

The information presented in this article is intended as a guide only, and the author takes no responsibility for any damage disfigurement or injury to persons (including but not limited to loss of life) or property that results from the use or misuse of the data or formulae presented herein. It is the readers' responsibility absolutely to assess the suitability of a design or any part thereof for the intended purpose, and to take all necessary precautions to ensure the safety of himself/ herself and others.

The reader is warned that the secondary voltages present in nearly all power supplies for amplifiers are potentially lethal, and to observe all applicable laws, statutory requirements and other restrictions or requirements that may exist where you live.

WARNING: All mains wiring should be performed by suitably qualified persons, and it may be an offence to perform such wiring unless so qualified. Severe penalties may apply.

All power supplies must be fused or protected by an approved circuit breaker, and all mains wiring must be suitably insulated and protected against accidental contact to the specifications that apply in your country.

References

1. Voltage Regulator Handbook - National Semiconductor Corp. - 1981 Edition
2. Manufacturers Report - "The Importance of the Power Supply" - May, 1997 (By

Nelson Pass)

3. Information supplied by Geoff Moss, my unofficial editor from the UK. As always, thanks Geoff!

Main Index

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Last Update: 18 Oct 03 - Minor changes only./ 24 Mar - some minor reformatting, and some additions./ 03 Mar - added protection, disclaimer and cap comparison info./ 25 Feb 2000 - created as a new page (formerly part of amp design)

The Audio Pages

Elliott Sound Products **Matching Power and Driver Transistors**

Copyright © 2001 - Rod Elliott (ESP)
Page Updated 24 May 2005

Articles Index

Main Index

Contents

- Introduction
- Test Equipment
- Basic Test Process
- High Current Test
- Conclusion

Introduction

Whether you use Bipolar Junction Transistors (BJTs) or Metal Oxide Semiconductor Field Effect Transistors (MOSFETs), there are many circuits that suggest (or require) matched transistor pairs. Some retailers sell matched devices, but they are rather expensive, and not commonly available.

Matching power transistors is theoretically easy, but in reality there are quite a few parameters that should be matched to get true matched pairs. The situation is made a little harder when you have PNP and NPN (or N-Channel and P-Channel) devices, since the simpler comparison (or bridge) techniques don't work due to the opposite polarities.

This article is not intended to cover all the possibilities, since the equipment needed is out of the range of the average hobbyist. Bear in mind that some of the tests are potentially destructive if the DUT (device under test) does not have an adequate heatsink, so the provision of a fairly substantial heatsink is essential. Ideally, a quick clamping method will be utilised so the devices

don't have to be screwed down each time - this can get seriously time consuming.

Manufacturers get around the heatsink problem by pulse testing (so the device does not get a chance to get hot), but this requires expensive equipment to do, so a simpler test is needed. Simpler actually makes it more complicated for high current tests, since you will have to mount the transistor before you start. For these tests, using mica washers is out of the question (too time consuming), so the heatsink will be at the potential of the collector (BJT) or drain (MOSFET).

You will also need a fairly substantial power supply if you want to test at high current. You must also be aware of the safe operating area of the device, since exceeding this will destroy BJTs very quickly. MOSFETs are a little more forgiving, but will still fail if pushed too far.

WARNING - The rated source to gate voltage of MOSFETs must never be exceeded

---

Test Equipment

Ideally, you will have two multimeters, but one can be used if you don't have two. The meters will most commonly be digital, and a current range up to at least 2A may be useful (but not essential).

The power supply should ideally be regulated, but again, if you don't have one, then unregulated will still work, although mains voltage variations will affect the accuracy of measurements. In reality, small variations will not cause great errors, and the final result will still have some variations - finding two identical devices is not normally expected or obtained.

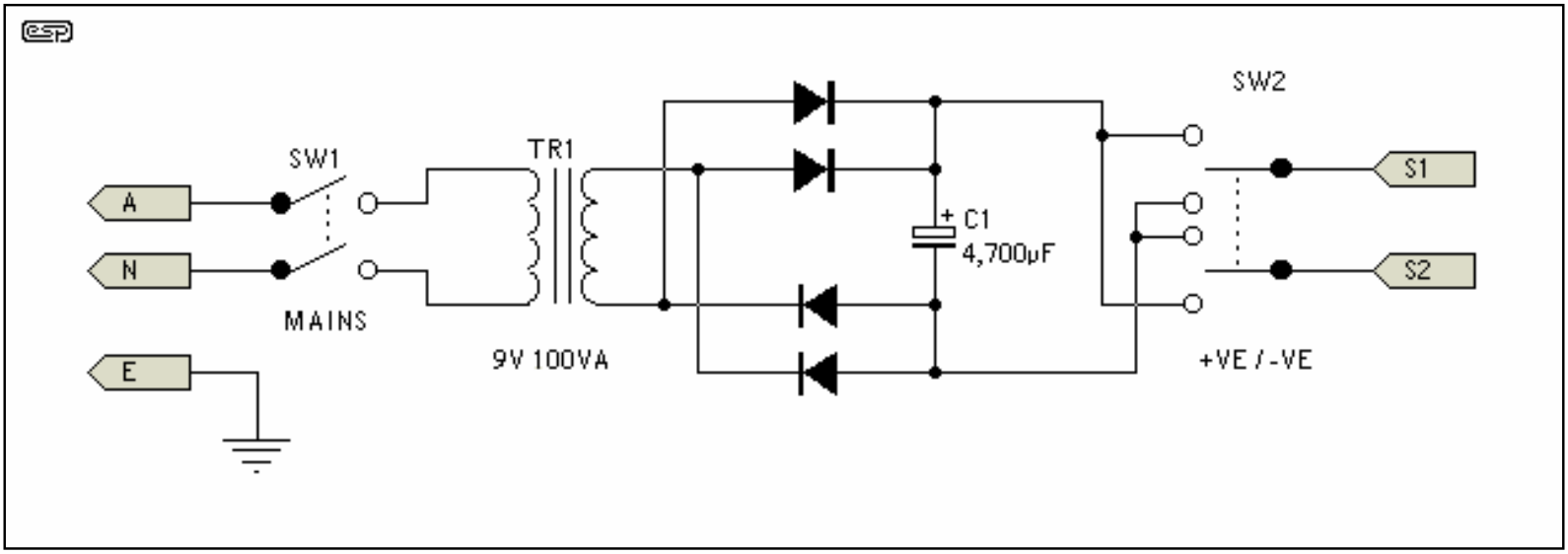


Figure 1 - Power Supply (Unregulated)

You will also need some high power resistors. The actual power rating depends on the supply voltage and the test current. For the suggested 12V supply, a maximum test current of 2A is

reasonable, so the resistors should be set up as shown - four 1 Ohm 5W resistors (as shown below) will be more than acceptable for the high current tests. The others should be as shown in Figure 2.

The bridge rectifier needs to be a high current type, and given that 35A bridges are quite cheap, this is a good choice. I suggest that the power transformer should be fairly substantial, or excessive voltage sag will be experienced during testing. A 100VA trannie will be more than sufficient.

The filter cap is not even required, and leaving it out will be very useful if you have an oscilloscope, since the waveform can be observed for matching as well as the current. Having said that, my preference is to install the capacitor anyway, and using an oscilloscope is likely to be more trouble than it's worth.

The heatsink should be designed so that transistors can be mounted and unmounted easily and quickly, or the task will become a chore very quickly. A strong spring clamp will normally be sufficient to keep the transistor clamped to the heatsink and nice and cool during the test, which should still be done reasonably quickly so that device heating does not skew the results. In all cases, the test duration should be kept the same for each device, and a fan is useful one the heatsink to ensure that heating over time does not cause errors (these can become significant with even a small increase of heatsink temperature). The fan can be powered from the supply shown.

Basic Test Process

The test itself is fairly simple. The first qualifier is for gain, and the emitter-base voltage for BJTs at a known current, or the source-gate voltage for MOSFETs, again for a known current. This first test should be done with the current set for the quiescent current (per device). Figure 1 shows the set-up, and it is deliberately fairly simple. This will work for BJTs and MOSFETs, without any changes. Remember that when testing MOSFETs, the gate is static sensitive, so appropriate precautions must be taken so the DUT is not damaged. Never exceed the rated source-gate voltage - ***ever!***

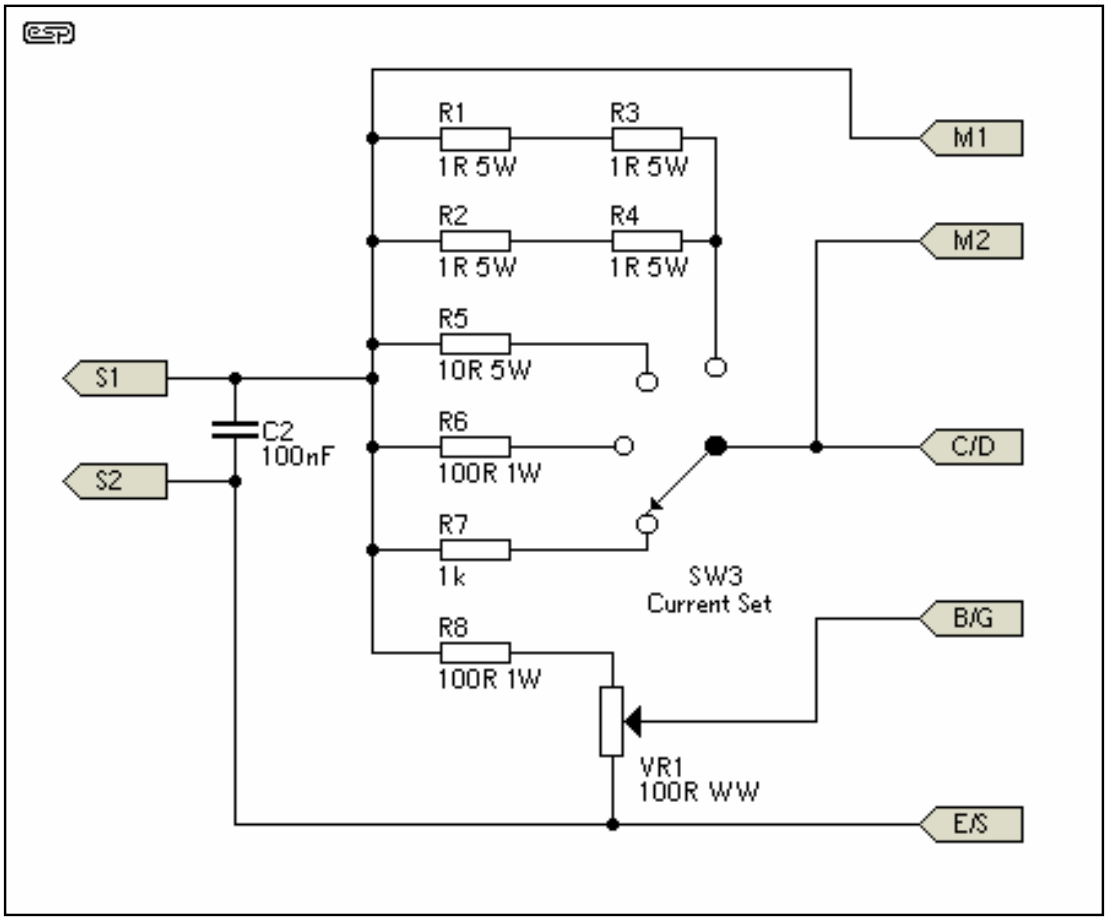


Figure 2 - Quiescent Current Gain / Voltage Test

VR1 needs to be a wirewound pot for testing BJTs, but can be a higher value carbon pot for MOSFETs. I suggest the wirewound pot as a matter of course, since it increases the usefulness of the test jig. The pot will get quite warm in operation (it dissipates nearly 1W), so make sure you switch off the power between tests. Test leads should be fitted with colour coded alligator clips for emitter/source, base/gate and collector/drain (BJTs and MOSFETs respectively).

Select a transistor at random from those on hand, and connect it into the test jig. Make sure that the pot is set to minimum, and that the correct polarity is selected first! Set the switch to the 10 Ohm range, and adjust the pot until the voltage across the M1 and M2 terminals is equal to 10 x Iq (in Amperes). If you want to use a quiescent current of 100mA, the voltage across the resistor will be 10 x 0.1 = 1V. Remember to divide the total quiescent current by the number of parallel output devices first.

Now you can measure the emitter-base or source-gate voltage. Note the readings, and mark the transistor just tested (so you can correlate the device with its measured characteristics). Repeat the test with the other devices you have, reversing the polarity as required - ***but do not adjust the pot*** - leave it exactly where it was for the first transistor. Bear in mind that the current in subsequent transistors could vary by several hundred percent, so it may be necessary to readjust the pot and re-run the test from the beginning :-(

For each device, note the emitter-base (or source-gate) voltage, voltage across the resistors (or

the current through them), and the reference number you marked on each device. When you have finished, you will have an array of voltages and currents (calculated from the resistor voltage if you don't use a current meter), and you can choose those devices that are the closest match. Generally, only the current is really significant except for devices in parallel, where the turn-on voltage becomes important. The source-gate voltage will not change if you are testing MOSFETs, so that only needs to be measured once.

By using a higher resistance, smaller devices can be tested in the same way. For driver transistors, the 100 ohm range would be satisfactory, and for small signal transistors, use the 1k ohm range. Be very careful that you keep the current and voltage within the device ratings! The rotary switch on the supply is designed for just this purpose.

The test will match devices so that they are approximately equal at the all-important crossover region - if desired, you can test at an even lower current, to ensure that there is the least possible error between devices, however the tests become very time consuming, and obtaining devices that are fully matched over the full operating range is somewhat unlikely.

High Current Test

Once you have a selection of transistors that have approximately equal low current characteristics, you can do a high current test if you want to - this uses the 1 ohm range on the rotary switch. I don't recommend that you exceed 2A, unless you are sure of what you are doing.

The test is set up in exactly the same way as before, except the current is increased to the desired test value. For each device tested, make sure that the test duration is maintained for the same time - say 10 seconds. You will need to wait for the heatsink to cool to the same temperature (or near enough to it) between tests. A fan will help here, and is highly recommended.

Again, you will note the exact current of each device with the pot in the same position as for the first transistor tested. At the end, you will have a set of figures that show the closest matching devices out of those available. I strongly suggest that you don't expect miracles - if you can get transistors that measure within 10% of each other for both the high and low current tests this is a good result. You may do better, but don't count on it, and don't get all depressed if you have to accept a wider tolerance.

For the brave (and those who have taken the time to create a really solid heatsink test set), you can run further tests at higher currents, but you will need to be extremely careful. Remember that 2A at 12V is 24W continuous dissipation, and this will heat the transistor under test very quickly - higher currents will create even more heat.

With the 1 ohm resistor, transistor dissipation is reduced only slightly, and even 24W is a lot to get rid of in any test environment. In extreme cases you will have to use thermal compound for

the DUT to prevent overheating, since it will not normally be possible to take a reading quickly enough unless you have access to a digital storage oscilloscope. If you have access to one, then pulse tests are recommended for all high current testing. The way to go about this is beyond the scope of this little article though :-)

An alternative approach is to take your measurements at high and low currents,, and then calculate (or plot) the transfer characteristics of each device. Although time consuming, this should give good results. Ideally, you will take measurements at a minimum of three points. Measure collector/drain voltage (and for MOSFETs, the source to gate voltage) at ...

- the expected design quiescent current
- close to the peak expected current, and
- a point midway between the two

Once you have these figures, you can then select those that represent the closest possible match. You will be very fortunate indeed if you get exact matching, but it should be possible to get several acceptably matched pairs from a batch of reasonable size (you will probably need a minimum of 10 devices, if possible from the same manufacturing batch.

Conclusion

The tests described are not the most accurate known, but will be well within the abilities of hobbyists. The results can be expected to be very good when used for matching, and the selected pairs will be much closer than random selection will ever give you.

The test set is fairly cheap to make, and can be used for all sorts of transistor testing - not only for matching, but to test that transistors are functional. Since it will work with bipolars and MOSFETs, it has enhanced usefulness over most conventional transistor testers for the majority of basic testing needs. The little extras added just make it that much more useful.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 06 Dec 2001

Elliott Sound Products **Project 56**

Variable Amplifier Impedance

Rod Elliott (ESP)

Introduction

Is this a project or an article? Since you may have arrived here from either index, you might be a little confused. This is fine, as it has elements of both, but lacks the in-depth analysis of other articles, and also lacks the application of a final design that is common to my projects. This is for the experimenter. No further work will be done to refine formulae or produce "magic" spreadsheets to allow you to determine the impedance that is best for a speaker / enclosure combination or anything else. Readers' contributions are of course welcome, and I would be interested to hear about any massive improvements that were made to systems using these techniques. (Or minor improvements, for that matter.)

The idea of being able to vary the output impedance of a power amplifier has been around for a long time. I have used these techniques for over 30 years in various designs, and much as I would like to be able to claim otherwise, I was by no means the first.

The technique has been used to drive spring reverb units, and various other transducers where current drive is preferable to voltage drive. The latter is what we all strive for with power amplifiers - a perfect (ideal) voltage amplifier has zero ohms output impedance, and the amplitude does not change as the load varies. Loudspeakers are very non-linear loads, and the impedance will change at different frequencies for all sorts of reasons.

So voltage drive maintains a constant voltage across the load, while a current drive

circuit will produce exactly the same current into the load - in both cases regardless of impedance. In reality neither approach is ideal ...

- Voltage drive will produce greater power into the load as the impedance falls, since the load voltage remains constant, and a lower impedance means more current and thus more power.
- Current drive has exactly the opposite effect, so when the load impedance rises, a greater voltage is needed to maintain the same current. So with the same current but a greater voltage there is again more power, but now it is with higher impedances rather than lower.

Most loudspeaker systems are optimised for voltage drive, however in many cases "optimised" is overly optimistic, since they are just assembled and the builder hopes for the best. A tweak here, an attenuator there, and a zobel network somewhere else rounds off the process and yet another "world's best" loudspeaker system is born.

Not so much a real project as a field day for experimenters, this article describes the methods you can use to tailor an amplifier to a specific driver. Much is empirical (i.e. design by experiment, trial and error), and although formulae are possible, the load presented by a typical loudspeaker is very complex and leaves big holes in the final result if the maths are allowed to take over. Having said that, good results can be obtained by either method (theoretical or practical), so here goes ...

Description

Figure 1 shows two amplifiers (shown as opamps), with (a) being conventional voltage drive, and (b) is current drive. The output impedances are in the order of zero ohms and infinite ohms in each case (this IS a theoretical discussion at the moment!). The gain will appear to be exactly the same for each into an 8 ohm resistive load, but will change based on load for the current amp. We shall see that this is not really the case - some power is lost in the series feedback resistor. The voltage amp has the same gain regardless of load.

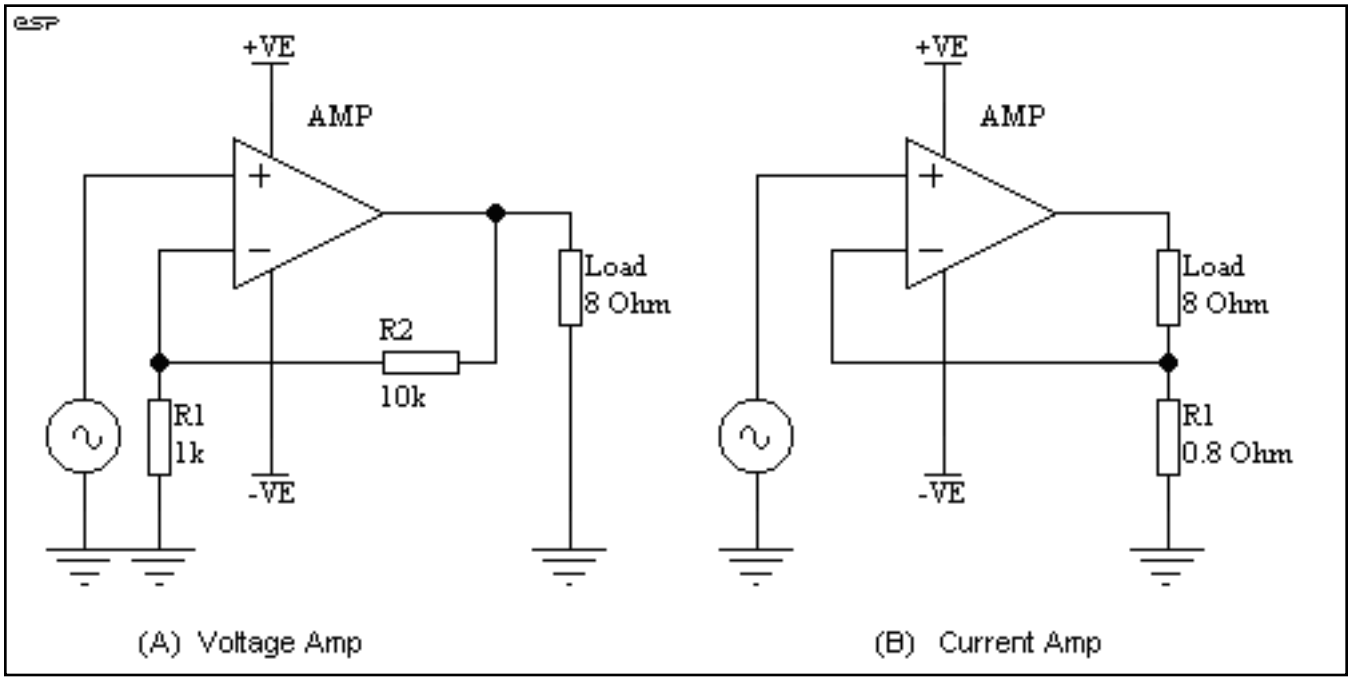


Figure 1 - Voltage and Current Amplifiers

The feedback is applied differently to achieve the desired result, as can be seen, although the difference is not at all subtle at first glance, as you look more closely you will see that it really is a simple rearrangement of feedback resistances. With the 8 ohm load, both amps have a gain of 11, but the series feedback in (B) actually means that the gain to the load is 10. With 1V AC signal applied, the voltage amp will develop 15.125W (11V at 8 ohms) into the load. The current amp will develop 12.5W (10V at 8 ohms) into the load, with 0.125W dissipated as heat in R1. This will remain constant for all load impedances.

Where things get interesting is when the load impedance changes. Should the load increase to 16 ohms (such as at resonance), the voltage amp will produce the same voltage, so power is halved. The current amp will provide the same current, so power into the load is increased fourfold to 50W, with the same 0.125W lost in the series feedback resistor. This of course assumes that the power supply voltage is high enough to allow the amp to do this without clipping.

Likewise, with a dip in impedance to 4 ohms, the voltage amp will provide double the power (30.25W - again assuming the the supply can now provide the necessary current), but the current amp will only produce 6.25W under the same conditions.

Which of these approaches is correct? For resistive loads it does not matter - both will perform identically except for the slightly lower gain of the current amp (this is easily compensated for). With loudspeaker loads, neither is ideal. Power to the load will vary widely depending on the impedance, which in turn depends upon frequency. The voltage

amp will create a dip in response at each frequency where the impedance rises (and a rise wherever there is a lower impedance), and the current amp will do exactly the opposite.

Mixed Mode Feedback

By applying a mixture of both forms of feedback, it is possible to define the output impedance at any value between the two extremes. To emulate a valve amp for example, an output impedance of about 4 to 6 ohms is needed, assuming an 8 ohm load (this is the assumed nominal load impedance for all examples cited).

Figure 2 shows the arrangement used, and with the values shown the output impedance is a bit under 4 ohms. We could simply use a 4 ohm resistor in the amp's output, but this will waste much of the amplifier power, which will be dissipated in the resistor instead of the load. The same power output is available from a current amp as a voltage amp (give or take a fraction of a dB to account for the series feedback resistor).

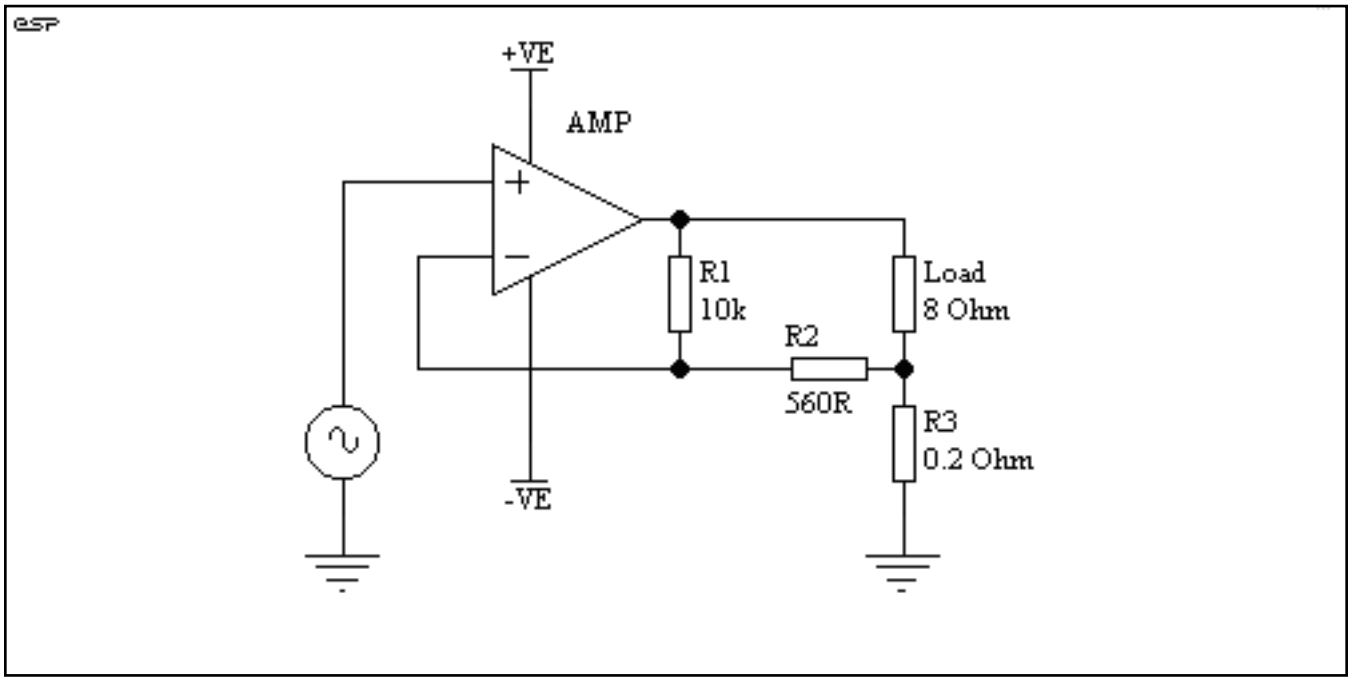


Figure 2 - A Four Ohm Output Impedance Amplifier

As can be seen, there is minimal additional complexity to achieve this result, and in my experience the final exact impedance is not overly critical, given the "real world" variations of a typical loudspeaker.

The no-load voltage is 18.8V with an input of 1V, and this drops to 12.8 at 8 ohms, and 9.7V with a 4 ohm load. These voltages are measured across the load, ignoring the

voltage drop of the series feedback resistor.

From this, we can calculate the exact output impedance from

$I_L = V_L / R_L$	(where I_L=load current, V_L=load Voltage and R_L=load resistance)
$Zo = (Vu - V_L) / I_L$	(where Zo=output impedance, Vu=unloaded voltage, V_L=loaded voltage)

$I_L = 12.8 / 8 = 1.6$

$Zo = (18.8 - 12.8) / 1.6 =$
$6 / 1.6 = 3.75$ ohms

So we have created an amp with an output impedance of 3.75 ohms, with very little loss (just over 0.5W is lost in the 0.2 ohms series feedback resistor with 20W output into 8 ohms). To see if this is useful, we will now have a look at what happens when the load impedance doubles or halves.

With a 16 ohm load, the power into the load falls to 14.6W, or about 1.37 dB. Contrast this with the conventional low impedance amp whose power will fall by 3dB (i.e. half).

When the impedance is reduced to 4 ohms, the output power is now 23.5W (an increase of 0.7dB), while a conventional amp would be producing 40W - an increase of 3dB.

Somewhere, there is a magical impedance that will give almost the same power into any load from double to half the nominal, give or take a dB or so. I am not about to test all possibilities, but having experimented with the concept for many years I am quite convinced that there are practical benefits to the use of modified current drive, where the impedance is defined. The exact impedance will depend to a very large degree on just what you are trying to achieve.

Further Applications

It has been suggested that loudspeaker intermodulation distortion is dramatically reduced by using a high impedance source (but by whom I cannot recall - I do know that one site I looked at was Russian, and a reader sent me a translation many months ago). I have experimented with this idea to some extent, but have been unable to prove that

this is the case - at least with the drivers I tried it with.

This does not mean that the claim is false, but I am unable to think of any valid reason that could account for such driver behaviour. It is interesting anyway, and some of you might like to carry out a few experiments of your own. I would be most interested to hear about your results should you decide to test this theory - preferably more thoroughly than I did.

By adjusting the impedance of an amplifier, the total Q (Qts) of a loudspeaker can also be altered, so driver behaviour in a given sized box can be changed. This can be used to adapt an otherwise unsuitable loudspeaker to a speaker enclosure, but does have limitations in terms of the overall variation that can be achieved.

More variation can be achieved by virtue of the fact that it is now possible to either retain or increase the power delivered to a loudspeaker at resonance, so that the ultimate -3dB frequency may be lowered from that theoretically claimed for a loudspeaker / enclosure combination. Care is needed, since too much additional power will make the speaker boomy, and usually additional internal damping is needed to compensate for the minimal damping factor provided by the amplifier. With the amplifier output impedance set at 4 ohms, damping factor into an 8 ohm load is 2 - a far cry from the figures of several hundred typically quoted. These (of course) fail to take into consideration the resistance of the speaker leads, and loudspeakers themselves are usually compromised by the crossover network, so the damping factor is not always as useful as it might seem.

The results of using modified impedance can be very satisfying, allowing a useful extension of the bottom end. My own speakers are driven from a 2 ohm amplifier impedance, and there is no boominess or other unpleasantness, but a worthwhile improvement in bass response is quite noticeable.

Negative Impedance

Again, this has been about for many years, but I have only found one driver type that seems to obtain any improvement from its use - horn loaded compression drivers. All cone speakers (including horn loaded) sound worse with negative impedance, but you might have some weird driver that can benefit from a negative impedance amp.

As the name suggests, when a negative impedance amp is loaded, the output voltage rises. The greater the load, the more output is applied. This is very risky, and negative impedance amplifiers can easily oscillate when connected to a reactive load such as a

loudspeaker. Indeed, negative impedance oscillators have been with us for many years in RF (and other) work, and there are quite a few electronic components that exhibit negative impedance. So, not new, but interesting.

Figure 3 shows how the circuit is rearranged to accomplish this most bizarre of ideas. By simply substituting the "real" earth connection for the junction of load and series feedback resistor, the voltage developed across the resistor now increases the gain by adding to the input signal voltage.

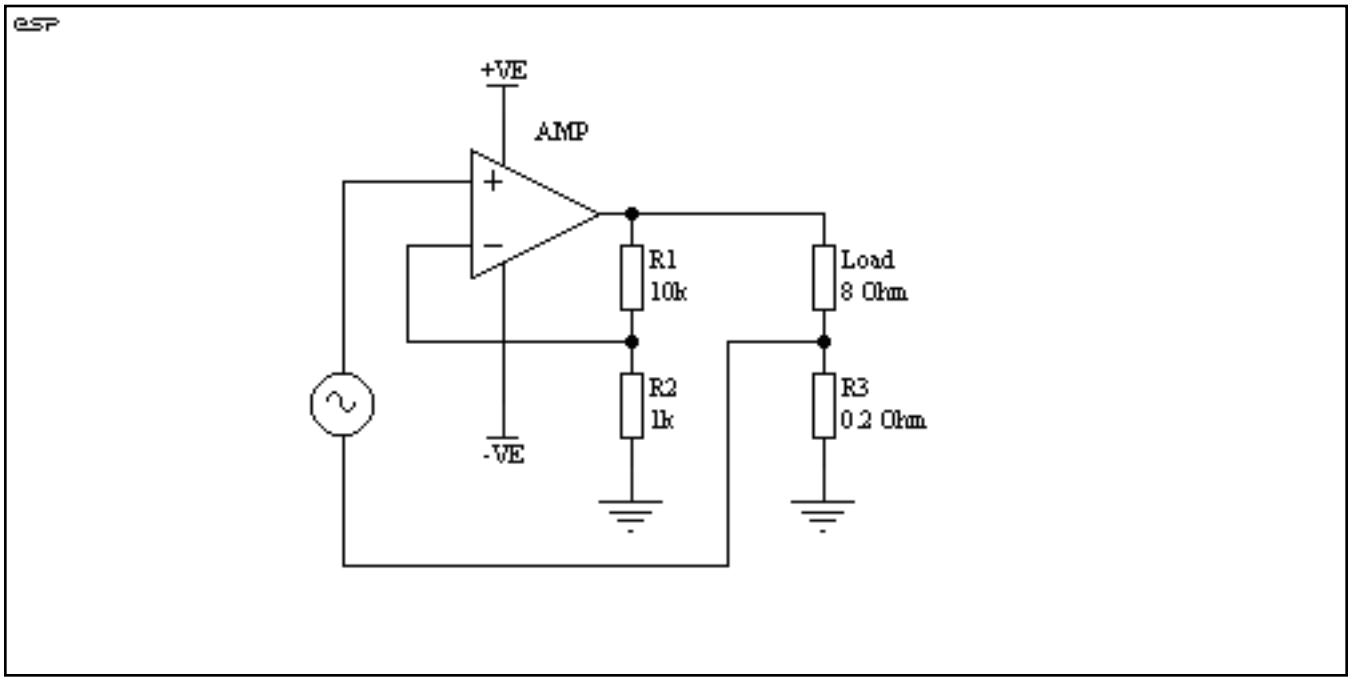


Figure 3 - A Negative Impedance Amplifier

This circuit with the values as shown will provide an open circuit load with 11V (as with a conventional amplifier as shown above), but when the load is applied the voltage increases as the load resistance is reduced. With the values shown, the load voltage will increase from 11V to 14.7V with an 8 ohm load, rising to 22V with a 4 ohm load. A quick calculation using the formulae above will show that the output impedance is about -2 ohms.

The negative impedance amp is by its very nature unstable - the output voltage will continue to rise as the load is reduced, until at some point positive feedback exceeds negative feedback and the circuit will oscillate. Another undesirable side-effect is that distortion is increased, because negative feedback (which reduces distortion) is being counteracted by positive feedback. Again, this is a non-linear function, and the results are unpredictable at best. Reactive loudspeaker loads can cause a negative impedance amp to oscillate, either at the frequency between the loudspeaker's resonance and box resonance (where impedance falls), or at some other frequency determined by

crossover components. Such oscillation can damage speakers (and ears!), so care is needed to ensure that this cannot happen. Low values of negative impedance (not less than -4 ohms) are strongly recommended.

I have found that only relatively small amounts of negative impedance are useful in practice. For example, one could use negative impedance to remove the resistance of the speaker cable and crossover components, although the results will not be as good as expected, and probably far worse. Even with a negative impedance of about 1 ohm, most speakers will show signs of "displeasure", and amplifier distortion will be increased - usually by an amount that is disproportional to the feedback factor.

Conclusions

This "project" is not really a project at all (which is why it has also been included in the articles index), but more of a starting point for experimenters. The circuits shown will all work with "real" amplifiers, but great care and considerable testing are needed to ensure that the results you actually obtain are providing a real benefit.

It is too easy to make a change such as shown here, and fully believe that the result is an improvement, where in reality (as eventually discovered after extensive listening and comparison) the opposite is true. While negative impedance may be found useful for horn compression drivers, it is unlikely (in my experience) to provide any benefit at all to cone loudspeakers.

Positive impedance can produce an improvement in bass response, but the cost can be high - boomy, over-accentuated bass at resonance, usually accompanied by a loss of bass "speed" and definition. There will be more freedom for the speaker cone to waffle about after the signal has gone ("overhang"), and it is rare that a speaker driven by a higher than normal impedance will perform well without additional damping in the enclosure.

There is no doubt that at output impedances in the order of 4 to 6 ohms your amp will sound more like a valve amp (but without the distortion), but it is up to you to decide if this is what you really want to do. The technique works well for guitar amps, as it allows the speaker to add its own colouration to the sound, which adds to the overall combination of distortion and other effects to produce pleasing results. For Hi-Fi the case is less clear, and experimentation is the only way you will ever find out for sure.

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright (c) 22/23 April 2000./ Updated 25 Nov 2003 - Corrected error in negative impedance section.

The Audio Pages

Elliott Sound Products **Baffle Step Compensation**

Copyright © 2001 - Rod Elliott (ESP)
Page Updated 08 Dec 2001

Articles Index

Main Index

Contents

- Introduction
- Baffle Step Response
- Baffle Step Equalisation
- Variable Equaliser
- References

Introduction

The so called "baffle step" is an increase in output from a loudspeaker as the size of the baffle becomes significant in terms of the wavelength of sound for a range of frequencies. At low frequencies, where the baffle (the panel the loudspeaker is mounted on) is small compared to the wavelength, the speaker is assumed to be operating with a spherical radiation pattern. While this may be the case should the speaker be situated up a tree in the middle of a field, it hardly qualifies as true when the same loudspeaker is installed in your listening room. There may of course be something about you that I don't know (for example that your speakers really ***are*** up in trees), but for the majority this will not be the case.

As frequency increases, the size of the baffle becomes significant, and the spherical radiation pattern no longer applies. This is also partly to do with the loudspeaker drivers

themselves - as frequency increases, typical cone drivers become more and more directional anyway - this occurs as the dimensions of the cone become significant with respect to wavelength.

Baffle Step Response

It has been shown by Olson et. al. that the optimum enclosure for a loudspeaker driver is a sphere - that is to say that there is no "baffle" as such, but that the driver is installed in a spherical enclosure. While this is all very well in theory, spheres are hard to build, and even harder to include in a typical lounge room unobtrusively.

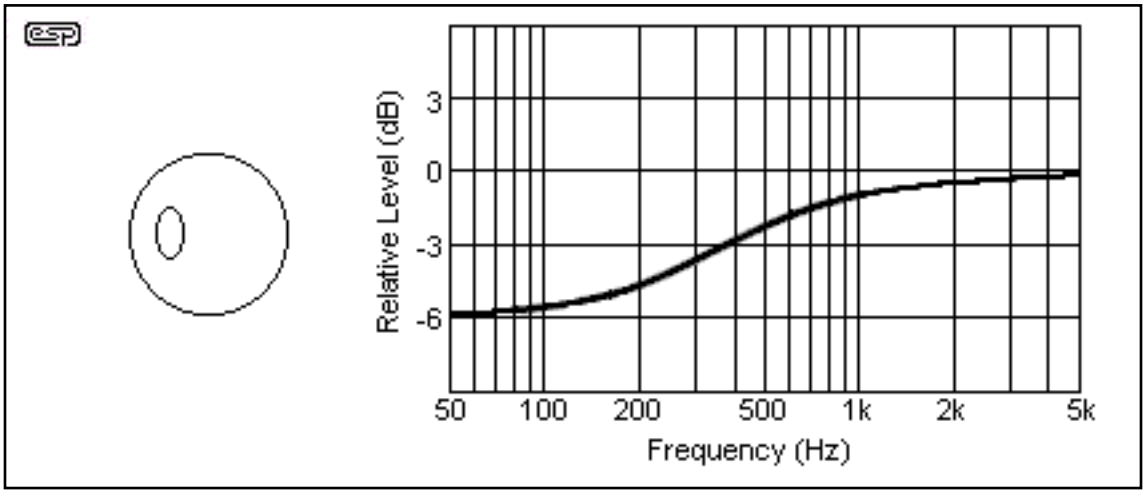


Figure 1 - Response Behaviour - 300 mm Sphere (After Olson)

The increase in output can be seen as the dimensions of the sphere approach the wavelength of the reproduced frequency. Once the wavelength is smaller than the diameter of the sphere, the response flattens out again, having risen by 6dB from the low frequency level. The radiating efficiency of the driver starts to increase earlier than one might imagine - there is a 1dB increase in SPL at the frequency where the baffle is about 0.2 of a wavelength, and the baffle step effect has no influence when the baffle width is greater than 3 wavelengths. At this point, the driver is effectively radiating into a hemisphere.

The absolute worst enclosure is a circular tube (cylinder), with the driver mounted in the exact centre of an end-cap. The response ripples caused by this are extreme, but with most systems this is not an issue, since this would be a very unusual shape for an enclosure. The response in a cylinder shows fluctuations in level that are quite unacceptable, and it is not a shape I would recommend to anyone. A cube is little better, and again the effect is made very much worse if the driver is mounted in the centre. Bacause these shapes are (or should be) avoided altogether, I have not

included any response graphs.

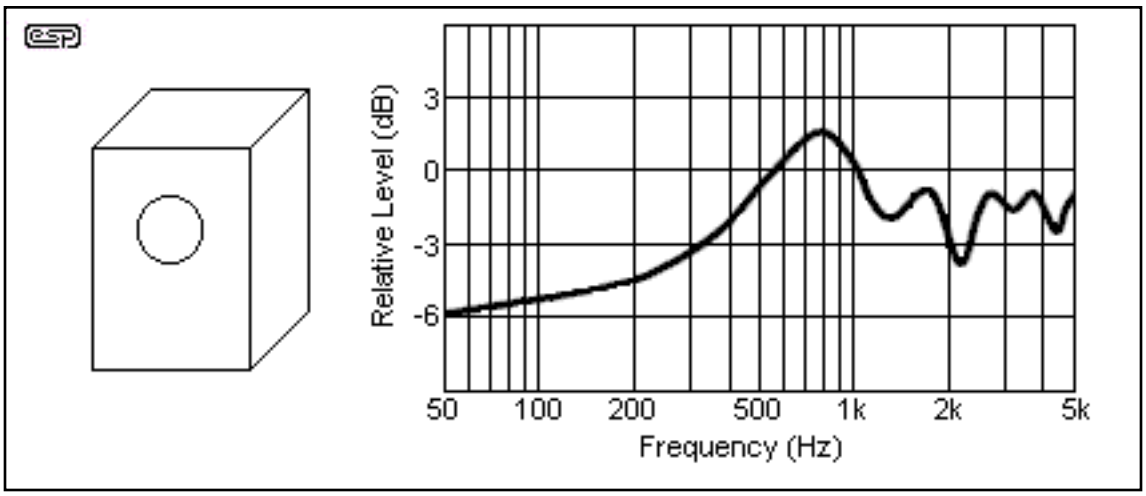


Figure 2 - Response Behaviour - 300mm Wide Rectangular Box (After Olson)

The overall behaviour is greatly improved with any mounting method if you ensure that the centre of the driver is a different distance from each edge of the box. These distances should be "non-harmonic", which is to say that they should be different, but in such a way that no one distance is an even multiple of any other.

When the distances from the driver centre to three sides are equal, the response will be much like that shown in Figure 2. This is better than a cube or cylinder, but much greater improvements can be made, just by changing the position of the driver. In addition, rounded edges will cause less refraction than square or raised edges, further improving the overall response, and particularly at the higher frequencies.

Baffle Step Equalisation

The baffle step is easily compensated for where the speakers are mounted in trees or are in other (more sensible) outdoor environments, but becomes much more difficult as a normal listening room is introduced. Adjacent walls, cabinets and other furnishings will all have an effect, and the outcome is unpredictable at best.

It has been determined that the baffle step may only be a couple of decibels in some circumstances, and sometimes less. In a normal room, it is unlikely that the step will be greater than 3dB, unless the speakers are a considerable distance from the walls.

To this end, I have determined a simple passive network that allows you to adjust the level with a pot, until the response is optimum. The frequency is fixed, and is easily determined based on the width of the baffle. The equaliser circuit should be placed

between the preamp and power amp - it is not suitable for use in the speaker lines. Although circuits exist for use in the speaker lines, I would not recommend them in any situation, since power losses are very high, and they cannot be adjusted easily to suit your listening environment.

Based on an excellent formula developed by John Murphy (True Audio), we can calculate the frequency easily for the box ...

$f_3 = 115 / W_B$ (where W_B is the baffle width in meters)

$f_3 = 380 / W_B$ (where W_B is the baffle width in feet)

Needless to say, I will use metres, and as shown in the graphs above, will use a baffle 300 mm wide. The frequency is therefore ...

$f_3 = 115 / 0.3 = 383$ Hz

Variable Equaliser

A circuit is needed that will provide a typical 3dB decrease in level at the calculated frequency, and the (very simple) schematic is shown in Figure 3.

The value of C1 is determined by the following (1/2 of the pot value is added to the resistor value, since both are in series, and a typical situation will have a pot setting of around halfway - there will always be some error, but it will normally be quite small in practice) ...

$C1 = 1 / (2 * \pi * (R1 + (VR1 / 2) * f)$

$C1 = 1 / 2 * \pi * 15k * 383 = 27nF$

Since the formula will regularly give values that are not obtainable, we may use the closest available value with little error. Certainly the error will be very much smaller than that created by the room acoustics and other influences. The sample circuit is shown in Figure 3, and as shown is valid for a box with a 300mm wide baffle. The only change that you will need is to the value of C1 according to the above formula.

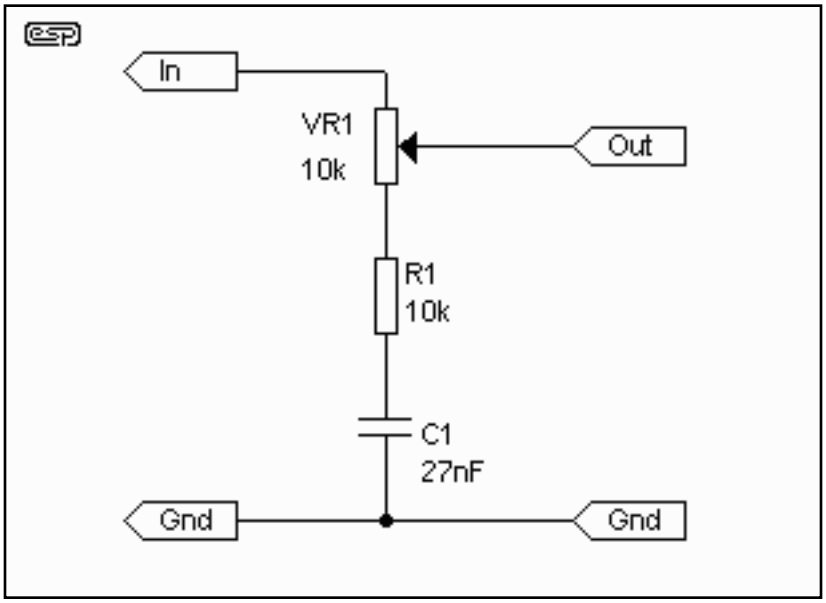


Figure 3 - The Baffle Step Correction Circuit

It is essential that the compensation circuit be driven from a low impedance source, and the load impedance should be reasonably high. There will be little error with loading above 20k, but basically the higher the impedance, the better. Opamp buffers at the input and output may be used if you cannot ensure that the source impedance is 100 ohms or less, and that the load impedance of the following stage is greater than 20k. My recommendation would be to use a buffer stage at the output with an input impedance of about 100k.

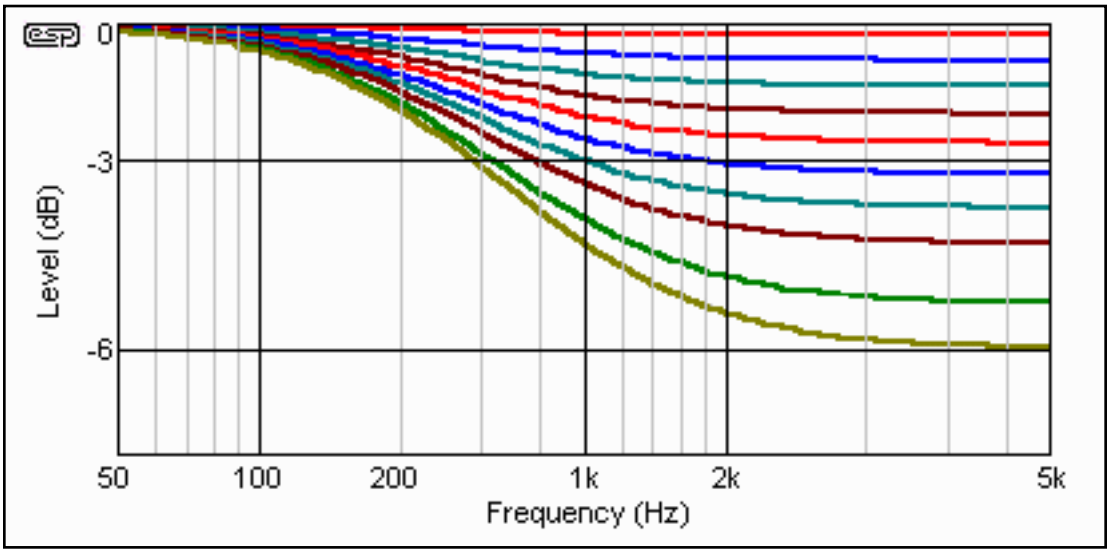


Figure 4 - Typical Frequency Response at 10% Increments

As shown in Figure 4, the pot may be moved from a flat response right through to full 6dB compensation. The graph shows 10% increments of the pot. When the pot wiper is at the top of its travel (based on the schematic), the circuit is inactive, and maximum compensation occurs at at the opposite end of the pot's travel. Nearly all normal

listening room environments will use a setting somewhere in between.

References

1. Harry F Olson - Direct Radiator Loudspeaker Enclosures, JAES Vol. 17, No. 1, 1969
2. John L Murphy - TrueAudio

Note that John Murphy's article was the inspiration for this offering, and there are many other references quoted in his original article
(See http://trueaudio.com/st_diff1.htm)

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 24 Jun 2001

The Audio Pages

Elliott Sound Products **Compliance Scaling**

Compliance Scaling and Other Techniques

(Fitting just about any driver to just about any alignment)

Robert C White / Rod Elliott

05 June 2005

Articles Index

Main Index

Contents

- 1 Introduction
- 2 T-S Parameters and Compliance
- 3 Using The Transforms
- 4 Using the Qb5 Table
- 5 Calculating kp
- 6 Increasing Qt
- 7 Auxiliary Filter Damping
- 8 A small Satellite Surround Speaker
- 9 The B5 Alignments
- 10 References

---

1 Introduction

Since Thiele published his seminal paper on the design of reflex enclosures [1], his original table of 28 'alignments' has been greatly extended, and these are widely available in both speaker design cook books and on the net.

The impression that seems to be given about these alignments is that they are chiseled in stone, and that if one has a particular driver then one is stuck with the f3 and box volume dictated by the alignment. This however is not true. The fact that the driver's suspension compliance is very insensitive to system parameters, i.e. ...

α, h, f3

means that very many alignments exist that are close to but not quite the 'classical' ones usually published and, as White states, 'virtually any driver can be fitted into an alignment', [2, p.892]. Thiele states that from certain points of view the suspension compliance is *'unimportant'* (his italics), [1, p.188].

The miracle of fitting just about any driver to just about any alignment is performed by a procedure called 'compliance scaling', and was, as far as the author is aware first published by Keele, in a 1974 AES paper, [3].

In addition to compliance scaling two other techniques are outlined, electronically increasing Qt and modifying auxiliary filter damping factor.

Programs available as downloads such as WinISD do a good job of system modeling, the difficulty is that once one strays from the standard alignments that these usually have as defaults, one is in the dark as to what to change and by how much, in order to achieve the desired response.

What follows is a simple method that yields some figures to enter into such programs to give a desired result, and allows you to design a box with a particular driver to achieve a target f3, and or a target box volume, also included is a method of increasing Qt.

But first a word about compliance in relation to the Thiele - Small parameters.

---

2 T-S Parameters and Compliance

The T-S parameters have become the standard way of specifying the performance parameters of low frequency moving coil drivers. This is because they clump the fundamental physical properties together in such a way as to ease the problem of calculation and specification.

they are related to the physical properties by the following relationships ...

Vas = *po*C² CmsSd²
Qt = √(Mms / Cms) (1 / Rmt)
Fs = 1 / (2π √(Mms Cms)

Where **Cms** = suspension compliance (m/N), **Mms** = moving mass (kg), **Sd** = diaphragm area (m²), ***po*** = density of air (1.2kg / m³), and **c** = speed of sound (343m/s).

As can be seen compliance plays an important role in all of the above, and also in the system parameters, **α, h, f3**

As stated by Keele, [4 p.254], typical batches of drivers have a 10-20% variation in T-S parameters, and this is largely due to differences in suspension compliance, but luckily for the speaker designer this has little effect upon the frequency response in a given box because of the insensitivity of this to alignment.

The actual scaling procedure entails applying a set of transforms to a 'seed' alignment to give use a new set of system parameters, and this uses a 'normalised compliance' constant, **Cn** defined as ...

Cms / Cms'

Where **Cms** is the driver compliance, and **Cms'** is the driver compliance that would be needed for the driver to be 'correct' for the alignment. **Cn** then gives us the transform set ...

h = h' √Cn
Qt = Qt' / √Cn
α = α'Cn
f3 / fs = f3' / fs' √Cn
fa = fa' √Cn

Applying these transforms to a standard B4 alignment gives the following results, (Data copied from, [2, p.892]), for the Cn values of 0.25, 1.0, 4.0 ...

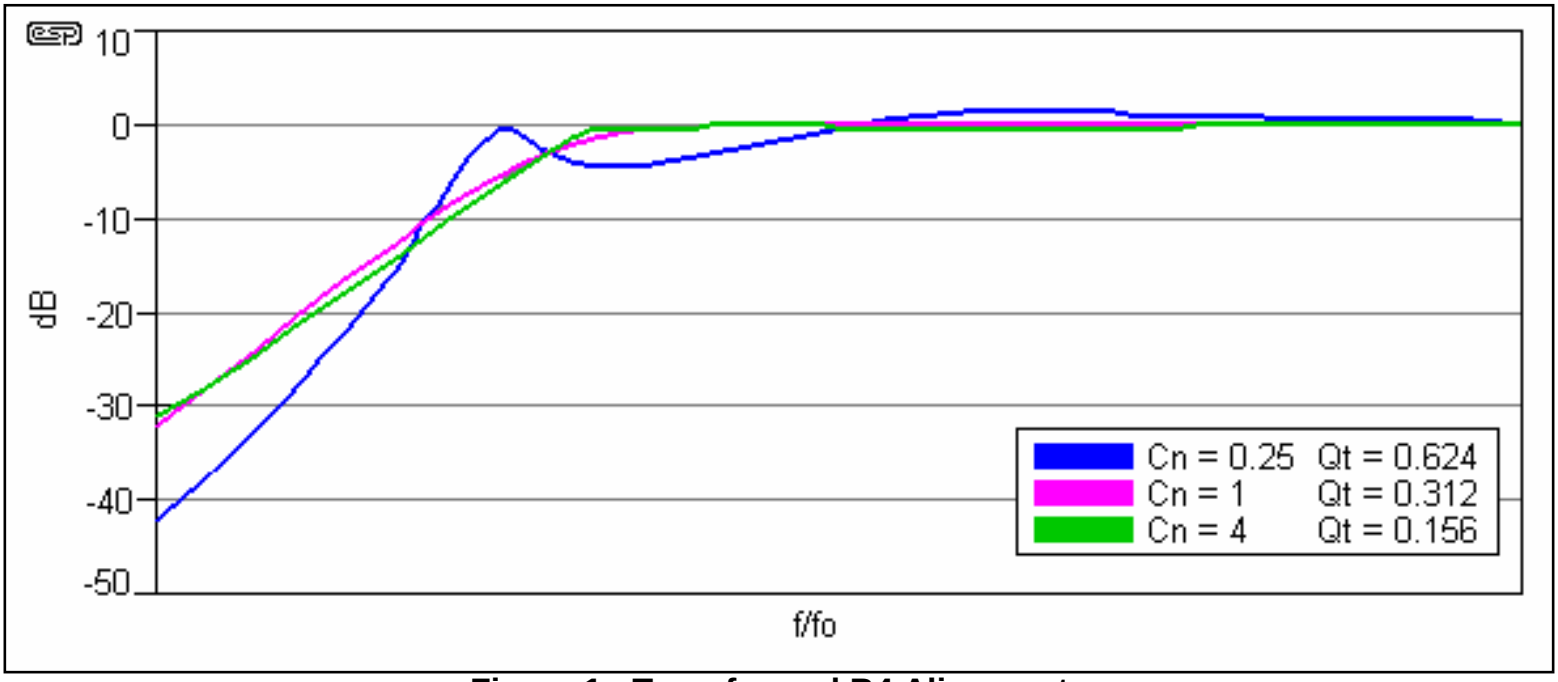


Figure 1 - Transformed B4 Alignments

In the above graph the blue plot is for a Cn = 0.25, this corresponds to a Qt of 0.624, the magenta line is for Cn = 1, Qt = 0.312, the exact B4 alignment, and the green is for a Cn of 4, Qt being 0.156 It can be seen that the greatest deviation is for low Cn values. Figure 2 is for a filter assisted B6 alignment ...

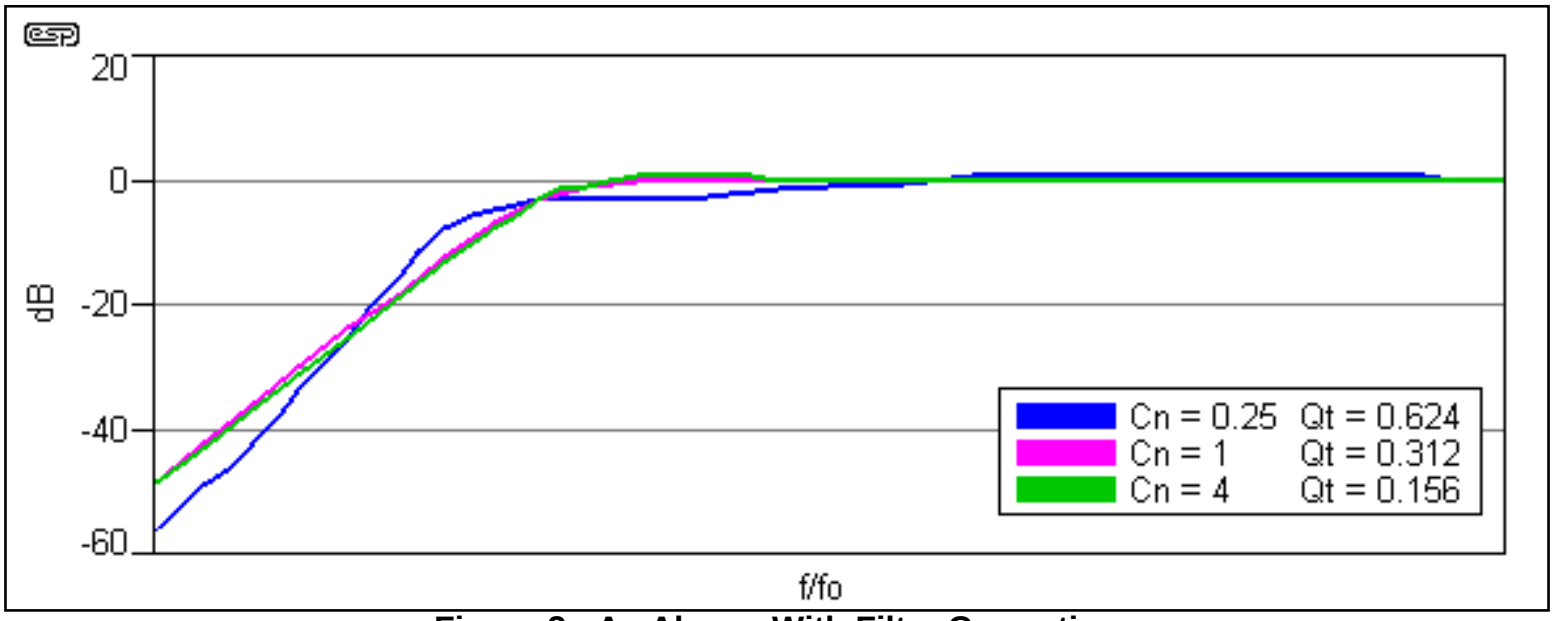


Figure 2 - As Above, With Filter Correction

The Cn values are the same for this plot, and the low Cn plot is somewhat smoother than for the non filter assisted case. Overall the Qt can vary over a wide range without any significant change in overall frequency response and at the low frequencies involved this slight aberration is completely swamped by room effects.

3 Using The Transforms

We usually have a specific f3 or box volume in mind when we design a speaker. The techniques that follow allow you to obtain a specific f3 or box volume. Also discussed is the method of alignment adjustment by means of increasing power amplifier source impedance by means of current feedback around the power amplifier, (see Rods article). To make it easier the tables that follow contain two constants, one enables a particular f3 to be obtained, another a particular box size. From the transforms we can write f3 as ...

f3 = (fs / f3' Qt') / (Qt fs')

This separates into two constants, one characteristic of the enclosure, the other of the driver, these are ...

kbf = (f3' qt') / fs' and **kdf = fs / Qt**

Likewise the expression for Vb ...

Vb = (Vas Qt² Vb') / (Vas Qt'²)
kbv = Vb' / (Vas x Qt'²) and **kdv = Vas x Qt²**

Table #1 has the values of kbv and kbf for the Qb5 alignments (alignment table in 'Satellites and Subs' article).

Qt	kbv	kbf	Qt	kbv	kbf	Qt	kbv	kbf
0.324	4.789	0.303	0.445	9.182	0.445	0.514	7.101	0.529
0.318	4.702	0.318	0.425	7.372	0.490	0.517	7.099	0.533
0.311	4.618	0.329	0.415	6.783	0.505	0.520	7.112	0.536
0.303	4.546	0.338	0.405	6.351	0.528	0.523	7.127	0.540
0.295	4.464	0.345	0.394	6.009	0.541	0.526	7.157	0.543
0.287	4.378	0.351	0.384	5.685	0.554	0.530	7.177	0.547
0.279	4.295	0.356	0.373	5.425	0.565	0.534	7.201	0.550
0.271	4.217	0.361	0.362	5.202	0.575	0.538	7.258	0.554
0.263	4.147	0.365	0.352	4.982	0.585	0.543	7.294	0.558
0.255	4.088	0.368	0.342	4.792	0.594	0.549	7.340	0.562
0.247	4.041	0.370	0.331	4.659	0.599	0.555	7.412	0.566
0.240	3.975	0.373	0.322	4.498	0.608	0.562	7.503	0.570

0.233	3.921	0.376	0.312	4.392	0.612	0.569	7.626	0.574
0.226	3.880	0.378	0.303	4.280	0.618	0.577	7.781	0.577
0.219	3.852	0.379	0.294	4.190	0.623	0.587	7.973	0.581
0.213	3.804	0.381	0.286	4.092	0.628	0.598	8.249	0.584
0.207	3.767	0.383	0.278	4.013	0.632	0.610	8.614	0.587
0.201	3.742	0.384	0.270	3.953	0.635	0.625	9.143	0.589
0.196	3.692	0.386	0.262	3.910	0.637	0.643	9.913	0.592
0.190	3.692	0.386	0.255	3.851	0.640	0.664	11.228	0.594
0.185	3.665	0.387	0.249	3.778	0.645	0.691	13.962	0.595

Table 1 - Qb5 Alignments

4 Using the Qb5 Table

In recent years the tower type loudspeaker system has become popular, these have the major advantage of a high SAF (Spouse Acceptance Factor). The larger of these generally has two 200mm drivers in either a two, two and a half or three way configuration. Two 200mm drivers have sufficient cone area and excursion to provide useful bass down to 30Hz, especially if used in a filter assisted alignment. A typical high quality 200mm driver is the Peerless Model CSX 217H, from the data in WinISD, This has ...

kdf = 27.4 / 0.281 = fs / Qt = 97.51

This needs ...

kbf = 30 / 97.51 = 0.308

From the QB 5 Class I alignment table the suitable alignment is No. 1 with a **kbf** = 0.303, so ...

√**Cn = 0.324 / 0.281 = 1.153**, therefore **Cn = 1.329**
a = 1.989 x 1.329 = 2.64
h = 0.995 x 1.153 = 1.322
fa = 0.944 x 1.153 = 1.255

multiplying or dividing the a, h and fa data in the QB5 table by **Cn** or √**Cn** as appropriate, gives the transformed values ...

Vb = (2 x 81.58) / 2.64 = 61.8 litres
Fb = 27.4 x 1.322 = 36.2 Hz

Fa = 27.4 x 1.255 = 34.4 Hz

Putting these figures into WinISD gives the result ...

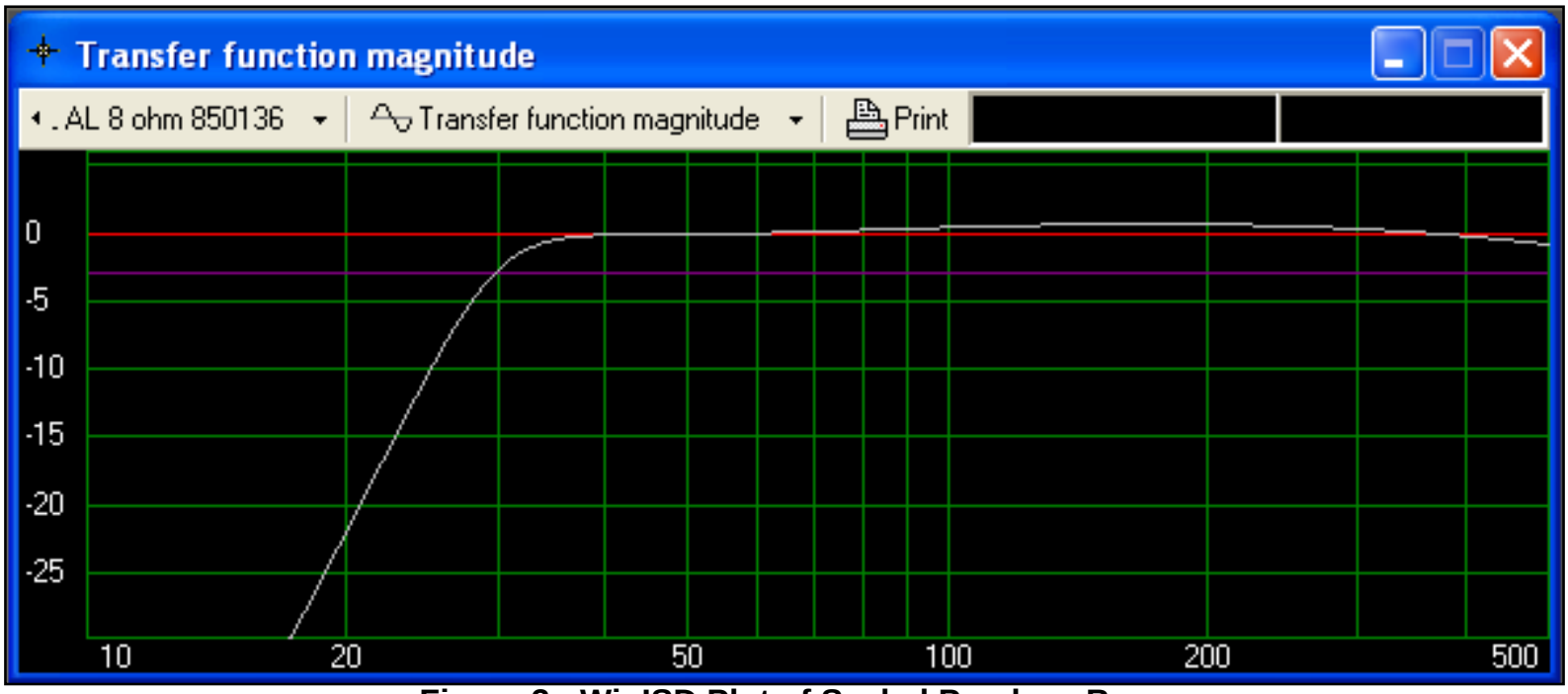


Figure 3 - WinISD Plot of Scaled Peerless Box

By using the compliance scaling and filter assistance this is a much better result than would normally be expected. The filter is a second order high-pass, having a -3dB frequency of 29.8Hz and a Q of 2.141

Should anyone simply select the same driver and run WinISD, the optimum result is actually far from optimal, having a -3dB frequency of almost 47Hz. While the box is smaller, it will be sadly lacking in the bottom end.

---

5 Calculating kp

Small's power handling parameter 'kp', is a very useful tool in determining the excursion limited power handling of a particular system. I have yet to find any reference to it on the Net except for my QB5 alignment article, I suspect this is because kp is difficult and tedious to calculate, and programs like WinISD provide a plot of maximum SPL. However, I still like to check by calculating kp, and the simplest way I have found is as follows ...

- Enter the data for the exact QB5 alignment for your driver
- Find pk. Excursion for this
- Enter the data for your transformed alignment

- Find pk. Excursion for this

Divide the peak excursion for your transformed alignment by the 'correct' one, and substitute in Small's expression, [5, P.439] ...

$kp1 = 0.425 / ((f3 / fs)^4\ j(xmax)^2)$

Where kp' is the correct 'seed', kp

$J(xmax)' = (0.424 / ((f3 / fs)^4 / kp)^{0.5}$
$J(xmax) = ((pkx / pkx')\ j(max))'$
$Kp' = 0.425 / ((f3' / fs')^4\ j(xmax)'^2)$

In our first example the low pk. Excursion is = 1.69mm and the high pk = 0.873mm, an average of 1.28mm. The 'correct' alignment has low pk = 0.532 and high pk = 0.977, averaging 0.755

The correct alignment has a kp of 9.349, giving j(max.)' of 0.131, j(max) is then = 0.222 giving Kp = 5.98

Giving a conservatively rated output of around 105db Peak for a pair before the linear excursion limit is reached, the SLS 213 driver will give more output before excursion limiting, but at the expense of a box twice as large. It should also be noted that since the class I alignments have an average of 6db of boost at around the f3, they need around four times the power that the nominal efficiency would indicate, this is the price we pay for a small box.

---

6 Increasing Qt

We can increase Qt by increasing Qe, and this can be done by increasing the driving amplifiers output impedance by means of current feedback,[6], (Rods article).

If we write Qe as [7] ...

$Qe = Rvc / Lces\ ws$

We can increase Qe by adding a source resistance to Rvc

$Qe' = (Rvc + Ro) / Lces\ ws$

If the Qe we need is = Qe', then The required Qe' is given by ...

Qe' = 1 / (1 / Qt' - 1 / Qm)

Where Qt'= the required Qt., The required source resistance is then ...

Kl = Rvc/Qe

Ro = Qe'Kl - Rvc

[ ] The following diagram shows the essentials of modifying the amplifier's output impedance, giving a circuit that is relatively easily scaled for any loaded gain and a defined output impedance. The calculations for this are fairly straightforward, but only if you can accept an apparently random loaded gain. If you want to specify the loaded gain, the calculations become extremely tedious. While a far better mathematician than the editor may be able to derive a suitable equation, I was unable to do so, and as a result have written a small Visual Basic application to perform the maths needed to arrive at a satisfactory set of component values. This is still under development to get it right - the full details will be published in a second part to the current page.

Although I did try this with a spreadsheet (and that's where the formulae I eventually used came from), it is a reiterative process. Those wishing to experiment are encouraged to do so. All the necessary formulae and the finalised calculator will be available with Part 2 of this article.

Essentially, the calculations involve solving for an unbalanced Wheatstone bridge network, with specific desired end results. Of course you can always cheat and use a series resistor of the appropriate value, but this will have to dissipate considerable power with any amplifier capable of reasonable output power.

In essence, the determination of the correct values is not easy. Because we generally need to specify the output impedance, as well as the normal loaded gain (so the amp is in line with others in the system), we end up with too many unknown variables - the circuit may look simple, but calculations for it are not.

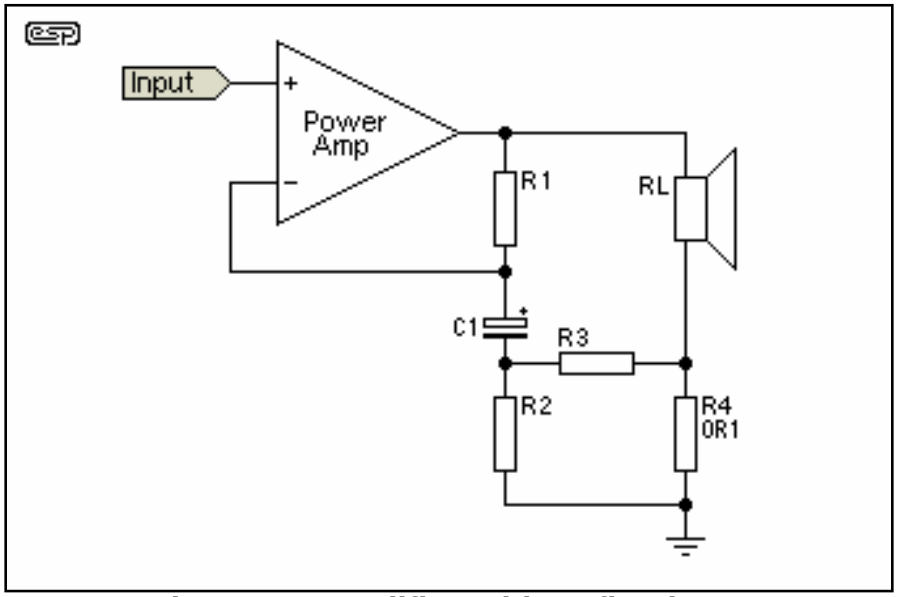
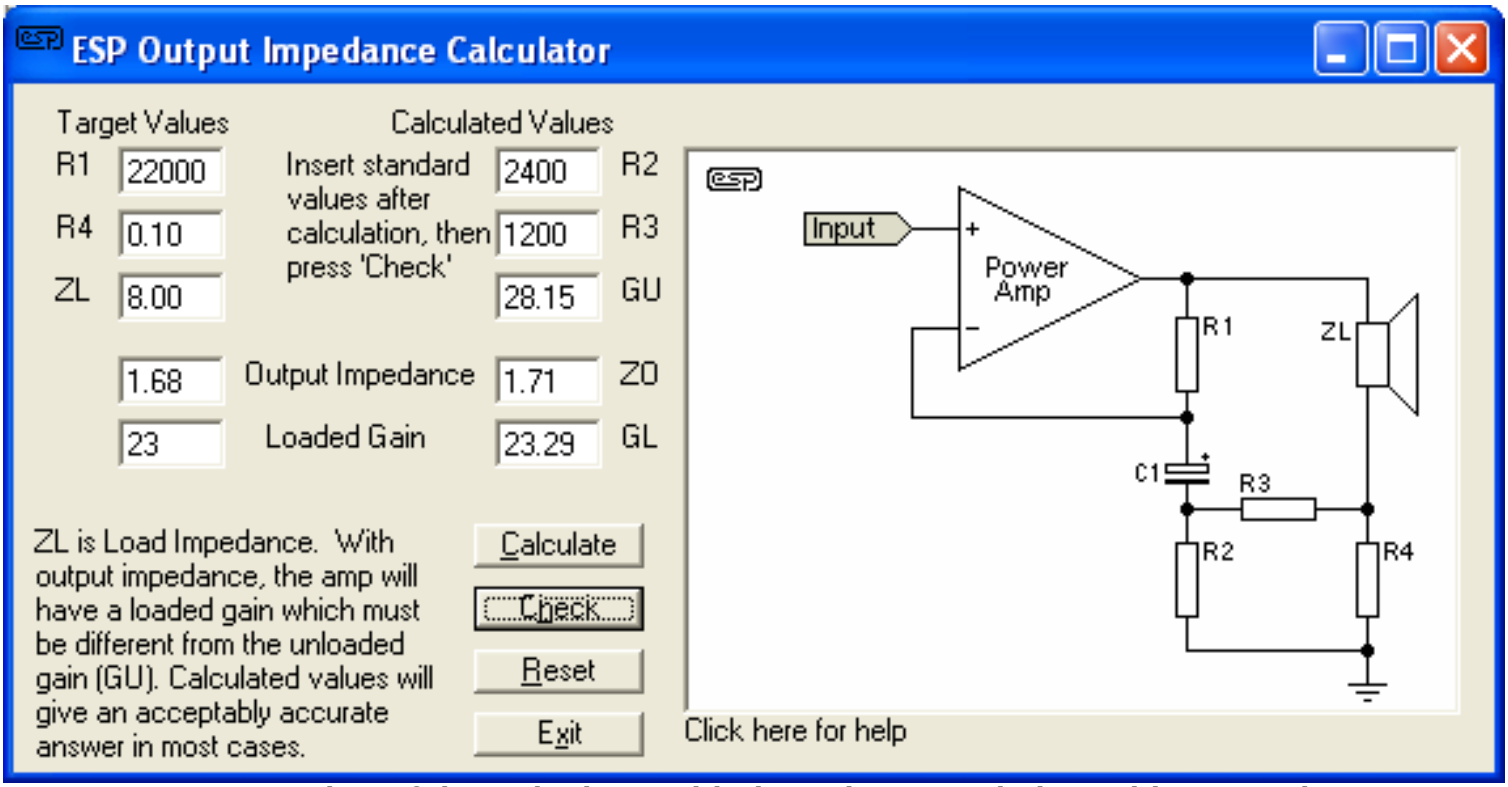


Figure 4 - Amplifier With Defined Z_{out}

The circuit above is almost identical to that shown in the article / project Variable Amplifier Impedance. By varying R2, R3 and R4, it is possible to achieve a wide range of impedances that will be usable in this application. The circuit can be made variable, however this is not normally useful except for ongoing design and experimentation.



Screen Shot of the Calculator with the Values Needed For this Example

The above screen shot shows the values determined for the example presented here using the first draft of the calculator program. Standard values were used for R2 and R3, and the results checked as shown. The difference between the desired 1.68 ohms and the

calculated 1.71 ohms is minimal and will not have a significant effect on the result.

Note that the loudspeaker driver's nominal impedance is used to determine the loaded gain, and this is useful because nearly all ESP amps are designed with a gain of 23 (approx. 27dB). The calculator program is available as a download (see below). Sorry, but because it is written in Visual Basic, it is only suitable for Windoze machines (or Macs running Wine emulation or similar - perhaps - this has not been tested).

As shown above, the values apply for the following example. R1 is 22k as used in most ESP amplifier designs, and the feedback resistor (R4) is 0.1 ohm. Using 2.4k (2 x 1.2k in series) for R2 and 1.2k for R3, the output impedance is 1.71 Ohms, and the loaded gain is 23.29 - almost exactly the same as other amps as may be used in the system. [/]

This scheme is especially useful for increasing the Qt of low Qt high efficiency drivers, such as the JBL K140. This driver is still much available on the internet as a re-cone, and is especially designed for bass guitar applications. If we put it into a filter assisted alignment the result is an increase power handling. Increasing Qt has the advantage that the drivers high efficiency is preserved whilst achieving a low enough f3 in a small box. We want f3 = 40Hz for the usual bass guitar tuning.

Looking at the QB5 class I alignments the required f3 / fs of 1.333 is achieved by the alignment No. 7, this needs a Qt of 0.271, and a box of 92 litres ...

Qe' = 1 / (1 / 0.271 - 1 / 5) = 0.287
Ro = 1.68

The above values are perfect, with only the smallest discrepancy which is of no consequence in the final design. WinISD gives the following ..

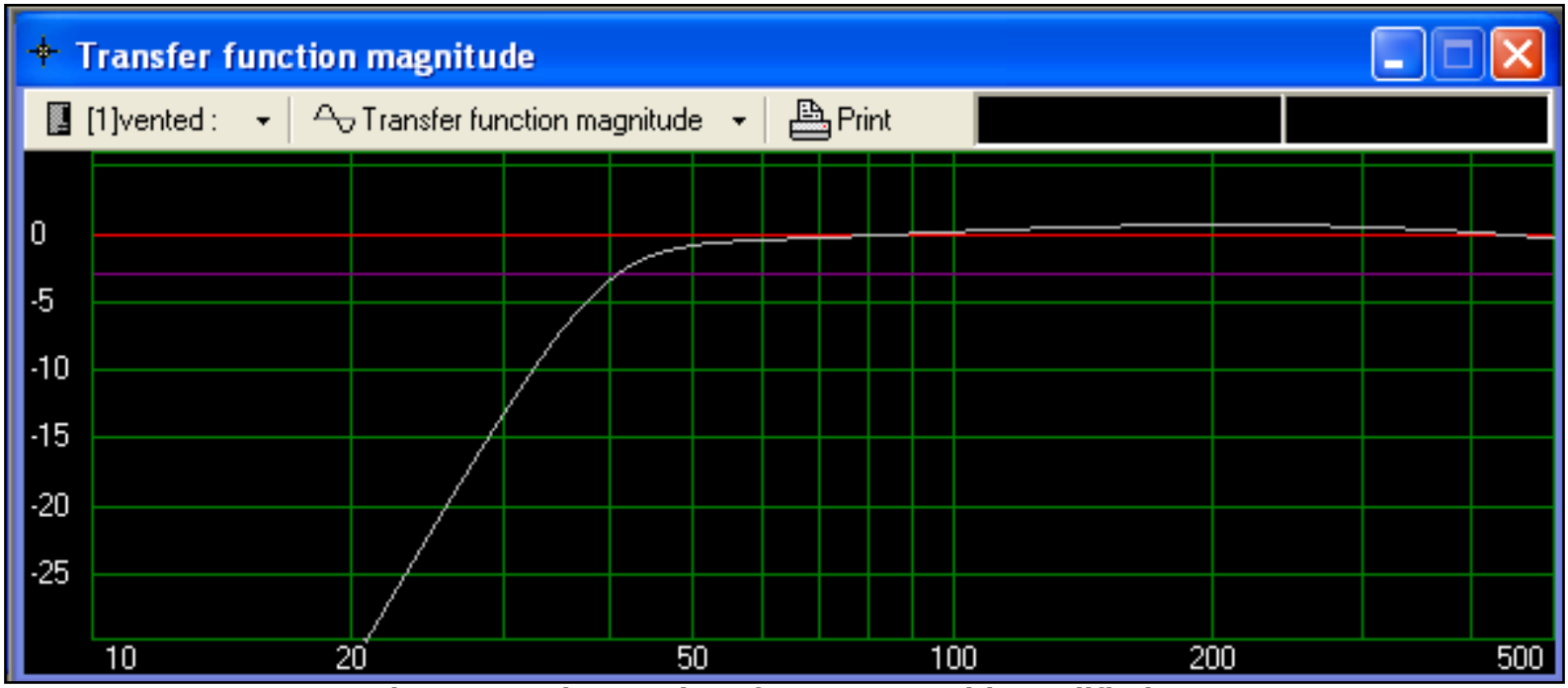


Figure 5 - WinISD Plot of JBL K140 With Modified Qt

This is achieved by using the amplifier set for an output impedance of 1.7 Ohms, and with a second order 40Hz high-pass filter having a Q of 1.658. It is worth noting that using any one or two of the techniques described will not work - the final result is derived only by the combination of compliance scaling, elevated source impedance and filter assistance applied as a complete system solution.

---

7 Auxiliary Filter Damping

Thiele points out that all other things being equal, doubling Qt results in a +6db peak at the resonant frequency, [1], i.e.

20 Log(Qt / Qt')

From this we can fit a driver to a particular filter assisted alignment by modifying the fa damping by the amount

dbfa' = dbfa - (20 Log(Qt / Qt'))

---

8 A small Satellite Surround Speaker

With the advent of five and now seven speaker surround systems, there is an increasing need for small speaker systems with good power handling, and the all important SAF

index improves with less intrusive enclosures. A typical high quality 150mm. Bass/mid. Driver is the Vifa P17WJ. In this case we fit it to a box that will give an f3 of around 80Hz, making it suitable for use as a satellite, and use a QB5 class II alignment, this optimizes excursion limited power handling.

If we use the QBQ class II alignment No. 14, we have

Vb = 12.6 litres, fb = 68.6 Hz, Fa = 78.1 Hz, 1/Qfa = 1.9

Giving the WinISD plot ...

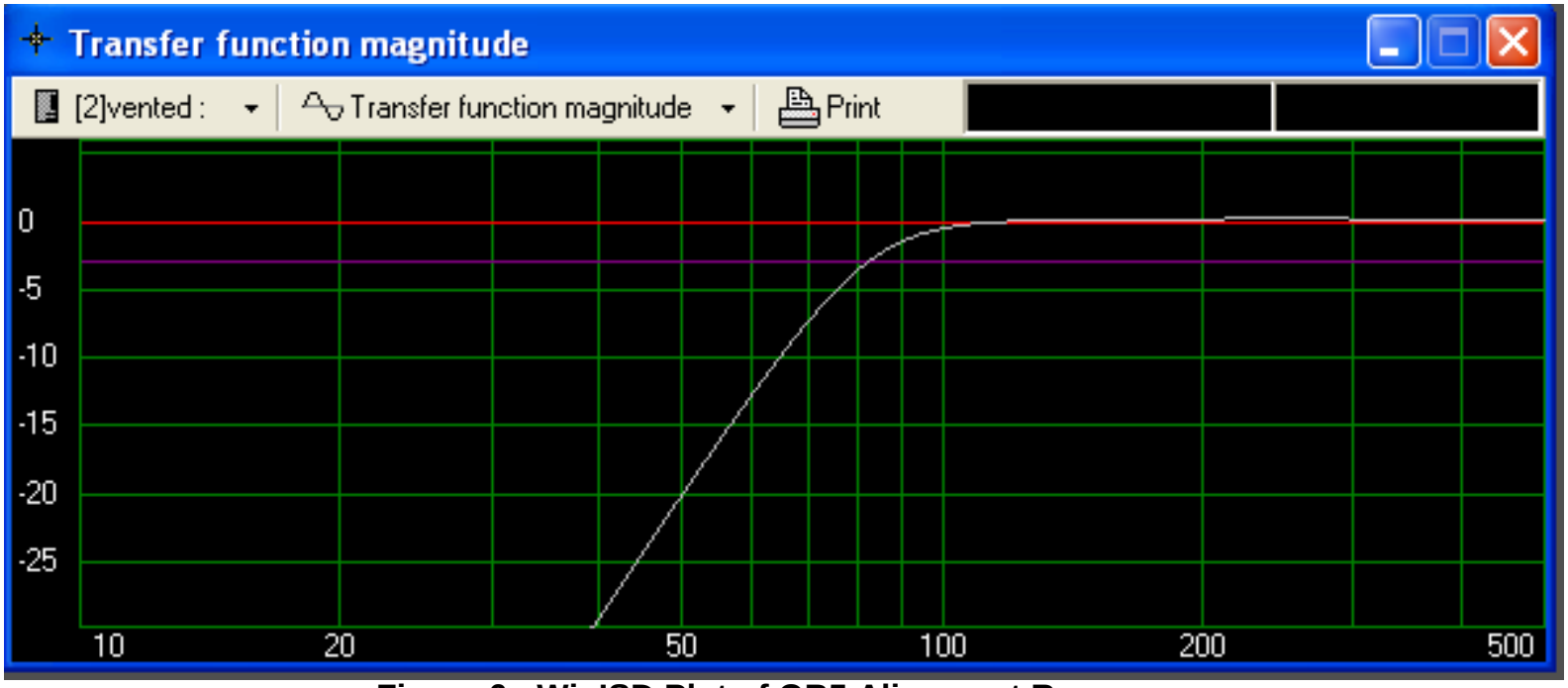


Figure 6 - WinISD Plot of QB5 Alignment Response

---

9 The B5 Alignments

In some cases we only want to use a simple passive first order filter, in others we only want to use two passive first order sections, in which case we have a fixed auxiliary filter damping factor of two. This is characteristic of the auxiliary filters in the QB5 class III alignments. In the case of being able to use only a single first order section, the B5 alignments [8], are useful, these are reproduced in table #2.

Table 2 - B5 Alignments

A driver that is popular for computer speakers is the Tangband, W3-926S. Using a B5 alignment with an F3 = 100Hz. Gives a box of 4.7 litres, tuned to 105Hz. This achieves its

rated x_{peak} with *one Watt* of input. If you put this driver into a sealed box the maximum output at 100Hz is 77.9dB, with the B5 box this is raised to 89.9db. Using a QB5 class III box gives a rather large 7.5litres.

It also should be noted that putting a capacitor between 150 and 220 uF. in series with the driver only changes the frequency response by around 2db. As shown in the Matlab plot of Figure 7 the frequency response with 220uF is not as smooth as with an isolated filter, but is within acceptable limits.

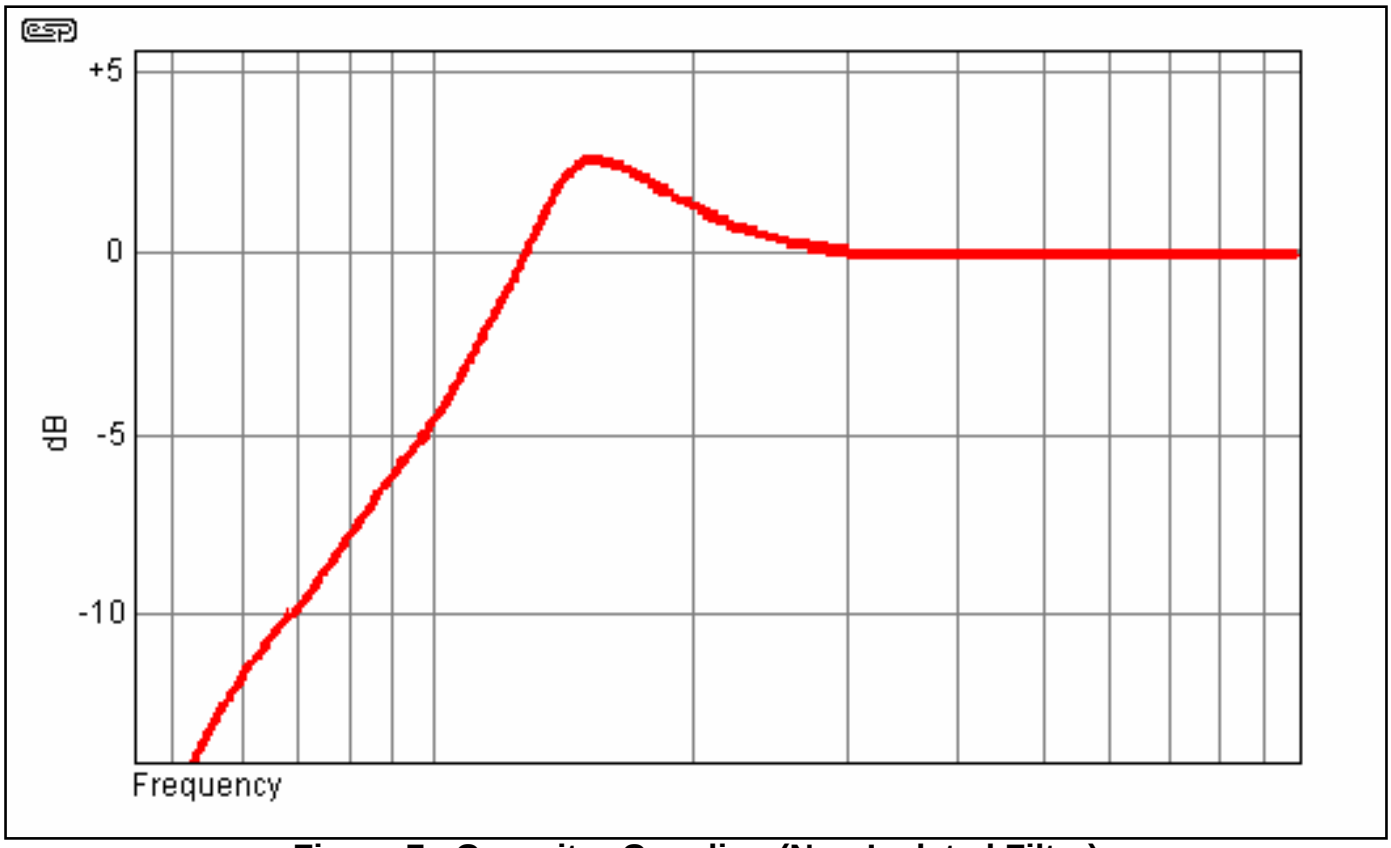


Figure 7 - Capacitor Coupling (Non-Isolated Filter)

In some instances it is convenient to provide one isolated and one non isolated filter, i.e. by means of an input coupling capacitor on the amplifier, and a capacitor in series with the driver. Using a first order isolated filter of twice the non isolated filters f3 gives a plot of Figure 8.

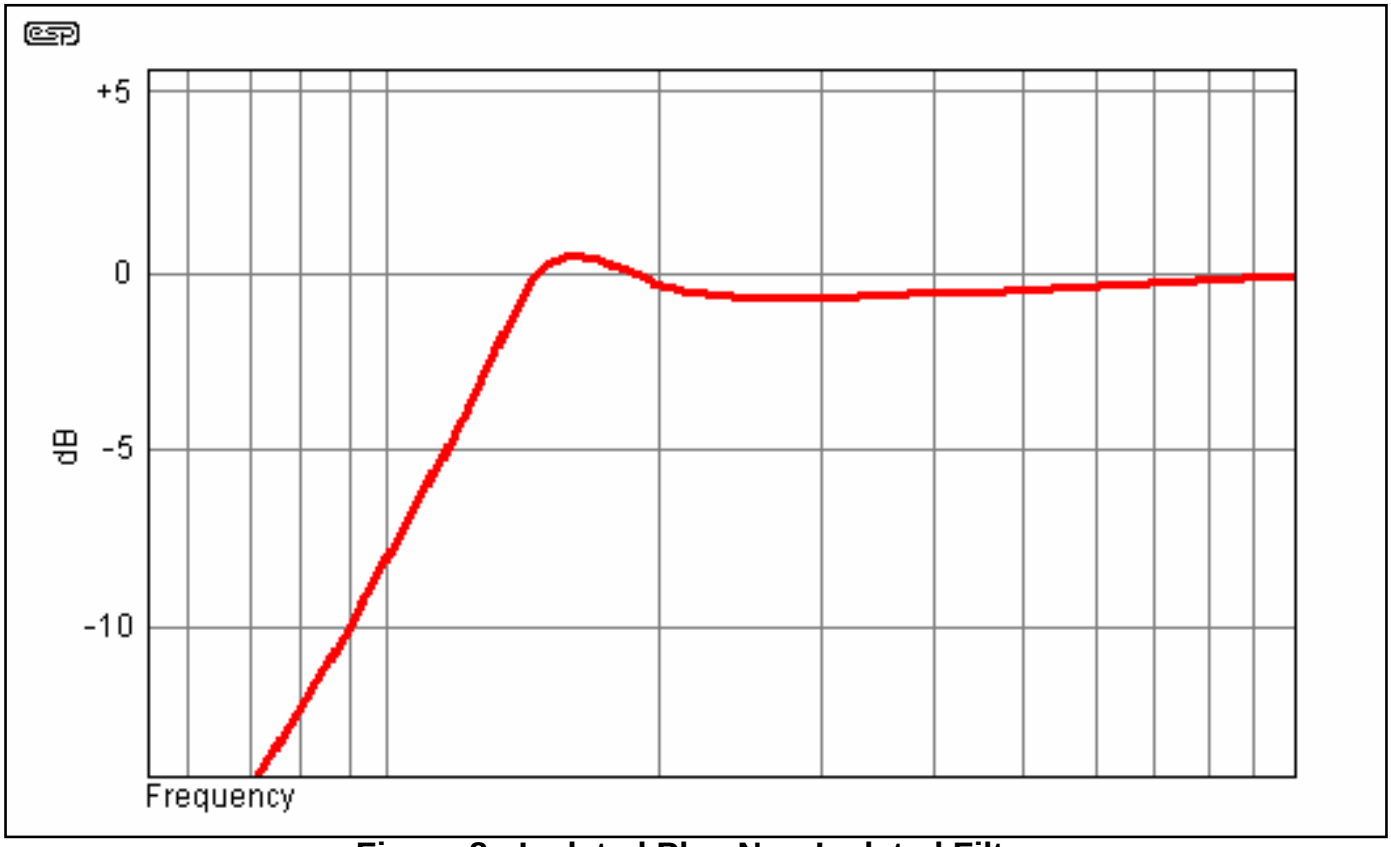


Figure 8 - Isolated Plus Non-Isolated Filter

As illustrated in this article it is possible to fit just about any driver to just about any alignment using the three techniques outlined, or indeed a combination of them.

All efforts have been made to make the calculations of the simple plug in the numbers variety, I apologise where this is not possible. Anybody is at liberty to turn this article into a spread sheet, or some other easy to use form, do not ask me for this however - programming computers is a thing that I dislike, and avoid at all costs.

---

10 References

1	A.N. Thiele, "Loudspeakers in vented boxes", (reprint) proc. IREE (Australia), Vol.22, (Aug.,1961)
2	J.N. White, "Loudspeaker athletics", AES Journal, Vol. 27, No. 11, (November, 1979)
3	D.B. Keele Jnr, "A new set of sixth-order vented box loudspeaker alignments", AES Journal, Vol. 23, (June, 1975)
4	D.B. Keele Jnr, "Sensitivity of Thiele's vented loudspeaker enclosure alignments to parameter variation", AES Journal, Vol. 21 (May, 1973)
5	R.H. Small, "Vented box loudspeaker systems Part II: Large signal analysis", AES Journal, Vol. 21, No. 6 (July/August, 1973)
6	R. Elliott, "Effects of source impedance on loudspeaker drivers", Articles, www.sound.com.au

7	J.E. Benson, "An introduction to the design of filtered loudspeaker systems", AWA Technical review, Vol.15, No.1 (1973)
8	R.A.R. Bywater & H.J. Wiebell, "Filter assisted loudspeaker systems with enclosure losses", "Loudspeakers, an Anthology, Volume 2" AES New York, 1984, pp.310-321

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Robert C White and Rod Elliott, and is Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The authors grant the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Robert C White and Rod Elliott.

Page created and copyright © 04 Jun 2005

Qt	Vas/vb	Fb/fs	Fa/fs	F3/fs	kbf	kbv
1.320	0.0438	0.695	0.431	0.651	0.859	13.10
1.230	0.0561	0.709	0.455	0.66	0.812	11.78
1.120	0.0711	0.725	0.485	0.671	0.752	11.21
1.010	0.0914	0.745	0.526	0.686	0.693	10.725
0.881	0.126	0.774	0.588	0.712	0.627	10.225
0.791	0.163	0.801	0.645	0.736	0.582	9.805
0.727	0.200	0.824	0.694	0.76	0.553	9.460
0.666	0.251	0.852	0.746	0.79	0.526	8.920
0.633	0.289	0.870	0.781	0.811	0.513	8.636
0.564	0.406	0.917	0.862	0.871	0.491	7.743
0.529	0.499	0.948	0.917	0.915	0.484	7.161
0.508	0.567	0.967	0.943	0.944	0.480	6.834
0.497	0.614	0.979	0.962	0.964	0.479	6.594
0.489	0.645	0.987	0.98	0.977	0.478	6.484
0.478	0.701	1.00	1.00	1.00	0.478	6.243
0.435	0.97	1.05	1.088	1.10	0.479	5.448
0.392	1.39	1.10	1.20	1.25	0.490	4.682
0.346	2.03	1.13	1.348	1.48	0.512	4.115
0.328	2.36	1.13	1.42	1.60	0.525	3.934

0.317	2.55	1.13	1.466	1.68	0.533	3.902
0.309	2.72	1.12	1.506	1.76	0.544	3.850
0.298	2.94	1.12	1.56	1.86	0.554	3.830

The Audio Pages

Elliott Sound Products **Subtractive Crossover Networks**

Copyright © 2005 - Rod Elliott (ESP)
Page Created 20 September 2005

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Basic Subtractive Crossover
  - 2.1 First Order Filters
  - 2.2 Second Order Filters
  - 2.3 Fourth Order Filters
- 3.0 L-R Filter Comparison
- 4.0 Conclusion
- References

1.0 - Introduction

A class of electronic crossover is variously described as a 'derived' or 'subtractive' filter is hailed by some users as the ideal. They have (apparently) perfect transient response, in that the summed output is not only flat, but a squarewave is also passed intact. This implies that they are the Holy Grail of electronic crossover networks. Almost by definition, no other crossover network should even be considered.

So, are they any good? Why aren't they used everywhere?

These questions are best answered by a full examination of the network, so that all the facts are available.

2.0 - Basic Subtractive Crossover

The general idea of the subtractive crossover is quite simple. If we have a filter, and subtract the filtered signal from the input, the result is a filter with the opposite effect (i.e. a low pass filter is 'derived' from a high pass filter and vice versa). Because of the subtraction process, the result must be perfectly in phase, and the sum must (by definition) be flat response.

There have been many variations on the general theme, some of which are claimed to provide better performance than others. Subtractive filters have been discussed in Elektor magazine [1] and some I have seen are quite complex. While the added complexities may suit a particular arrangement of specific loudspeakers, they generally don't add anything that changes the overall performance.

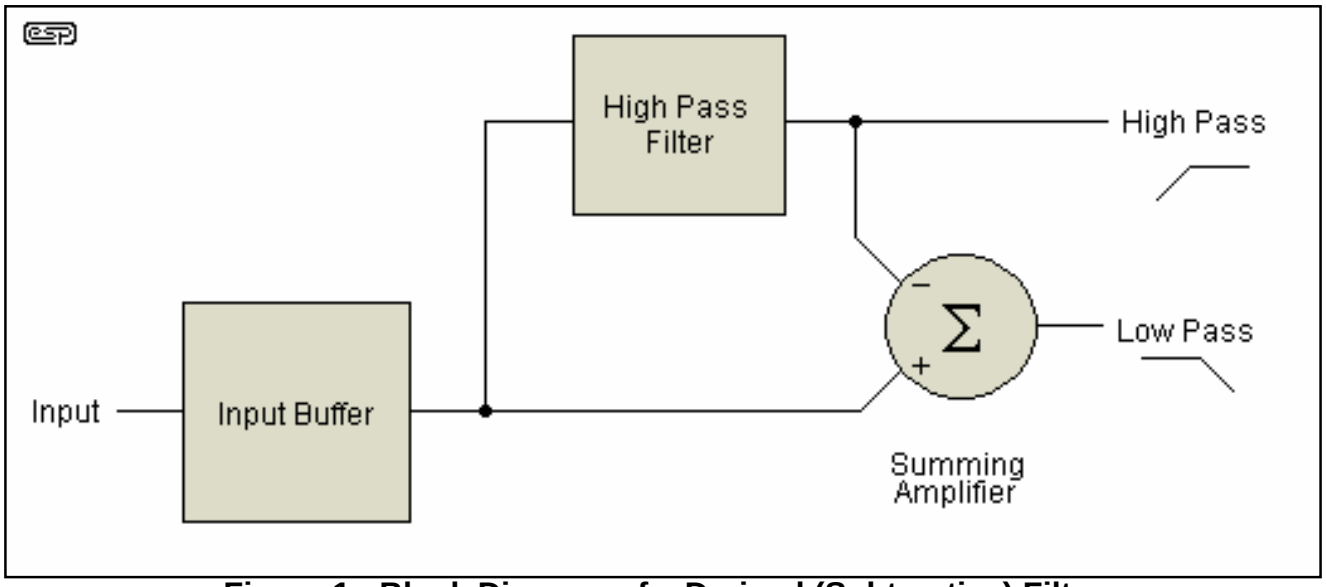


Figure 1 - Block Diagram of a Derived (Subtractive) Filter

While all circuits shown in this article are configured as shown in Figure 1, the filter itself can be a low pass section. No other changes are needed, but this connection may give rise to performance limitations that at the very least must be classified as undesirable (see below for more information).

In the circuit diagrams below, all buffers are unity gain, and all circuits are driven from a low impedance (voltage) source. This is a requirement for all filters, so the input buffer is not shown for clarity. The voltage source shown is an ideal voltage source - zero ohms output impedance.

Likewise, for clarity, the power supplies are not shown. All the results below are from the SIMetrix simulator, and while somewhat idealised, are representative of reality with any reasonable opamp in a real world circuit - especially within the audio band.

Within this article and for the simulations used to get the graphs shown, the same values were used for filter tuning, regardless of the filter order. While this does change the crossover frequency, as you will see it is actually of little consequence.

2.1 - First Order Filters

While there is little point looking at a first order (6dB/octave) network, it *is* the simplest to examine, and this will make it easier to follow the more complex filters. A first order crossover is already 'phase perfect', so making a subtractive version should give an identical result.

As shown below, this is the case. The only advantage of the subtractive method is that only one reactive element (the capacitor) is used, and this is highly debatable as an 'advantage'. This is especially true with the increase of overall complexity of the circuit.

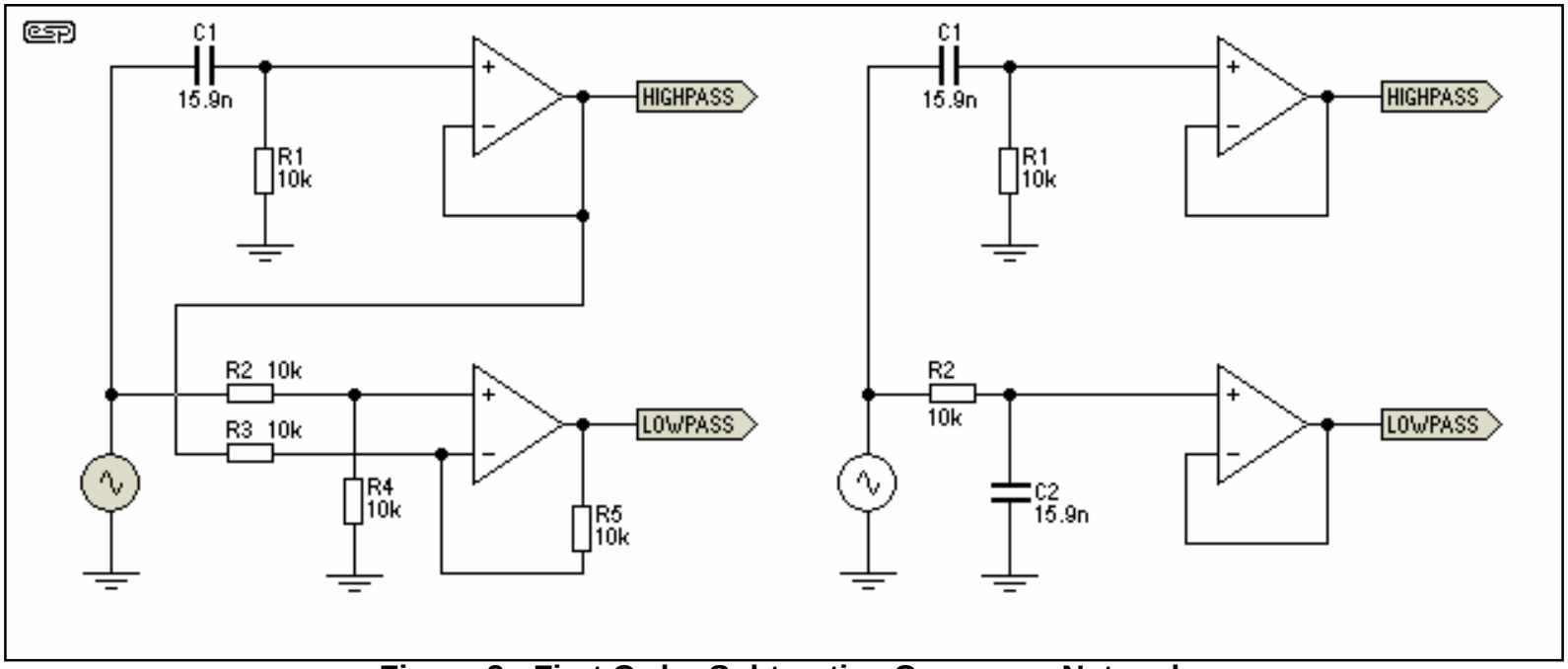


Figure 2 - First Order Subtractive Crossover Network

Figure 2 shows the schematic of the subtractive filter, and for comparative purposes, a conventional filter is also shown. A conventional high pass first order filter is used, although a low pass filter can also be used and gives identical overall results. The subtraction circuit is simply a common balanced amplifier, which only amplifies the difference between its two inputs. The frequency and phase responses are shown below, and they are identical to a 'normal' 6dB filter response. The summed output is flat, having no peaks or dips at the crossover frequency. Since a straight line is hardly inspiring to look at, this has not been included for this or any of the graphs that follow.

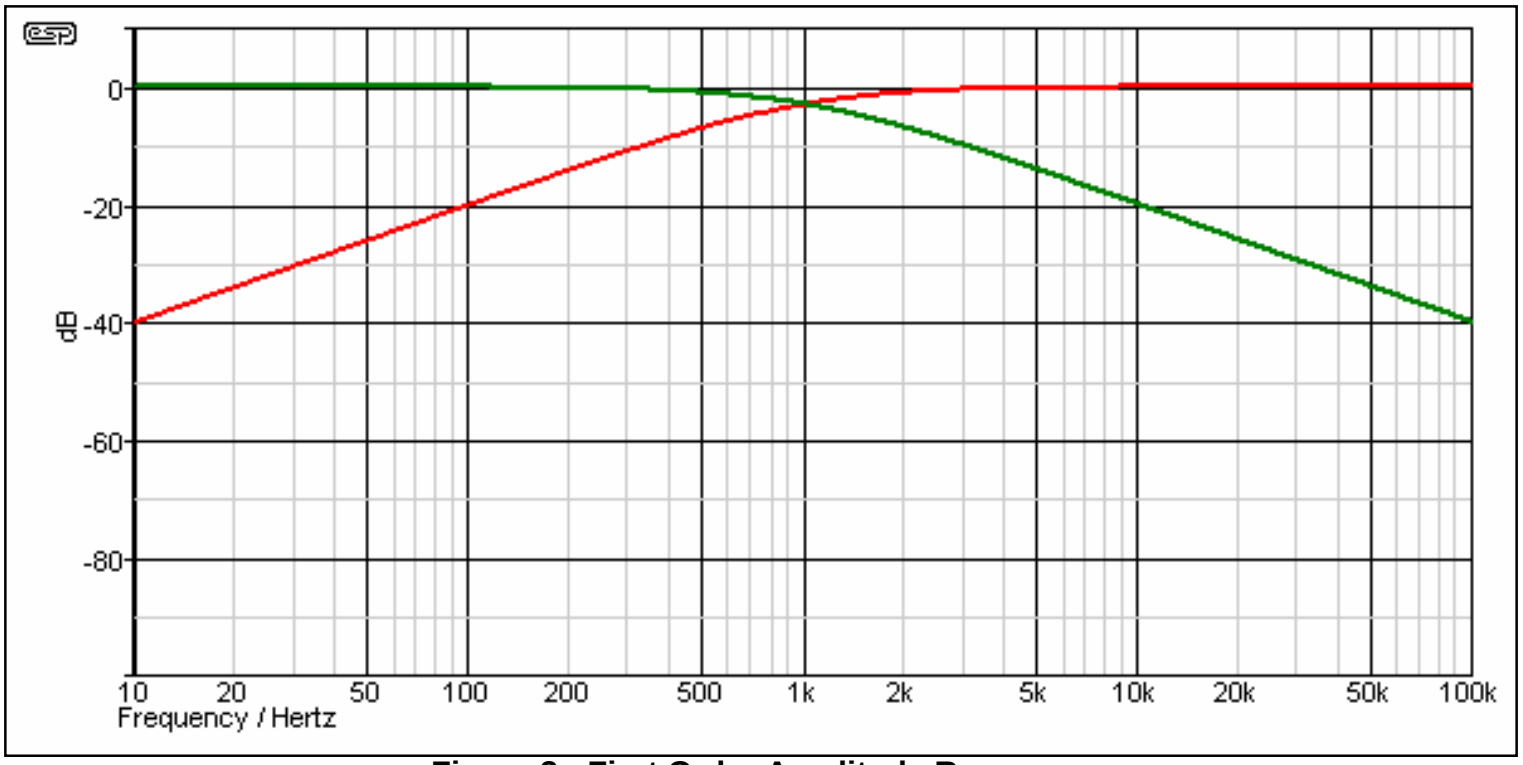


Figure 3 - First Order Amplitude Response

The amplitude response is as one would expect, and requires little or no further comment. As stated above, this is identical to a conventional first order filter response.

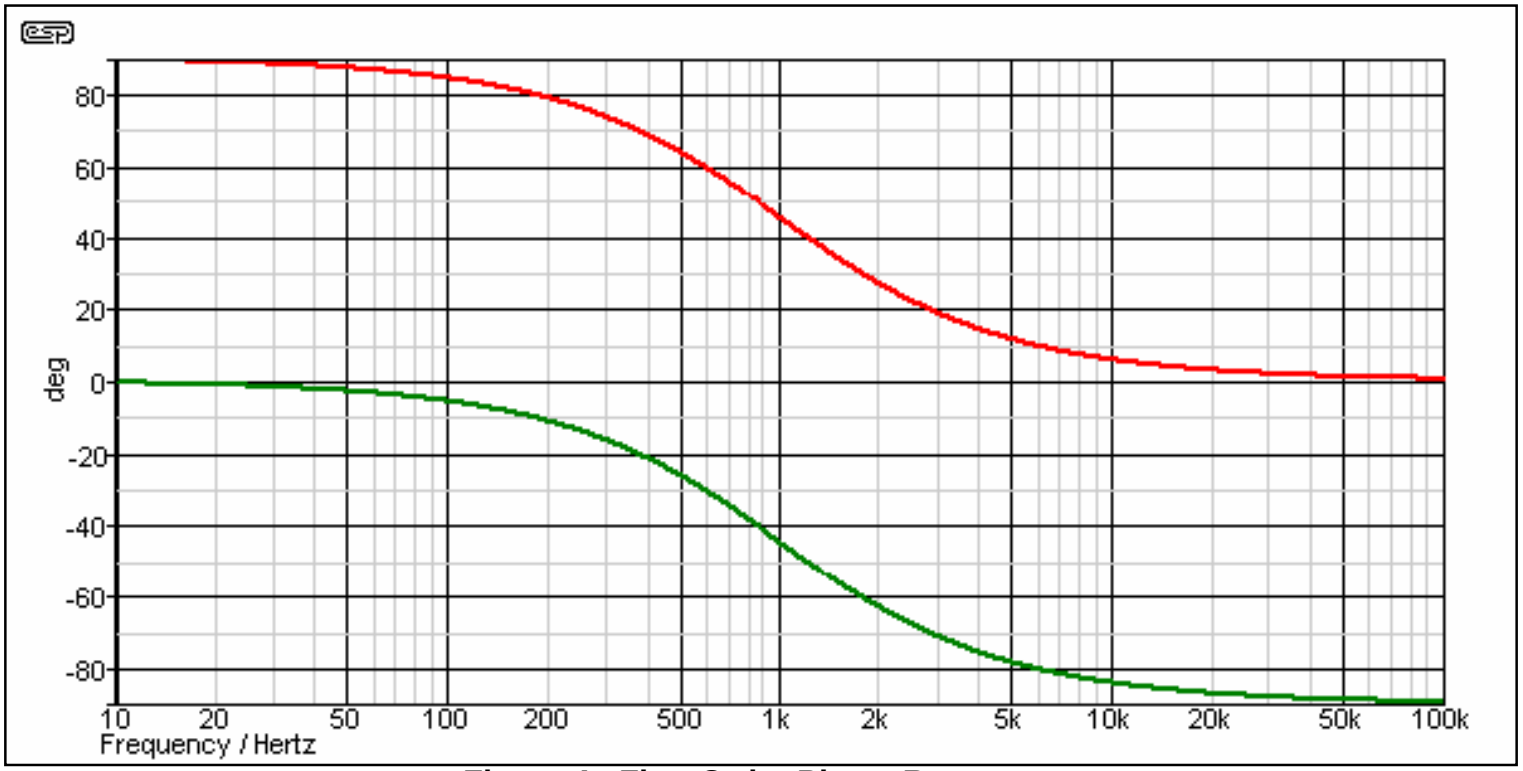


Figure 4 - First Order Phase Response

Phase response again shows the normal behaviour for a first order filter. In all cases in this article, the red curve is for the high pass filter, and the green curve is the low pass. As noted

earlier, there is no point using the subtractive method for a 6dB/octave crossover - the above is by way of example only.

2.2 - Second Order Filters

When we look at second and higher orders, we start to see real difference between the subtractive filter and other more conventional crossover networks. Figure 5 shows the schematic, and it must be admitted that it is a little simpler than a standard Linkwitz Riley filter (for example). While the difference in complexity is not great, the summed response is better, and unlike nearly all filter networks above first order, it is not only phase coherent, but the summed signal reproduces a perfect squarewave.

It is at this point that some people get excited - any filter that can pass a squarewave *must* be better than one that cannot, and in truth, virtually no conventional filter above first order can reconstruct a squarewave. The subtractive versions therefore must be better.

As we will see later on, this is not necessarily true, and the ability to reproduce a true squarewave is vastly overrated. Apart from anything else, we rarely listen to any audio signal that even approaches a squarewave, but there are other relevant factors that will wait until the conclusion of this article.

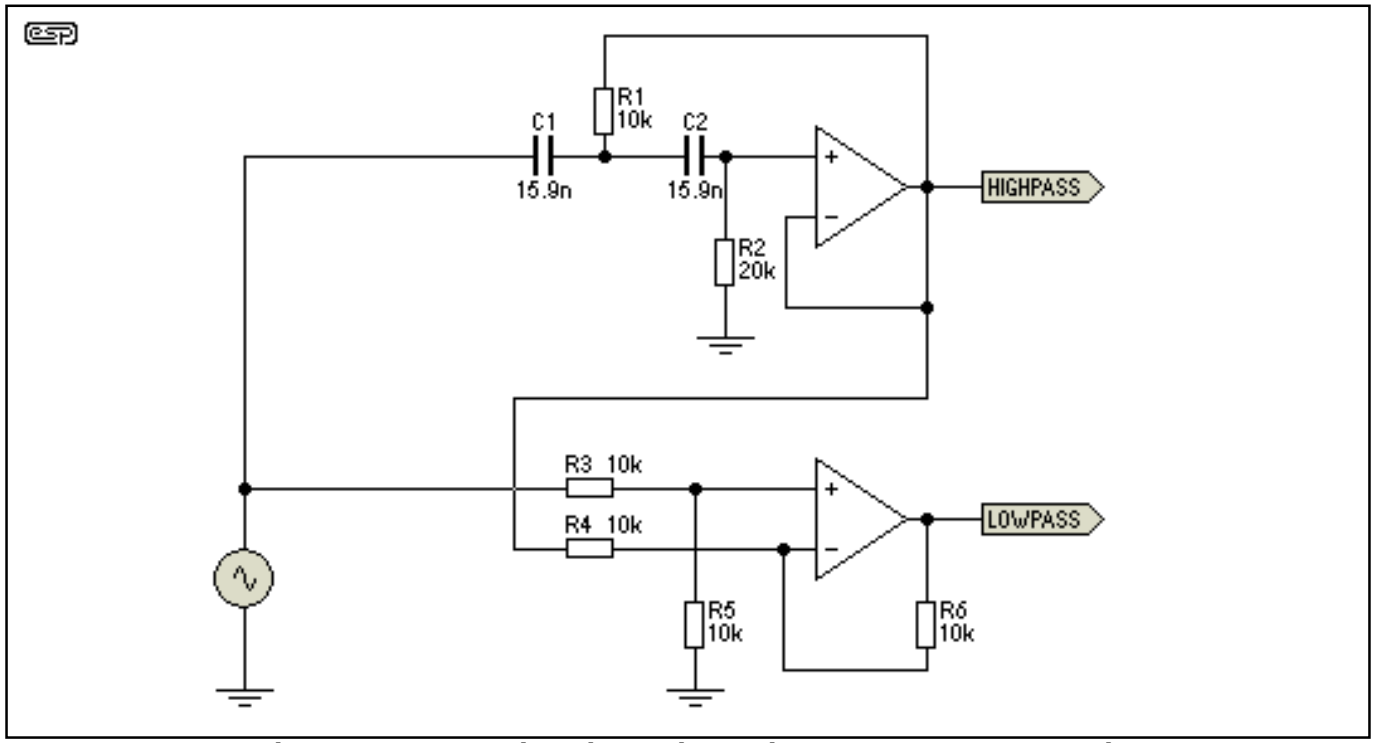


Figure 5 - Second Order Subtractive Crossover Network

Above, we see the schematic for a second order (12dB/octave) derived crossover. A single second order Butterworth highpass section is used, with the difference amplifier subtracting

the filter's response from the input signal. One would think that by doing this, the derived filter would match the rolloff characteristics of the filter, but this is thwarted by phase shift. It is phase shift that causes the derived rolloff slope to remain at 6dB/octave, and although it is possible to include a phase shift network to equalise the slopes, this will no longer allow the filter to recreate a squarewave, and it will behave the same as any other filter network.

In fact, various magazines (Elektor being one that I know of - thanks to a reader) have published projects that use a combination of a standard subtractive crossover with a phase shift (all pass) network. This does equalise the rolloff slopes, but the network behaves in the same way as a conventional crossover network.

The 'saving' is two capacitors, but you need more resistors and one additional opamp (not much of a saving). The circuit complexity is greater than a conventional filter because the repetition is replaced by a relatively complex phase shift network plus a summing amp. This makes it more likely that a mistake will be made while wiring the circuit. IMO there is absolutely no benefit, and it is far easier to build a conventional Sallen-Key based filter network such as that shown in Project 09.

Unless there is significant demand, no further details will be provided for this design. It is interesting from an acedemic perspective, but contributes very little in real terms.

Look carefully at the graph below ... as explained above, while the high pass section certainly rolls off at 12dB/octave, the derived low pass section is indeed only 6dB/octave. This is one of the greatest disadvantages of the subtractive crossover. The derived filter is *always* 6dB/octave, regardless of the rolloff slope of the filter itself. (However, see note above.)

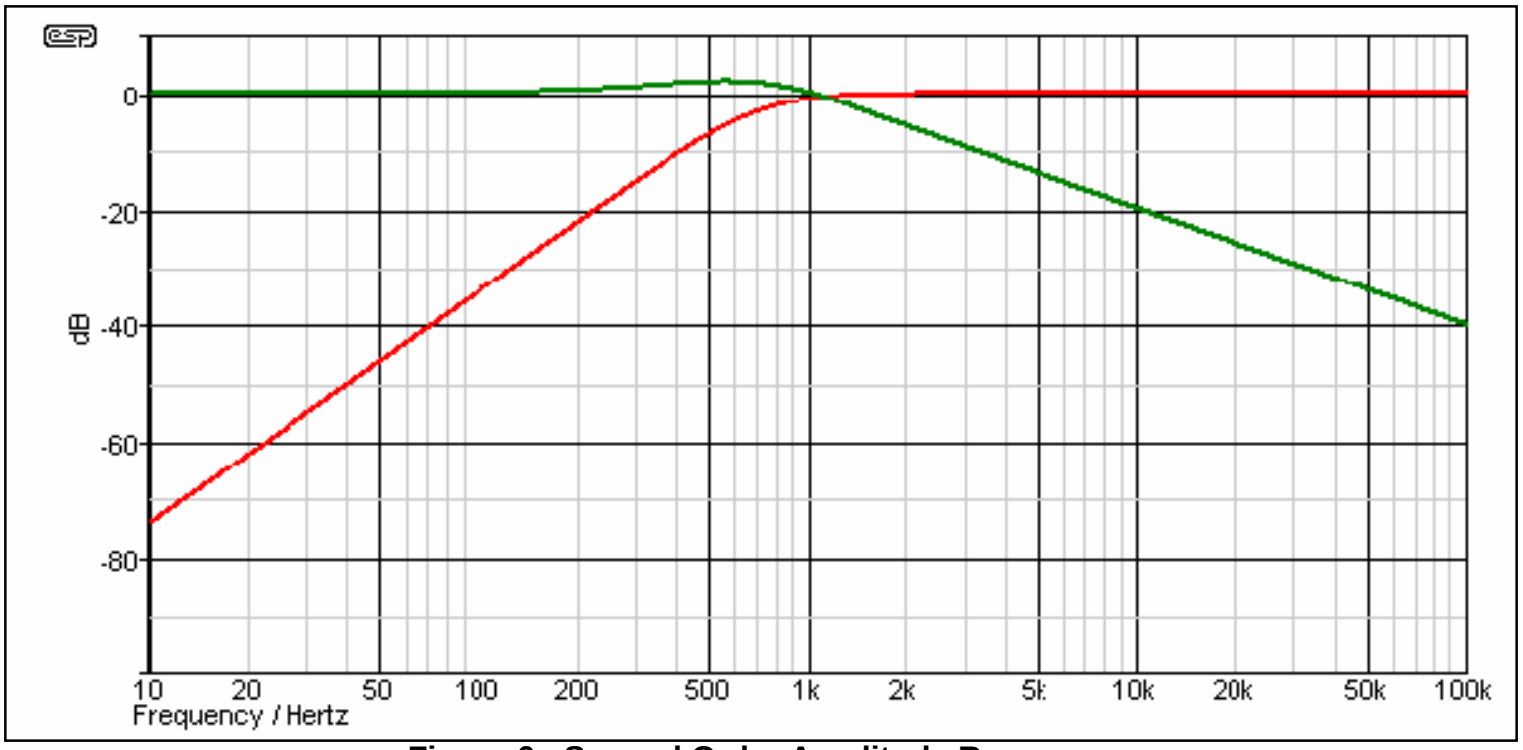


Figure 6 - Second Order Amplitude Response

Potentially of some concern is the peak in the low pass response, just before it starts to roll off. This can be reduced by reducing the Q of the filter. While it is not serious, the expectation is that the tweeter will have sufficient output at this frequency to cancel the acoustic peak, thus restoring flat response. As discussed in greater detail below, this may be wishful thinking.

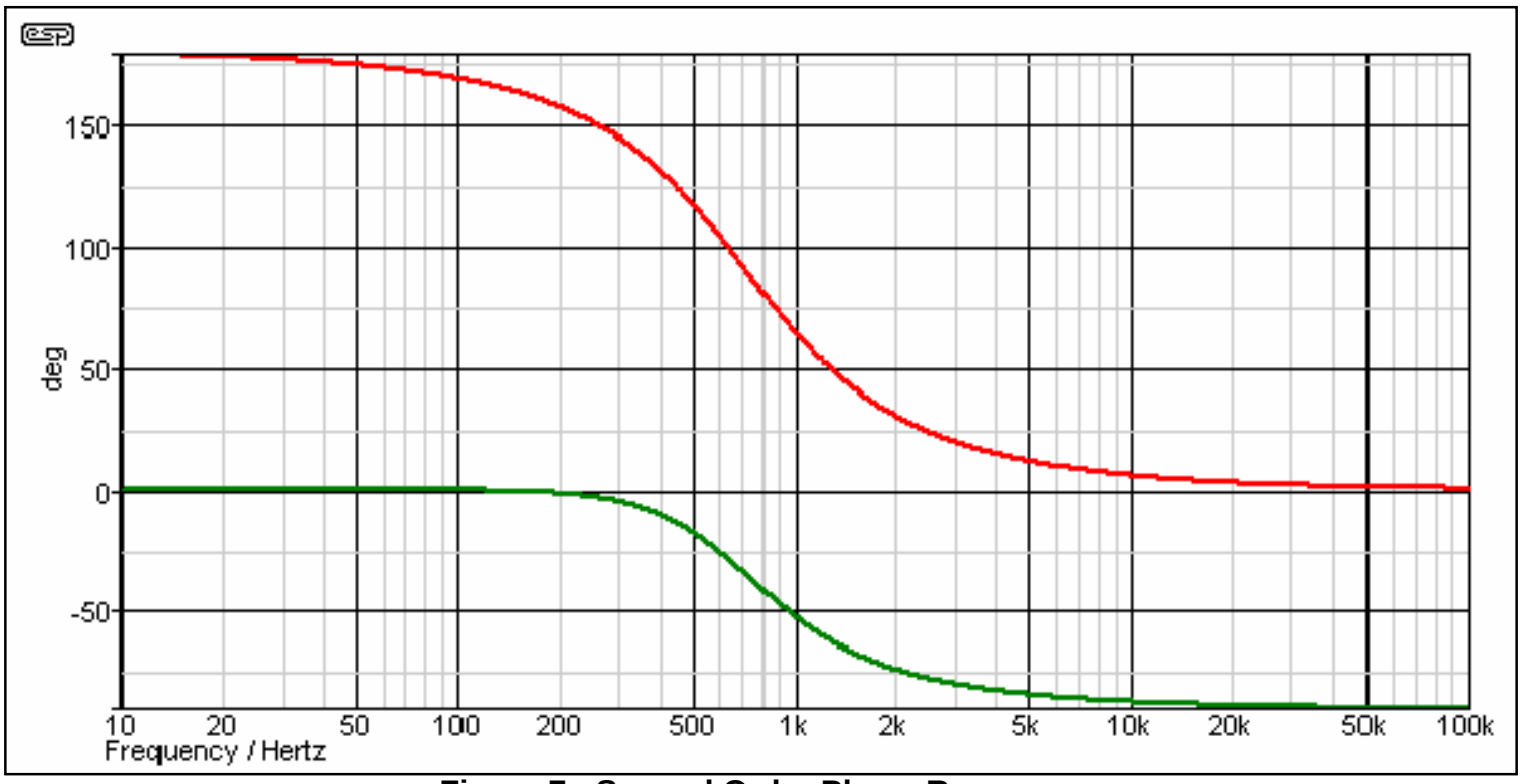


Figure 7 - Second Order Phase Response

The phase response is shown above. It is seemingly impossible that two outputs with such frequency and phase responses could possibly be summed to a flat response, but they do, and this filter (just like the first order network) can pass a perfect squarewave when the outputs are summed. Likewise, the summed response is completely flat, with no peaks or dips.

It is rather unlikely that the acoustic outputs from the drivers will be able to match an electrical summing network, so it is less likely that the acoustic output will be flat. Electrical and acoustic summing are not the same thing, and although electrical summing is an effective way to find out the ideal response of the system, what happens in reality is likely to be altogether different.

2.3 - Fourth Order Filters

Finally, the circuit diagram below shows a derived 24dB/octave (fourth order) network. Where this should offer the best response, in fact it is the worst of the three shown.

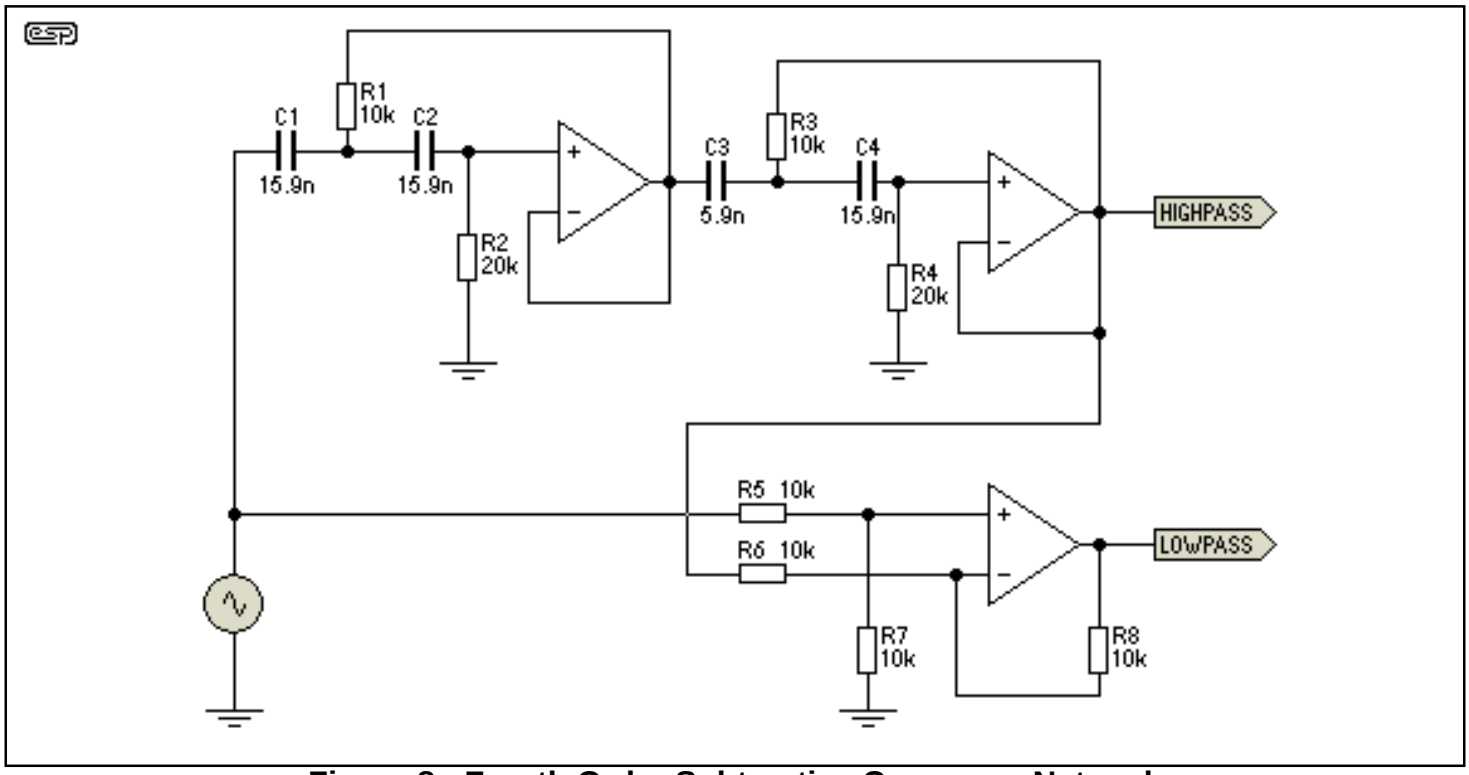


Figure 8 - Fourth Order Subtractive Crossover Network

The amplitude response (below) shows that there is a substantial rise in the response of the low pass section (the derived part of the network). If (and that is a very big ask indeed) the drivers sum as flat as an electrical network, then there isn't much of a problem. It is highly unlikely that the drivers will be able to produce a flat response in reality.

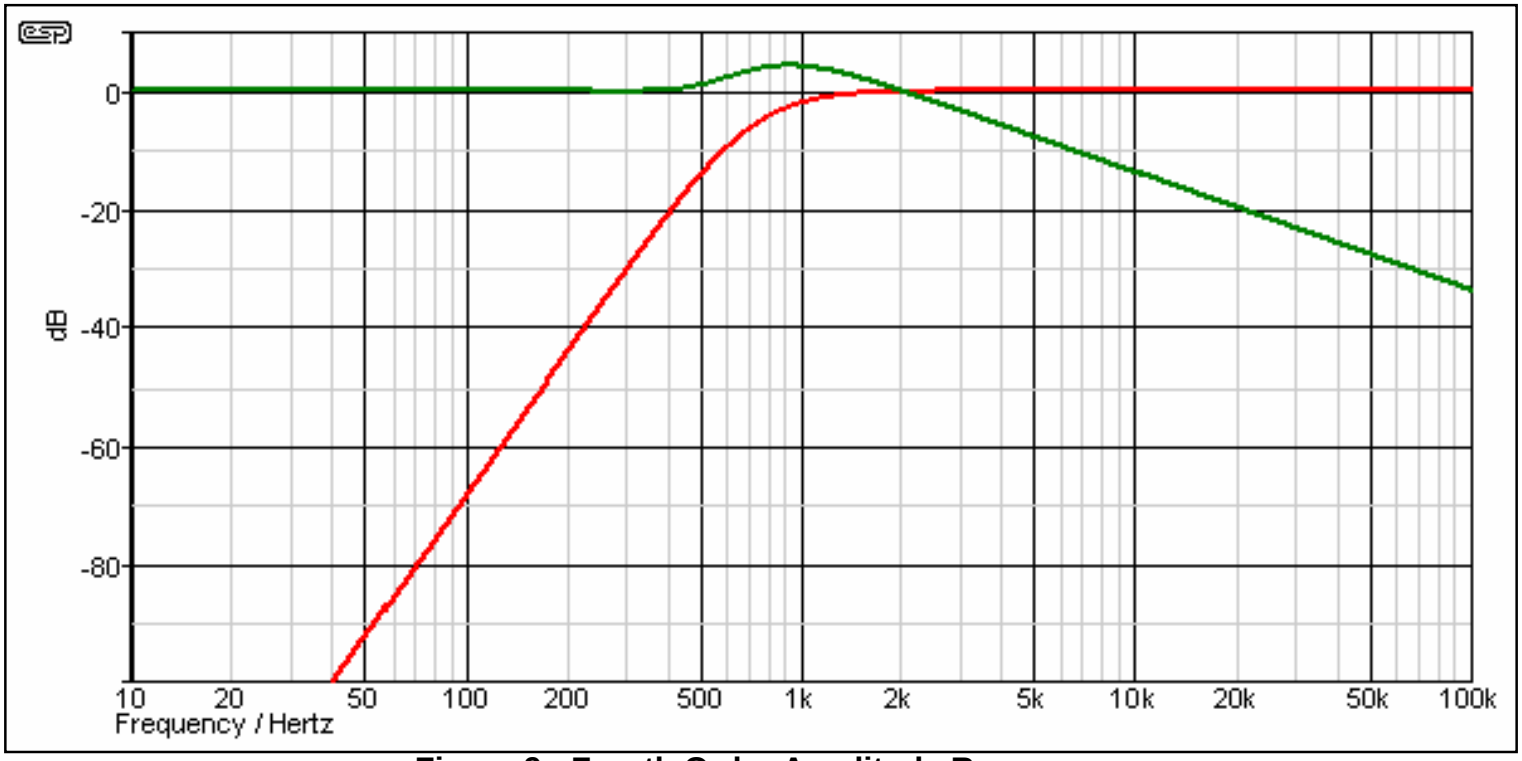


Figure 9 - Fourth Order Amplitude Response

The response peak is about 4dB, and that represents more than double the power applied to the driver over that frequency band. The amount of frequency overlap is (IMO) completely unacceptable, and a system built using this crossover would have to use accurate time alignment. Great care would also be needed to ensure that the polar response of the drivers is very similar over at least 3 octaves across the crossover frequency.

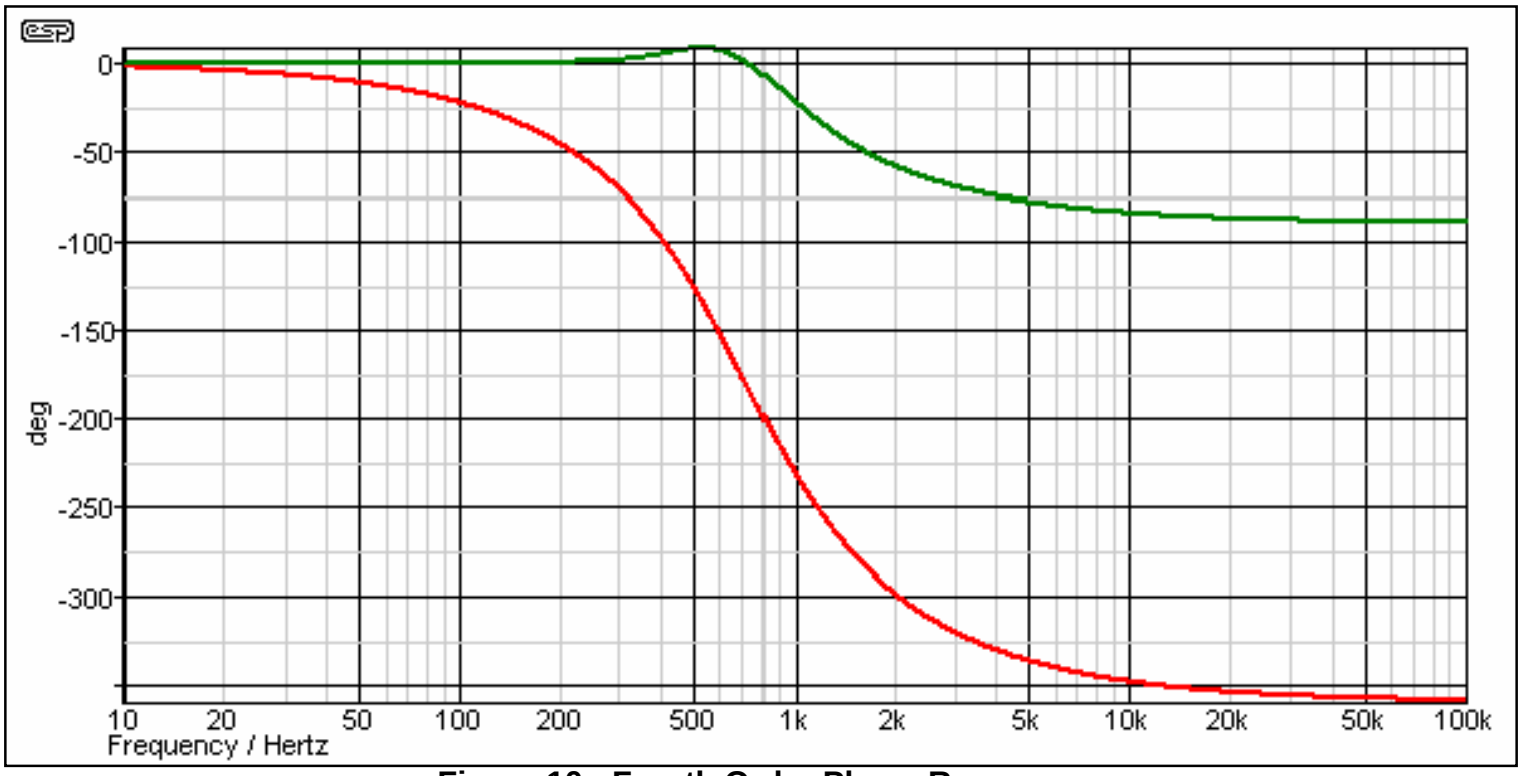


Figure 10 - Fourth Order Phase Response

The phase response also shows a peak, but this is of less consequence than the amplitude peak. In all subtractive filters, they are *not* phase coherent. That is to say that the phase of the signal applied to each driver varies, and the two are not in phase across the crossover region. Unless the phase response of the drivers is very predictable (no phase shift from voicecoil inductance or resonances) the two signals can no longer sum flat - even electrically. Acoustic summing will be worse than electrical summing in all cases.

3.0 - Linkwitz-Riley Filter Comparison

So that everything can be seen in the one article, I have included a schematic of a 24dB/ octave crossover, along with the amplitude and phase response. The first thing you will see is that there is actually little additional complexity. Rather than a complex circuit, it is simply repetitious. This is the simplest of the topologies that will give the desired overall response.

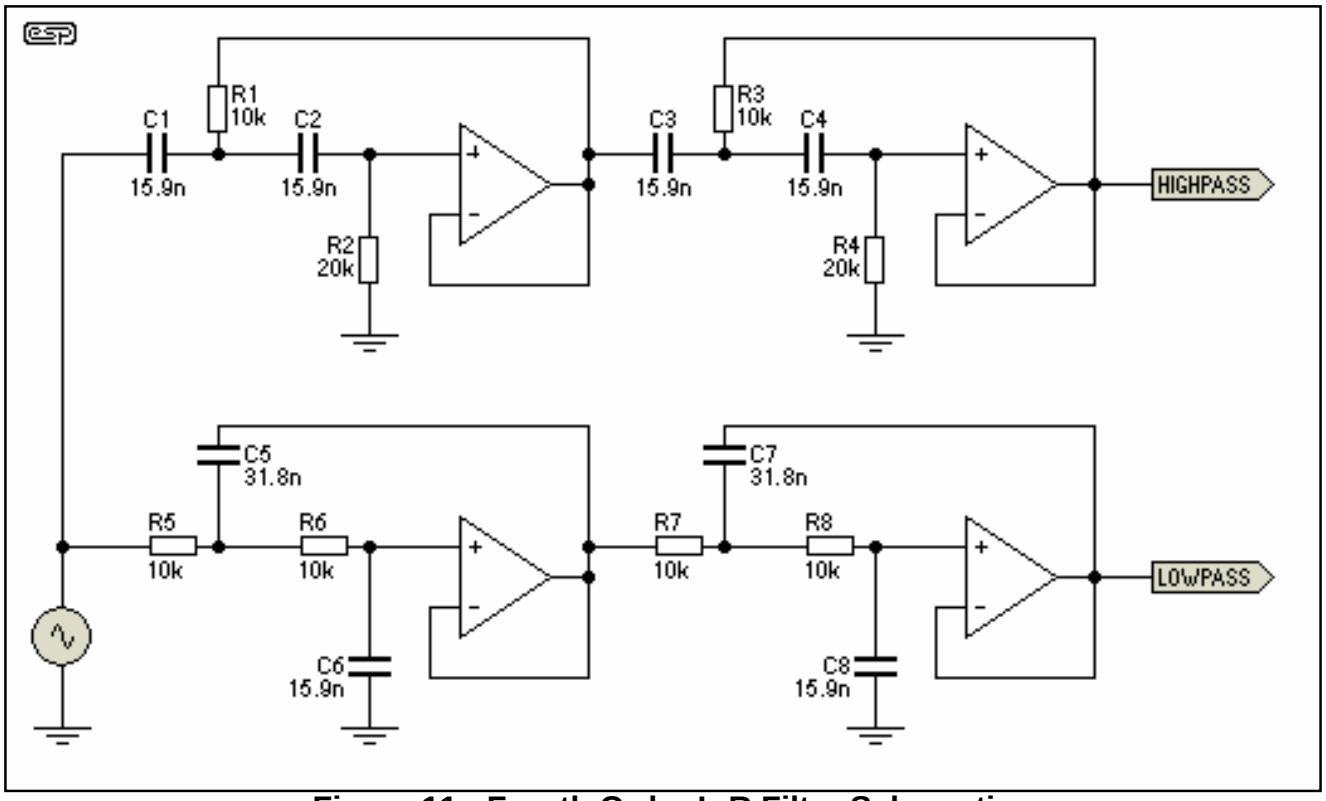


Figure 11 - Fourth Order L-R Filter Schematic

The high pass section is at the top, with the low pass section at the bottom. This is identical to the circuit used in Project 09, which has been a popular project from the time it was first published.

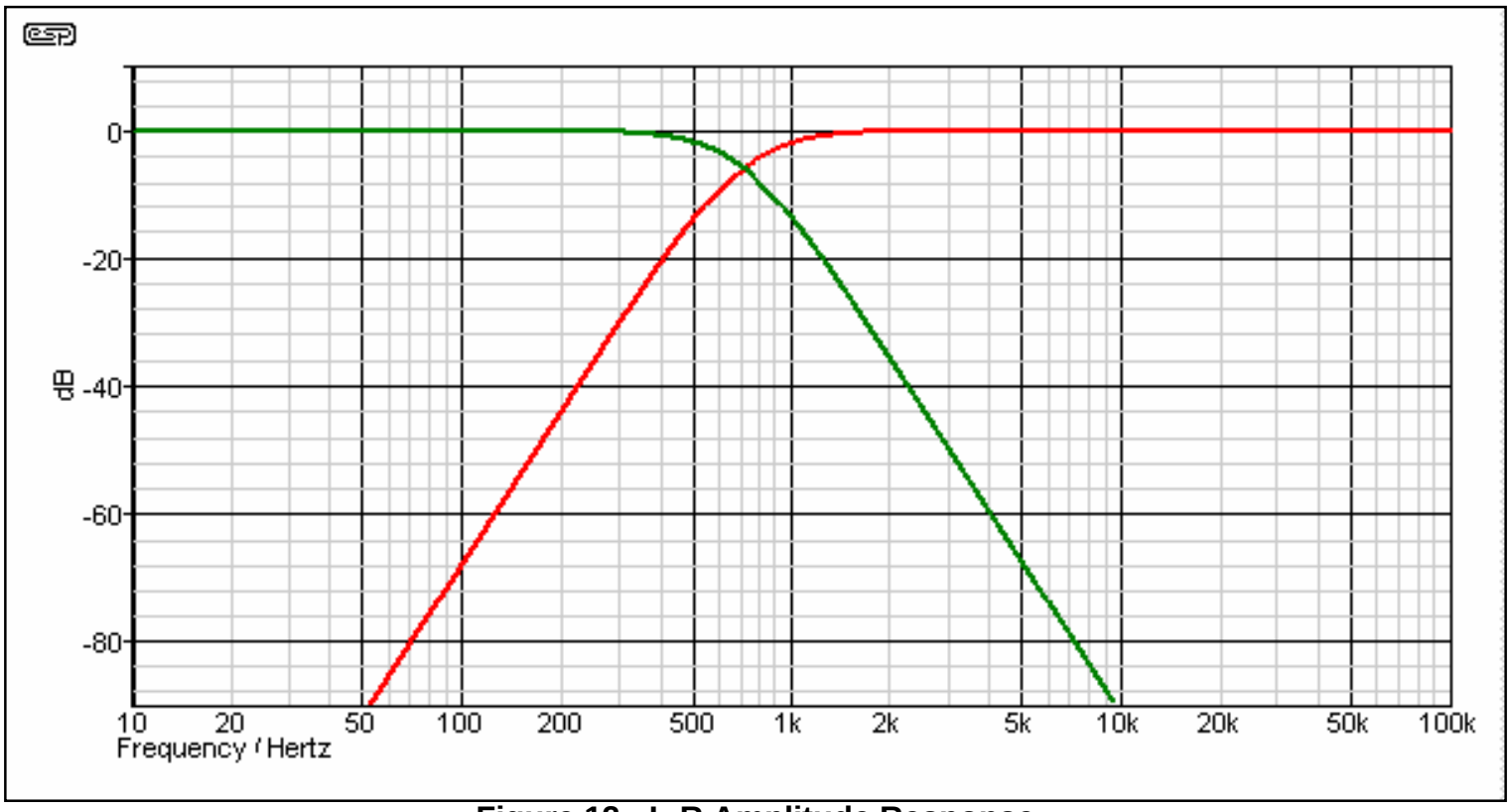


Figure 12 - L-R Amplitude Response

Amplitude response is exactly as we would expect. A nice steep rolloff for both sections, and a clearly defined crossover frequency. Because of the Linkwitz-Riley alignment, the summed output is completely flat (just like the subtractive filters), but without any of the associated problems of excessive overlap. No, it won't pass a squarewave, but the summed output still contains every frequency that made up the original squarewave, and testing by ear reveals that it is not possible to positively identify the squarewave from the modified version. While there is almost always a difference, the nature of the difference has more to do with the loudspeakers than the crossover.

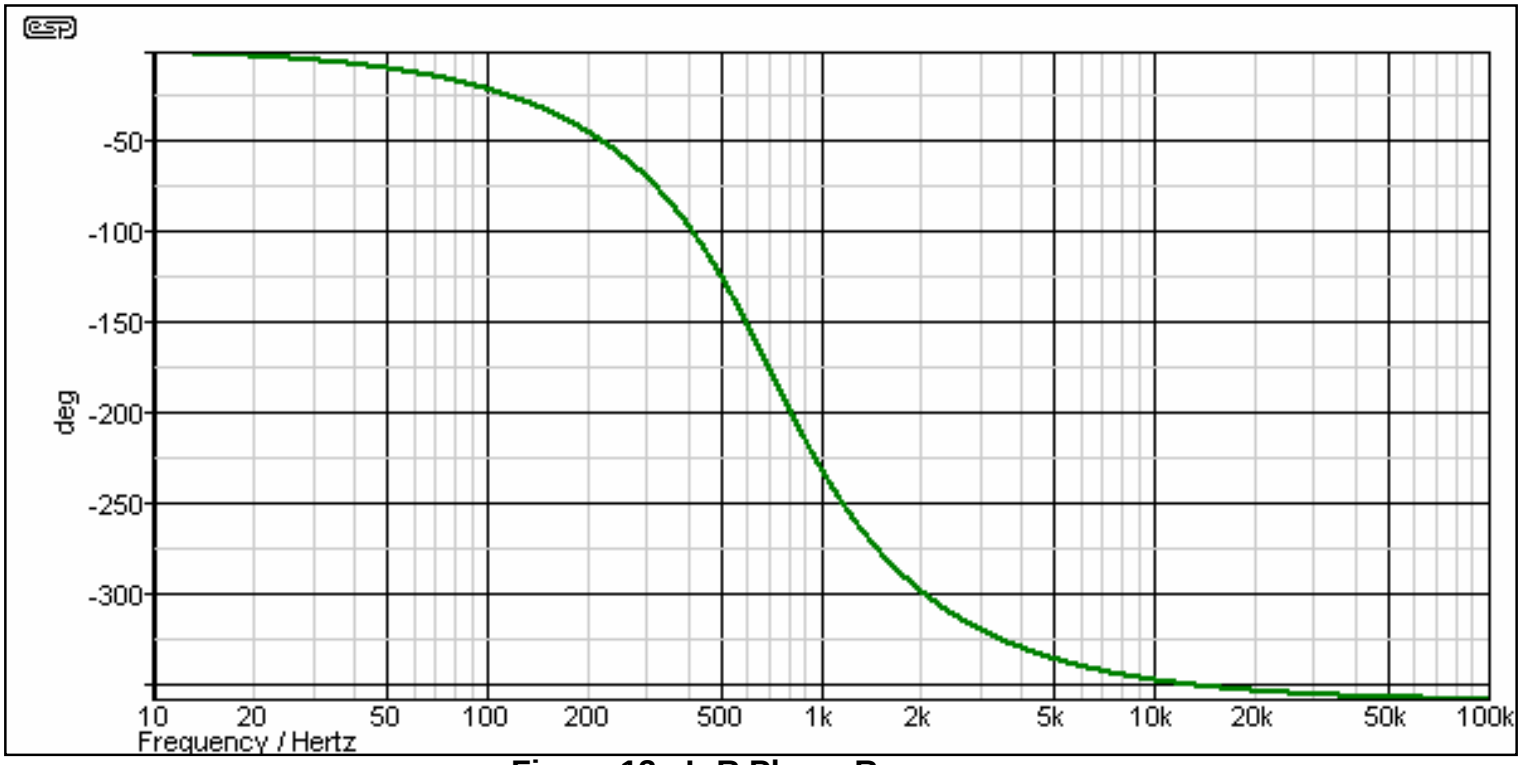


Figure 13 - L-R Phase Response

In case you wondered, no, I didn't leave out the high frequency phase response. It is simply overwritten by the low frequency graph - they are perfectly overlaid. That means that the two drivers remain in phase over the entire frequency range. This filter network relies on proper filtering, rather than hoping that the acoustic outputs of the drivers will complete the job that in reality is only half done by a subtractive filter.

4.0 - Conclusion

The first - and possibly the most important - thing that must be understood is that electrical and acoustical summing are not the same thing. Just because a crossover network sums flat electrically, this does not imply that it must also sum flat acoustically. With subtractive crossovers, the very worst scenario is presented to the drivers, where there is considerable frequency overlap between the adjacent loudspeaker drivers, and unless they have identical polar response over the entire overlap region (and at least an octave either side), the

combined acoustic output will be anything *but* flat. This seems to have been missed by many of the proponents of these filters.

Unlike conventional filters, where the higher the order sections have less overlap than low order, the subtractive networks present the opposite case. The derived section using a 24dB/octave high pass section has the greatest overlap, and we can see from the above that the 6dB network is actually the best in this respect. Let us simply say that this is less than desirable (note careful use of understatement).

The next issue is the derived filter section's rolloff slope - 6dB/octave. All the circuits above derived the low pass section, because that gives the greatest protection for tweeters (and midrange) drivers against excessive excursion. Quite a few published circuits over the years have derived the high pass section, and this places extreme demands on the drivers because of the power delivered below the crossover frequency. In addition, there is the peak at the very frequency where it is least desirable - at the lowest frequency the driver is meant to handle. It gets additional power at that frequency, increasing excursion and hence intermodulation distortion.

Speaking of crossover frequency, it is almost impossible to predict exactly where it is. It is obvious in the first order example, but as the filter order is increased, so too is the overlap region. One might want to use the -3dB frequency of the actual filter as a guide, but that's all it really is - a guide.

So, it should now be obvious that subtractive crossovers are most certainly not the Holy Grail, and in my opinion are virtually useless. Increased overlap at crossover may cause excessive beaming because the drivers are working as a mini-array, poor rolloff slope of the derived filter section can allow cone breakup (or if reversed, will probably cause excessive intermodulation), all because they can reproduce a squarewave. I think not.

The phase shifts caused by conventional crossover networks may seem extreme, but they are generally inaudible. Provided the phase of each driver is controlled and maintained (such as with a Linkwitz-Riley crossover), there are no audible effects. While phase anomalies may be audible if two different speaker systems are operated alongside each other, this is not a problem for home audio systems. The subtractive crossover network still has overall phase shift between drivers, so it doesn't solve that particular problem anyway.

So, if anyone was ever mildly curious, now you know why I have not (and will not) publish a project based on what I consider to be a seriously flawed design.

References

"Active Phase-Linear Crossover Network" , Elektor Electronics, September 1987

The above may not be entirely correct - I have not been able to verify the exact article title nor the year of publication. Unfortunately, the Elektor site only maintains its on-line index to 1998.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 20 Sep 2005

The Audio Pages

Elliott Sound Products **Doppler distortion in loudspeakers - Real or Imaginary?**

Copyright © 2004 - Rod Elliott (ESP)
Page Created 20 Aug 2004
Updated 09 Dec 2004

Articles Index

Main Index

Contents

- 1.0 Preamble
  - 1.1 Introduction
- 2.0 The Doppler Effect
  - 2.1 Frequency Modulation Characteristics
  - 2.2 Phase Modulation Characteristics
  - 2.3 Test Methodology
- 3.0 Test Results
- 4.0 Conclusion
- 5.0 References

1.0 - Preamble

In order to save everyone time (including and especially me), I must make a couple of points before anything else is discussed ...

- Had the 'Doppler distortion' effect in loudspeakers been referred to as Phase Modulation Distortion (as I suggest below) from the outset, the vast majority of the arguments denying its existence would never have occurred, since it can be demonstrated by simple reasoning.
- I *know* that the Doppler effect as it is commonly described involves phase shift, and that this is simply a different way of 'looking' at the phenomenon.
- I do not expect you to agree with me - the following is simply my suggestion. It is based on simple logic, describing a well known and real distortion mechanism in loudspeakers, but in a manner that (hopefully) makes a bit more sense than the purely theoretical coverage that most websites present.

Please make sure that you read and understand the above before you decide it is necessary to send me an e-mail pointing out things that I have already acknowledged either above or in the body of the article itself. Should you get the impression that I am sick of e-mails pointing out that the Doppler effect involves a change of phase, or telling me that 'Doppler effect' is the best way to describe the loudspeaker distortion, you are quite correct. I know the phase effect, and need to point out that had the effect been called Phase Modulation Distortion from the

outset, this article (and probably hundreds of others) would never have been written.

To say that this article has stirred up a hornets' nest is putting it mildly. Since the initial article was published, there has been a great deal of correspondence, both for and against what I have written. Many people seem to have missed the point, so to make absolutely sure that it is not missed, I will make it right at the beginning ...

Frequency shift in a loudspeaker is real - it does happen, and the amplitude (frequency deviation) of the shift is *identical* to that predicted by 'conventional' Doppler theory. The peak phase shift occurs at the extremes of the cone movement as one should expect, and indeed, there is zero phase shift at the low frequency zero crossing - the cone's rest position. The rate of change of phase is greatest at the zero crossing point, and this corresponds to the maximum frequency shift.

> In the previous version of this article, I was under the impression that there was no phase variance at the zero crossing point, however, after some correspondence and some further (rather tedious) tests, it can be shown that the peak frequency shift does indeed occur at the point of peak cone velocity - this is not readily apparent on the waveforms shown below, because the resolution required is so great that it could not be achieved using my oscilloscope. As it transpires (and in hindsight - always an exact science), this makes sense, since the peak rate of change of phase (and therefore frequency) *must* occur during the transition from one phase angle to another. My apologies to anyone who may have been mislead.

The Doppler effect is generally applied to moving vehicles (etc.) and the most common explanation is that the wavefront is 'compressed' as the object approaches the listener, and 'stretched' as it heads away. Another way to look at this is that the *phase* of the signal is constantly changing, resulting in the familiar frequency shift. The main point here is that the effect lasts for some relatively considerable period of time, and the observed frequency shift is directly related to the velocity of the approaching (or receding) sound source.

Consider a sound source approaching at 30m/s (108km/h). The sound source is generating a frequency of 1kHz, so as it approaches, the listener will hear not 1kHz, but about 1095Hz - a clearly audible change. As the sound source recedes, the perceived frequency will be 912Hz - again, readily audible (see [Ref 4] for the calculator that I used). As each successive compression or rarefaction of the wavefront reaches the listener, the phase is effectively advanced (or retarded) by approximately 108us or 39°, thus creating 'new' frequencies that are proportional to velocity and the original frequency. I shall leave it to the reader to analyse this further if desired, but suffice to say that this phenomenon has been known for a very long time.

While the principle is well known, the 'phase method' of describing it is not - it is far more common to see the explanation deal with the compressed or expanded wavefront (and wavelength). There is no anomaly here - both descriptions are equally valid, and explain exactly the same phenomenon.

Based on the measurements shown below, it is obvious that the conventionally explained Doppler effect does not occur with a loudspeaker - the periods where the cone is moving towards or away from the listener are too short, and the cone is not moving *through* the medium (air). It could also be stated that it is not the *cone* that is making the sound, but rather the *movement* of the cone. Several correspondents have complained that what amounts to 'movement of the movement' of the cone cannot possibly be seen to be a moving sound source per se. This can create a philosophical conundrum if one delves into it deeply enough, but fortunately this is not necessary. The phase shift model explains this, and it is obviously impossible for the cone to move *without* causing phase shift.

In short, the term 'Doppler distortion' (for a loudspeaker) is completely real, but is also a simple misnomer (IMO) - the frequency shift exists, but it can be shown that the shift in phase is static as well as dynamic. Naturally, the same can be said for any sound source that moves with respect to the listener, regardless of velocity. Minimising cone excursions for any driver that must reproduce higher frequencies is the key to obtaining better sound reproduction - this much is very clear.

1.1 - Introduction

There seems to be a new area of debate on the Net, and it can be found in newsgroups, forum sites and websites in general. For a long time, the existence of so-called Doppler distortion in loudspeakers was taken for granted by most (including the author), but there are now many challenges to this claim - some are well reasoned and have considerable merit, while others are based firmly on the basis that the challenger simply doesn't *believe* that the effect exists at all.

Doppler distortion in loudspeakers is mentioned in so many articles, web sites and references that for years it was considered a simple fact. A great many authors have applied their (not inconsiderable) mathematical skills to the subject, and it is fairly obvious that many (most) of the published results are based on the assumption of cone velocity, rather than absolute position.

The effect of Doppler shift in loudspeakers is believed by many to be audible, and it is (relatively) easy to measure the frequency spectrum and see sidebands that seem to show modulation consistent with frequency modulation. Indeed, the existence of (velocity based) Doppler distortion in loudspeakers was originally established many years ago [Ref 1] and is maintained to this day.

At this point, I would like to thank Siegfried Linkwitz [Ref 2] and Art Ludwig [Ref 10] for their input to this article. My initial published data caused both gentlemen a great deal of work (which I'm sure they could have well done without), but in the spirit of international co-operation they have been very helpful, and contributed a great deal to this (final?) version of the article.

In e-mail correspondence with Siegfried Linkwitz, he said ...

> *The considerable effort you put into this difficult measurement should not go unnoticed and lead to a deeper understanding of the Doppler effect. I certainly had never thought about and made the connection between the now obvious phase shift of the high frequency tone with cone displacement and the Doppler effect. In the same way that the frequency shift is more easy to observe under certain conditions, so is the phase modulation in the case of a loudspeaker. Both are just different descriptions of the identical physical phenomenon to which Doppler's name has become attached.*

Siegfried also said that "*I have not seen the associated PM pointed out in text books, but it follows from FM or PM modulation theory*" - this is really the crux of what this article is about - that no textbooks describe the phase change (and resulting phase/ frequency modulation) as the cone moves closer to or further away from the listener, and this leads to a general misunderstanding as to the real cause of the effect, and obscures other effects that could have been extrapolated had this been more widely known. More to come on this aspect shortly, but rest assured that it will not make a large difference to the way we design (or listen to) loudspeakers, but simply explains other phenomena that were not seen or predicted previously.

This paper has been a long time coming, and has had a considerable amount of input from others - both 'believers' and 'non-believers'. Indeed, it was a non-believer who first took me to task because I referred to 'Doppler distortion' in one of my other articles, and it took some effort on his part for me to see that there was something amiss. Since he prefers to remain anonymous, he will not be named, but suffice to say that his insistence was such that I had to work very hard to prove him wrong. I initially thought that I had failed, but it turns out that the effect was real from the very beginning, but explained in such a way that made no logical sense (to me and presumably to him). Papers such as that by John Kreskovsky [Ref 3] are very convincing, but unfortunately work from the wrong premise (i.e. velocity rather than phase) in the first place. (Note that the page has been updated at some stage, and some of the original claims (as I recall) are no longer there, or have been

modified.)

The original experiments I performed - well over a year ago at the time of writing - were inconclusive, and despite a lot of effort, it was not possible to determine exactly what was happening. Even during tests to try to locate evidence of a frequency shift, while there *was* some evidence that a shift existed, the resolution of the equipment I had at the time was insufficient to be certain. In addition, I did what I suspect many others have done before me, and *removed* the low frequency component from the captured audio signal - this was a mistake, because that information is needed to be able to see exactly where the shift occurs.

This study is as much theoretical as practical - the theory was needed to predict an outcome, and to examine other possibilities. I must thank those who have listened to my rants and hypotheses and supplied additional ideas ... it was one simple point that brought the whole issue into focus, and allowed this page to (finally) become reality ... phase!

2.0 - The Doppler Effect

The Doppler effect is very real, and has been heard by everyone at some stage. The classic example is of the siren or horn on a moving vehicle, which is increased in pitch as the vehicle heads towards you, and reduced in pitch as it heads away. There are countless references and descriptions in books and the Web, and I shall not even try to cite references except one - [Ref 4]. A Web search will tell you everything you ever needed to know (plus a lot more). Doppler shift requires that the sound source is moving with respect to the listener, which means that the listener is either stationary, or moving at a different speed from the originating sound source. An ambulance driver or passenger does not hear any Doppler shift, because s/he is moving at the same speed as the siren, so there is no relative difference in velocity.

There is a requirement for the Doppler effect to exist that there be relative movement through the medium (between the sound source and the listener). It is generally assumed that the medium itself is not moving, or is moving at a lower velocity than the sound source. Should the medium be moving as well, this may increase or decrease the effect from the listener's perspective (information is rather scarce on this point).

That the phenomenon of the Doppler effect is real is not in question - what *is* in question is whether the same effect occurs in a loudspeaker reproducing more than a single tone simultaneously. Furthermore, we should clarify the term 'distortion', since the word is normally applied to a non-linear function. Should the effect be demonstrated to exist in a loudspeaker, then it is obvious that it will appear in an ideal (or theoretically perfect) driver as well, so there *is* no non linearity. Based on this, the effect probably should not be called 'distortion' (although it must be said that anything that adds frequencies that did not exist in the original recording actually *is* distortion, but that is probably a philosophical debate rather than one to be considered by the engineering fraternity).

The first thing to consider is the maximum velocity of a loudspeaker cone, when driven by a signal of any given frequency. This is generally much lower than we might imagine, and the velocity may be calculated by ...

$$V_p = 2 * \pi * f_L * X_p \qquad \text{[Ref 2]}$$

where V_p is the peak cone velocity, f_L is the low frequency and X_p is cone displacement. For example, a cone that is moving ±5mm (10mm total) at a frequency of 50Hz will have a peak velocity of 3.14m/s (11.3km/h) - this is certainly not fast, and in itself would account for a rather small frequency shift. In a system where (conventional velocity based) Doppler effect does change the frequency, the change is given by ...

$$\Delta f_H = 2 * \pi * f_L * f_H * X_p / c$$ [Ref 2]

where f_H is the high frequency (modulated) tone and **c** is the velocity of sound (nominally 345m/s). Using the velocity calculated earlier, we get a shift of 9.1Hz for a high frequency tone of 1kHz (less than 1%). Bear in mind though, that the rate of change needs to be maintained for a reasonable period of time before the frequency shift will become apparent, and 10ms (the time it takes a 50Hz signal to swing from maximum positive to maximum negative) is probably insufficient to cause an audible frequency shift. Our hearing is such that it needs a reasonable number of cycles (which varies with frequency) before we can identify the pitch of a tone.

Remember that for the (conventionally explained) Doppler effect to exist, the 'carrier' (vehicle, train, etc.) or listener must be moving *through* the medium, but a loudspeaker cone does no such thing. The medium (air) will be pushed outwards and sucked inwards by the low frequency movement of the cone, so there is no (or very little) relative movement between them - the medium is moving *with* the cone! What is heard (and most commonly incorrectly attributed to Doppler effect) is a combination of things (in descending order of importance) ...

Intermodulation distortion
Amplitude modulation
Phase shift

Of these, the only one that comes close to frequency modulation (as predicted by much existing theory and practice) is phase shift, and this is caused by the relative position of the cone at any instant in time in relation to the listener. That the shift is small is obvious, and equally obviously it depends on the peak to peak travel of the cone. Assuming a large cone movement for the low frequency signal of (say) 10mm in each direction, this represents a complete cycle (360°) shift at 34,500Hz, or 36° at 3,450Hz (for example). It must be considered that any loudspeaker that is expected to have anywhere near 20mm travel at low frequencies, *and* is expected to reproduce high frequencies as well, will almost certainly generate considerable intermodulation of the higher frequencies.

Interestingly, the phase shift observed will be exactly the same at DC as at any low frequency that causes the same displacement - even moving the whole box relative to the microphone will generate the same amount of (static) phase shift. This is simply a function of the velocity of sound in air, and the wavelength of the high frequency tone. A 10mm change in the position of the box will cause a 36° phase shift (of a 3,450Hz tone) in exactly the same way that 10mm of cone movement will - this is a simple physical relationship.

An interesting effect of amplitude modulation (within the range we can expect from a loudspeaker) is that when you hear it, it sounds like frequency modulation. Again, this is a well known effect, and references abound. The effect is caused by the way our hearing works, and this may be the predominant effect that is heard by listeners who have 'proven' that Doppler distortion exists because they have 'heard' it. In general, the pitch seems to reduce slightly as amplitude is increased for frequencies below around 2kHz, while for frequencies above 2kHz the pitch increases with increased amplitude [Ref 5].

The majority of studies (on websites or elsewhere) that have shown that Doppler distortion *does* exist, have used a spectrum analyser (or FFT - Fast Fourier Transform) to show the sidebands so generated. The spectrum analyser is completely the wrong tool to use for this, as it only shows frequency information - the effects we want to capture are in the time domain, and are best seen with an oscilloscope.

The theory of Doppler shift is velocity based (although as noted above, it can also be explained by phase) - the higher the (relative) velocity between source and listener, the greater the frequency shift. This is extremely difficult to even attempt to measure, since the peak velocity of the cone is so low. With phase shift, velocity is not

relevant - only the variation of distance between source and listener determines the shift, and it can be shown that the peak phase shift coincides with the peak excursion of the cone (positive and negative).

In order to dispense with the idea of loudspeaker Doppler distortion, we must find the flaw in the seemingly intuitive and rather persuasive argument that has been used historically to 'prove' that Doppler distortion does in fact exist. Simply making the point that the sound source must move *through* the medium with respect to the listener is not sufficient, and it has been argued that the cone does in fact do just that. That this action would violate the laws of gaseous matter when subjected to a force is (apparently) insufficient explanation, so further proof is needed.

That there is a shift is not in question - what is at question is the exact nature of the shift, and where it occurs on the low frequency waveform. If the shift observed is indeed in frequency, and caused by the Doppler effect, then it must occur at the point where cone velocity is at its peak - at the zero crossing point of the waveform. However, if the shift is simply one of phase, then the peak (phase) shift will occur at the extremes of excursion, where the cone is effectively stationary. So which is it? Indeed, are the two effectively identical?

One of the biggest difficulties with measuring any of these effects is the fact that the delays (and associated phase shifts) are very small. . If we assume that the effect is indeed Doppler induced and that the deviation is 9Hz as predicted for our demonstration system, then the delay is only (approx.) ±28us. Measuring this in conjunction with ambient noise, distortion, and instrument resolution limitations is not easy, since the waveform period is 1ms - the absolute maximum shift possible is ±2.8%, but is likely to be a lot less. While these things are not a limitation with a simulator or mathematical approach, real life makes it a lot harder for us.

2.1 - Frequency Modulation Characteristics

It is useful to look closely at the effects of frequency modulation, so that we know what we are looking for. A lack of understanding at this level must result in a flawed conclusion, so a thorough understanding of what we *should* find is imperative. If we find exactly the signal characteristics as described below, then the original theory is proven - this is not open to conjecture, since we are dealing with simple facts.

However, should we get results that are quite different, then the observed phenomenon cannot be frequency modulation, and is therefore not related to the Doppler effect.

The signal frequencies used for the physical tests performed later will be different from those used in the theoretical discussion that follows, for reasons that (I hope) will become clear later, but driver total excursion will be as close to ±5mm (10mm total travel) as possible.

For the theoretical examination, the low frequency will be 50Hz, the high frequency 1kHz and the peak to peak cone excursion shall be 10mm (±5mm), therefore frequency shift will be 9Hz (as predicted by the formulae above). The high frequency tone will be modulated at the 50Hz rate, to obtain a maximum of 1009Hz and a minimum of 991Hz. This represents a Modulation Index of 0.18 [Ref 6]. Let's do a quick examination of wavelength, based on the formula above that predicts a frequency change of ±9Hz.

W = c / f

where **W** is wavelength in metres, **c** is the velocity of sound (nominally 345m/s) and **f** is frequency in Hertz.

W = 345 / 1000= 0.345m = 345mm
W = 345 / 991= 0.34813m = 348.13mm
W = 345 / 1009= 0.34192m = 341.92mm

The total change in wavelength is therefore 6.21mm. We could analyse this data forever (or until we became bored), but that is probably enough for now.

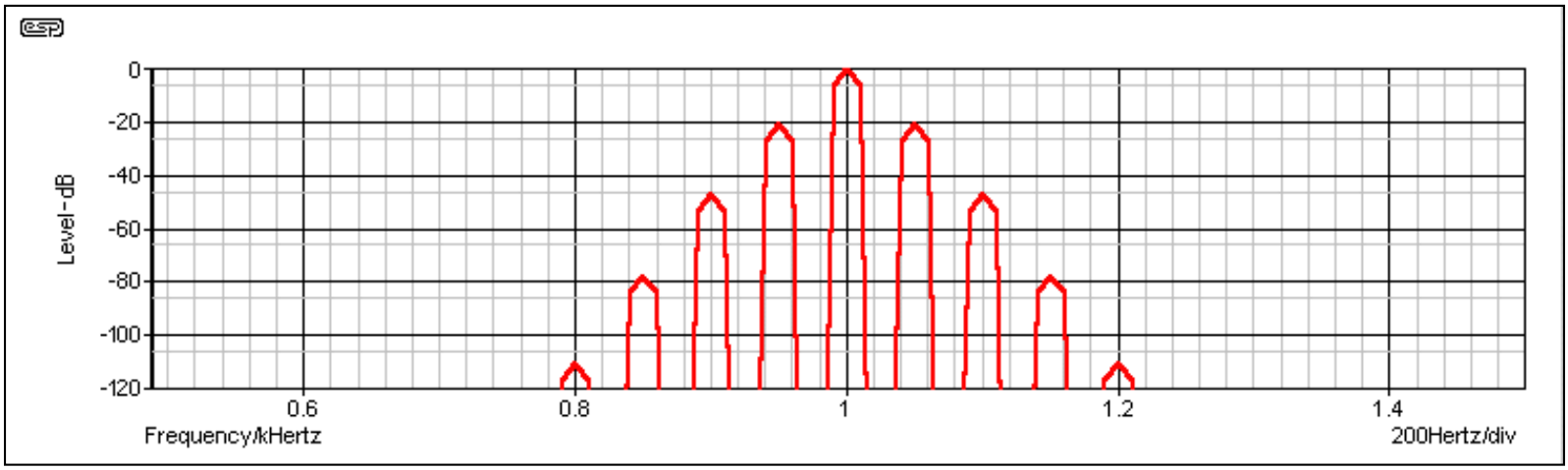


Figure 2.1 - Frequency Spectrum of 1kHz, Modulated by 9Hz at 50Hz Rate

As can be seen from the spectrum display (from a synthesised FM waveform), a frequency modulated (FM) waveform modulated as described has multiple sidebands - although only a limited number is shown for clarity, they are actually infinite and diminish into the noise floor quite rapidly. It is easily determined that the fundamental is at 1kHz (at a level of 0dB), with the first set of sidebands at -20dB (950Hz and 1050Hz). The next set is at -48dB, at frequencies of 900Hz an 1100Hz. The final set is measured at -78dB, at frequencies of 850Hz and 1150Hz. After that, the sidebands are at -111dB or lower - well into the noise floor. A pure sinewave predictably shows a single peak at 1kHz - no surprises there, and it is not shown.

The above assumes that the shift will actually be 9Hz - it has already been explained that this will almost certainly not be the case. In a listening test on a 1kHz tone, modulated by the full 9Hz at a 50Hz rate, it is possible to hear the FM as a slight roughness in the tone ... however using headphones (to eliminate standing waves, room reflections, etc.), the roughness is barely audible. if we assume that the frequency shift is only 4.5Hz - probably still unrealistic, but useful for our purposes - the signal through headphones sounds almost as clean as a pure sinewave.

2.2 - Phase Modulation Characteristics

Phase modulation [Ref 7] is actually very similar to frequency modulation - so much so that some narrow-band FM communications systems actually use PM (Phase Modulation) because it is easier to implement. Unfortunately, while it may well be easier for radio frequency signals, it is harder to synthesise than FM in audio, and no sound editor that I have access to is capable of creating a PM signal. The end result is sufficiently similar though (audibly speaking) that it is almost immaterial.

What is certainly *not* immaterial is the spectrum of a PM signal, so I was forced to build a phase modulator (using my Simulator and a standard analogue phase shifting circuit) that would operate satisfactorily at the modulation frequency of 50Hz. This proved irksome, but the results were gratifying, as it is now possible to display a spectrum analysis of a PM waveform, with (almost) the same characteristics as the FM signal referred to in section 2.1. This is an important step - without this, we do not have a useful reference and may be left with conjecture. Imagine the embarrassment if it had transpired that PM has a spectrum similar to intermodulation.

For the sake of this exercise, the spectrum of the PM signal is not shown - it is essentially *identical* to that shown above for FM. For all intents and purposes there is no difference whatsoever. Although there actually is a (very)

subtle difference, it is not visible until it is below the -120dB level, so is irrelevant in practical terms.

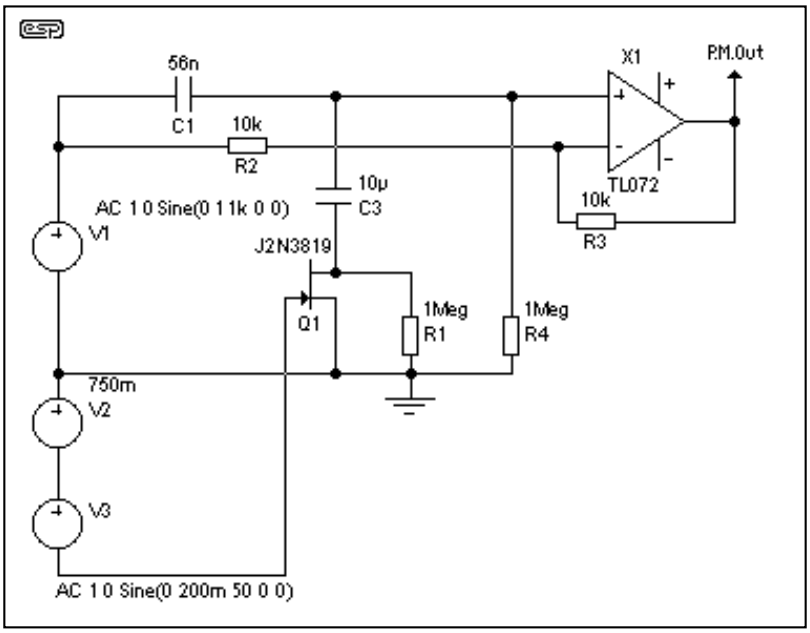


Figure 2.2.1 - Phase Modulator as used to Generate Phase Modulation

For reference, Figure 2.2.1 shows the schematic for the phase modulator that I used to make the comparisons. This was created in Simetrix [Ref 8] and is not too different from a real-life working circuit.

2.3 - Test Methodology

The next stage is to determine a test method that will show the effect - this is not a simple task, because we know that any shift will be small, and is invariably accompanied by amplitude modulation and intermodulation. Because of this, a spectrum analyser is useless, as it will show intermodulation sidebands that will completely mask any frequency variation. The instrument of choice is an oscilloscope, because that is capable of displaying information in the time domain, where the problem occurs.

If the (velocity based) Doppler theory is correct, we should expect to see a set of waveforms similar to that shown in Figure 2.3, where a low frequency tone is mixed with a high frequency tone, and fed to a loudspeaker. This signal is also fed to one channel of the oscilloscope, and the other is used to display the audio waveform reproduced by the loudspeaker, and picked up by a microphone. The gains of the two are carefully matched so that the waveforms are perfectly overlayed - any discrepancy should be visible. This can only be done using a digital storage oscilloscope, because the waveform must be captured and expanded so that any shift is visible.

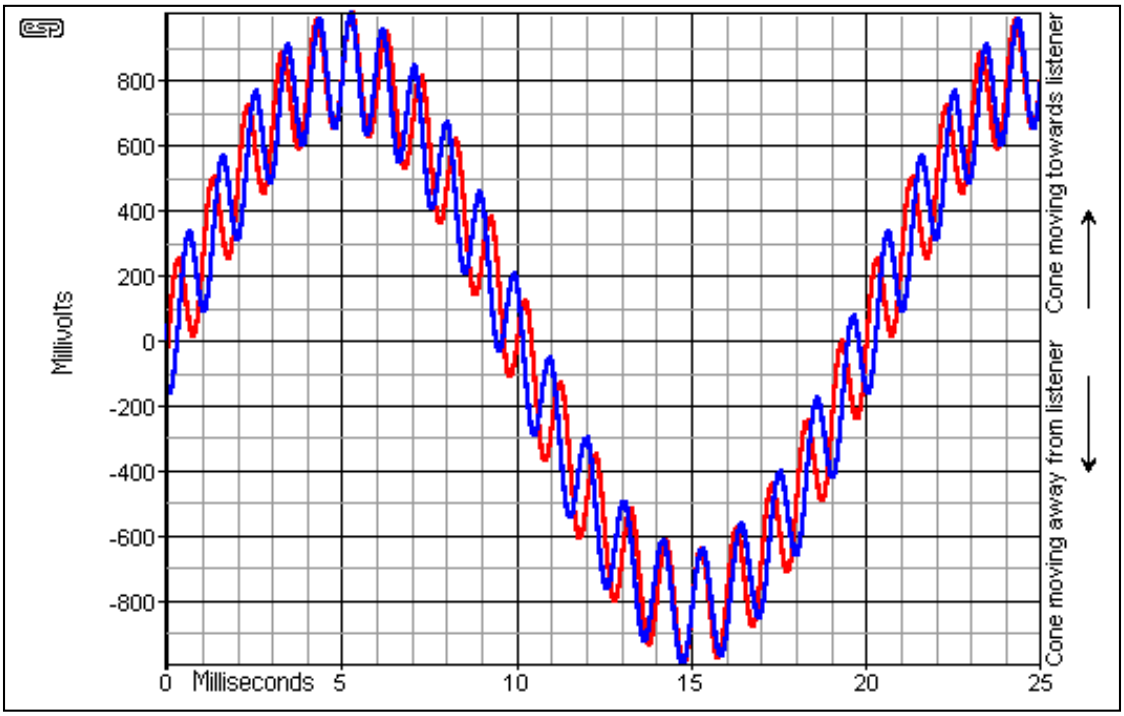


Figure 2.3 - Expected Oscilloscope Display for Doppler Modulation of the High Frequency

The red trace shows the unmodulated carrier riding on the LF waveform (i.e. the electrical signal sent to the loudspeaker), and the blue trace shows what we should expect to see on the oscilloscope from the audio signal picked up by the microphone. Note that the frequency deviation has been exaggerated for clarity - if it were that easy to see there would be no argument. It is quite apparent that with FM, the modulation is at its maximum at the LF zero crossing (maximum acceleration), and is at minimum (signals in phase) at the LF signal peaks.

If the shift is a genuine Doppler shift, then we will see the signal compressed or stretched at the zero crossing points of the LF waveform, and at the peaks of the LF waveform the HF signal should be in phase, since LF cone movement is effectively zero (for a moment in time, at least).

On the other hand, if the shift is one of phase, then the maximum variation between electrical and acoustical signals will occur at the *peak* of the LF waveform (positive or negative), as shown in Figure 2.4 below. This stage of theoretical analysis is very important - it is far easier to figure out what one should look for in advance than it is to try to figure out what one is looking at with no background.

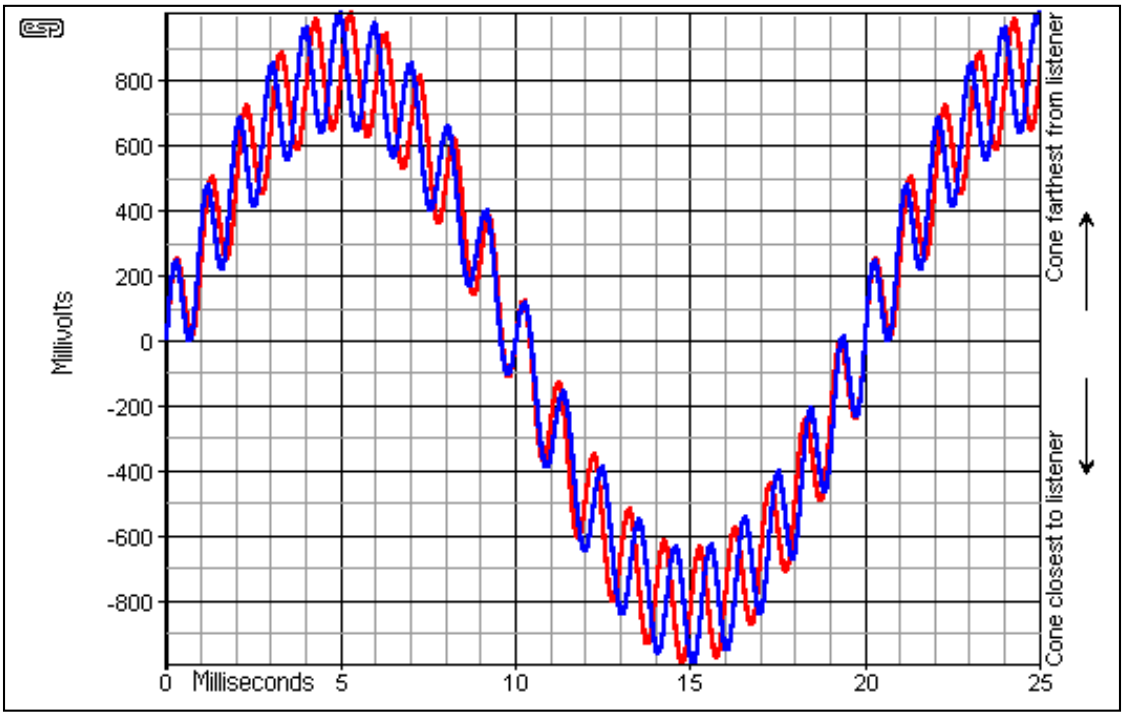


Figure 2.4 - Expected Oscilloscope Display for Phase Modulation of the High Frequency

The red trace again shows the unmodulated carrier riding on the LF waveform (the electrical signal), and as before the blue trace shows what we should expect to see on the oscilloscope from the audio signal. Note that the phase modulation too has been exaggerated for clarity. You can see easily that for phase modulation, the maximum variation is at the LF waveform peaks, where the distance from the cone's rest position is at its greatest. There is no modulation at the zero position, despite this being the point of maximum velocity for the LF signal.

This alone should start to give the astute reader a hint - the cone is *definitely* moving closer to and farther away from the microphone, so there must be a variation in the relative phase. After all, the microphone position had to be adjusted to ensure that the acoustic and electrical high frequency signals were in phase during the setup phase. If you are anything like me, at this point you'll be saying '*Of course! This **must** be so*.' (or words to that effect).

Figure 2.5 shows a schematic of the test setup. Several things are critical to get a usable result, and these include the following ...

- Woofer mounting - the woofer was not in an enclosure, to maximise the excursion. At the same time, it must be firmly held to prevent any movement that could give false readings.
- Microphone placement - the positioning of the mic is very important, since it is essential to have the electrical and acoustic signals in phase when the cone is at its rest position.
- Signal frequencies - the woofer frequency was chosen to give high excursion (close to resonance), and was adjusted slightly to alter the phase of the electrical vs. acoustic signal. The upper frequency was selected to be as high as the woofer could reproduce, so that any shift would affect as many cycles of the HF waveform as possible, making it easier to see any variance.
- Signal levels - just as important as phase, the electrical and acoustic signal levels have to be as close as possible. This allows a direct comparison with minimum error. It also allows measurement (or least a visual impression) of amplitude modulation, as this is useful for the final analysis.
- Setup position - the experiment must be set up so that it is far enough from the signal generating and measuring equipment that physical movement of the person (or persons) conducting the test don't affect the phase relationships. An experiment of this nature relies on accurate phase relationships, and these are

easily disturbed by movement near the experiment itself.

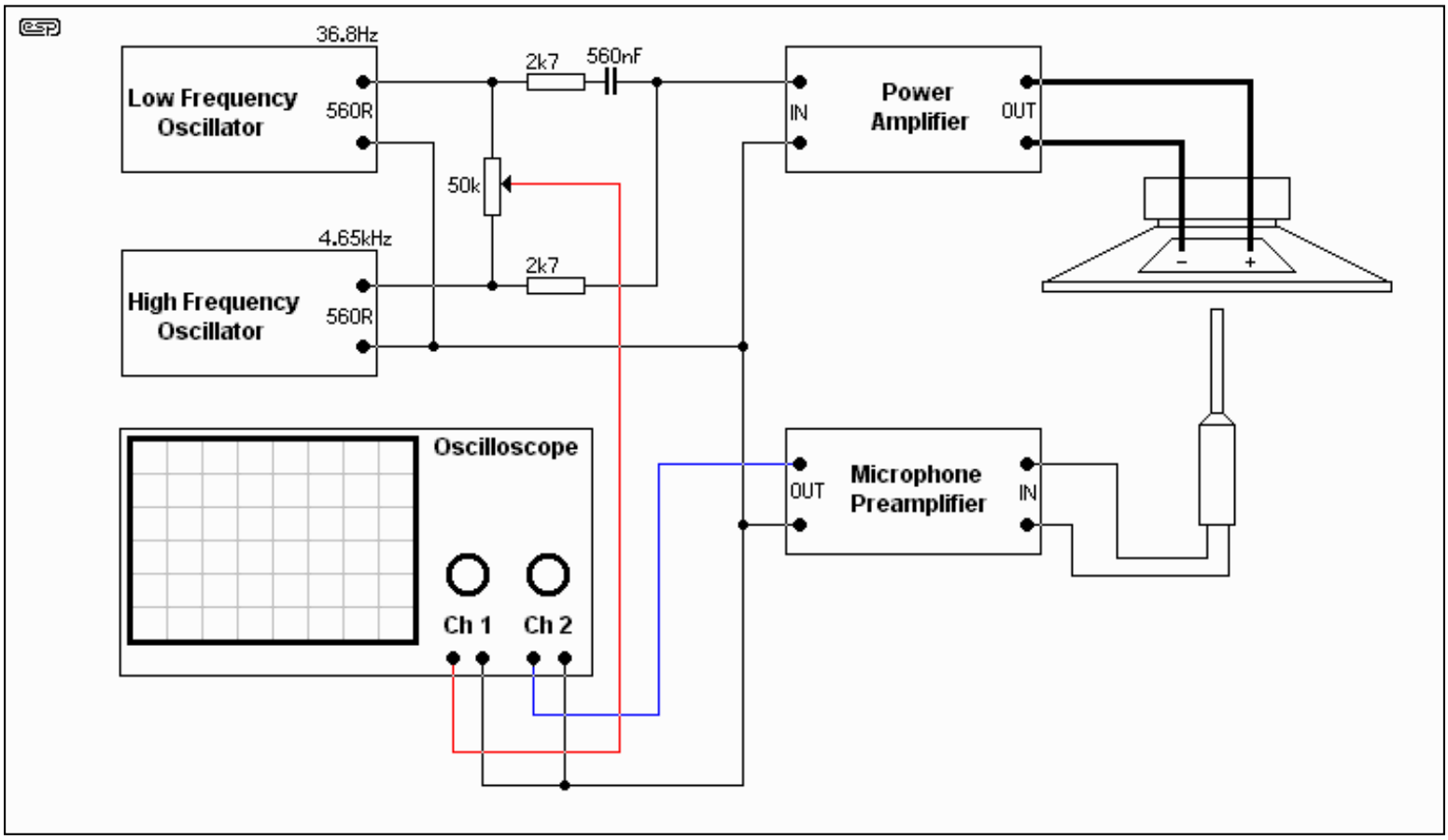


Figure 2.5 - Test Setup Circuit Diagram

The 50k pot allows the exact 'blend' of low and high frequency signals for the reference trace (Channel 1 on the oscilloscope). By varying the levels from the two oscillators, the loudspeaker drive signal is changed to suit requirements, and the mic preamp also has a gain control so the acoustic signal level can be exactly matched to the reference - in both phase and amplitude. The 560nF cap was needed to rotate the LF phase enough to be able to use a sensible frequency.

It is important to understand the exact test setup used to perform the test. First, the speaker is excited by the HF tone only, and the microphone gain and physical position adjusted so that the electrical and acoustic signals are exactly in phase. This is the reference signal, and it is imperative that this is done very accurately, or the end result will be nonsense. The HF signal is then removed, and the LF signal is applied - because of the use of a close microphone, there will be little LF phase shift ... but the speaker must be operated at (or near) resonance, or the reactive load will cause the electrical and acoustic waveforms to be out of phase.

This can be used to our advantage - by changing the low frequency slightly, we can advance or retard the phase just enough to make sure the signals are again perfectly aligned. By operating the woofer near resonance, the best relationship between cone excursion and power is possible - setup and adjustment take some time, and a smoking woofer voicecoil is not helpful. Likewise, for the high frequency signal, it is advantageous to use the highest frequency possible, so that any shift that occurs is maximised, and becomes easier to see on the oscilloscope. Because of the relatively short wavelength, mic positioning will correct all phasing errors (the voicecoil is inductive, and causes phase shift at high frequencies). Also note that the woofer will not be a linear piston at such a high frequency, so measured phase variations may not agree 100% with those calculated.

One might imagine that the test would be easier if the low frequency were filtered out, since that allows a better resolution of the HF, but then we would lose the essential information that shows us where the LF zero crossing

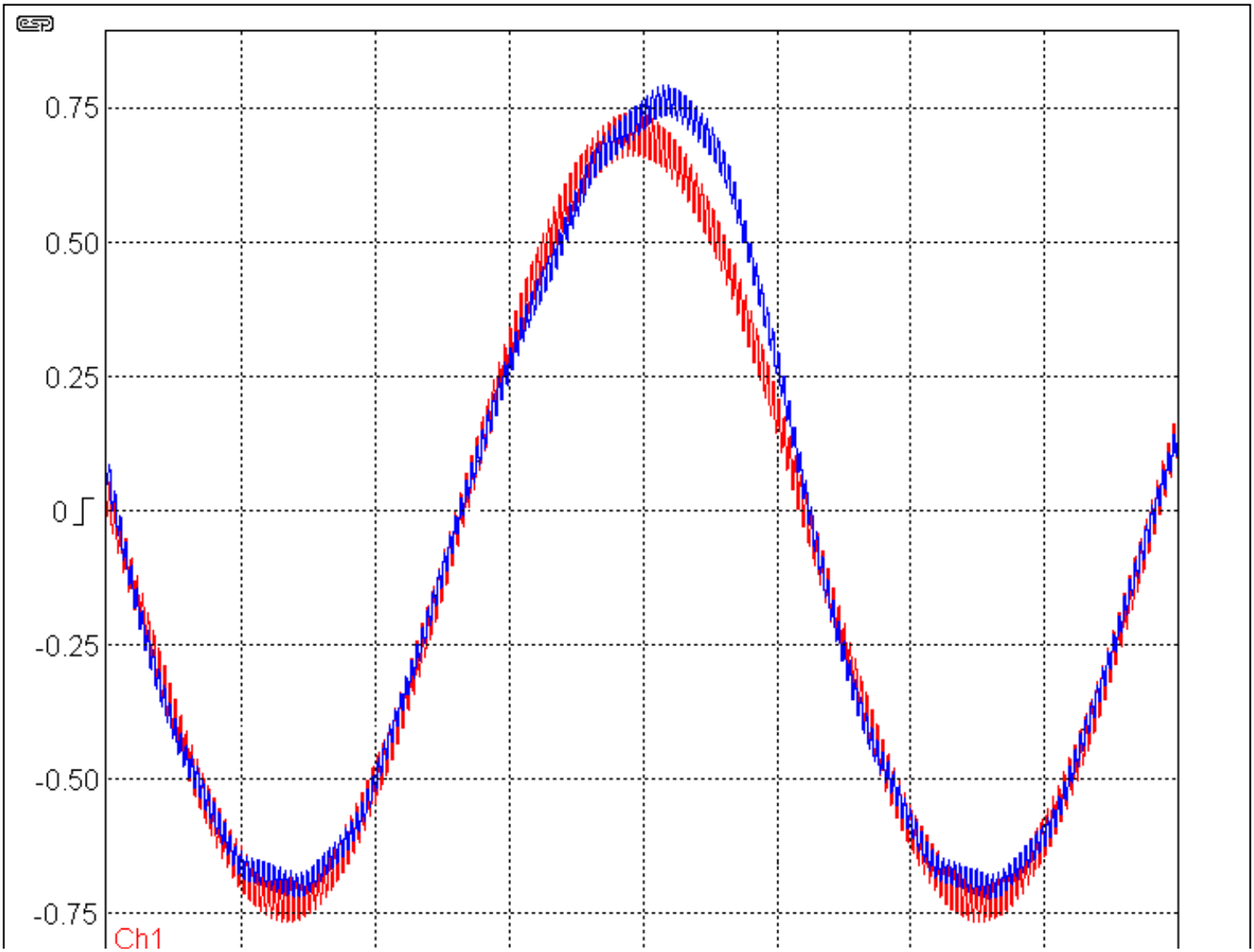
points and waveform peaks are - these are the parts of the waveform where we need to see what is happening, and we *need* the LF data.

As a point of interest, it must be stated that the microphone's diaphragm does not in any way influence the test results, since it is an electret capacitor mic. These microphones respond to pressure, and diaphragm movement is virtually nil - it is probably around the same as the movement of the human ear-drum, so any thoughts that this may introduce an 'anti-Doppler' effect are simply untrue. In comparison to cone movement, the travel is infinitesimal, and can safely be ignored.

Not so the woofer however - it proved necessary to wedge the woofer (sitting in a plastic crate filled with fibreglass) with a piece of wood from the ceiling to stop it from walking around the workshop floor! With an applied frequency of 36.8Hz (required to obtain the correct phase relationship and sufficient travel), there was no way it wanted to stay put. Even the high power 'digital' amp I used proved to be a problem as it kept cutting out, and I finished up using one of my P101 MOSFET amps - this was more than up to the task, and didn't get above warm, despite the load.

3.0 - Test Results

This is almost an anti-climax, since it should be fairly obvious by now what we will see. The results do not show a great deal of phase shift, but it *is* visible. It was necessary to adjust the signal levels slightly to get the best alignment, since as you can see in Figure 3.1 there is significant LF waveform distortion.



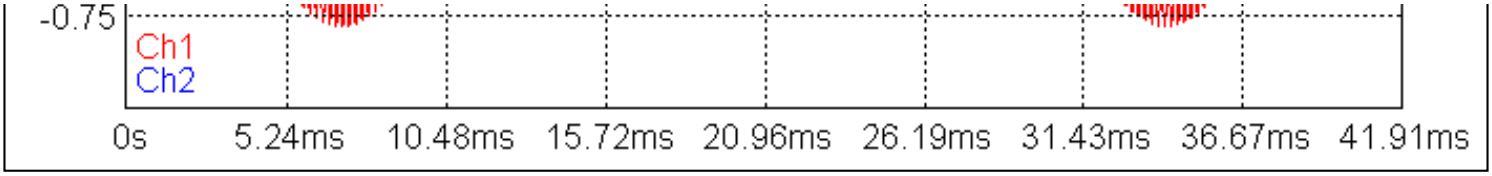


Figure 3.1 - The Composite Waveforms - Signal and Acoustic

As with all the following plots, the electrical signal is in red, and the acoustic signal in blue. It is obvious that any phase shift in the acoustic signal is not visible in the above, so the following three graphs show the expanded waveform at different points. The oscilloscope used is a Tiepie PC based instrument, and for this work was operated at a sample rate of 2.5Ms/s (2.5 million samples/ second) at 12 bits resolution, and with a record size of 130,972 samples. It exceeds the resolution available from my normal bench oscilloscope by a wide margin.

Amplitude modulation is easily seen, and was also quite audible during the test - even with earmuffs. Look at the peaks of the waveform, and you can see that there is a big difference between the electrical and acoustic signal HF amplitudes - in a perfect woofer, these would be identical. Likewise, the low frequency signal would not be distorted (and this was from a high excursion subwoofer). Although the waveform looks a little like the power amp was clipping, this is not the case - the distortion seen is from the loudspeaker itself.

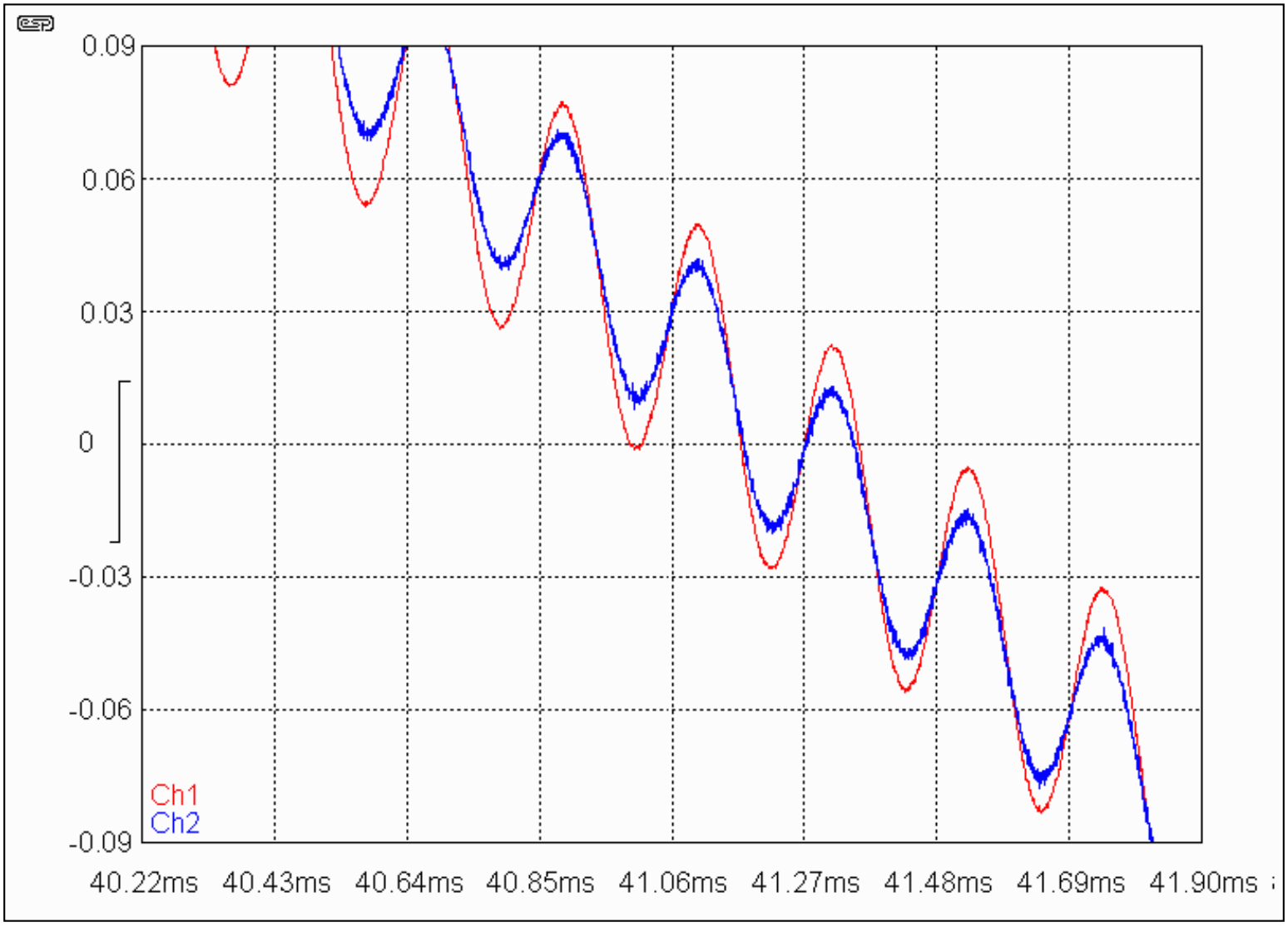


Figure 3.2 - Signals at LF Zero Crossing

It is quite apparent that there is no phase shift at the LF zero-crossing point, even though this is the area where cone velocity is highest. While not really visible here (much higher resolution is needed to be able to measure it),

there *is* a change to the period of the reproduced waveform - unfortunately, it only amounts to a pixel at best in the chart, and is easily missed. If the periodic time of a waveform changes, so does its frequency. For a 1kHz signal (having a period of 1000us), we can expect to see a change of only a few microseconds in the period (typically less than 10us, or 1%). This is a challenge to measure, as should be obvious.

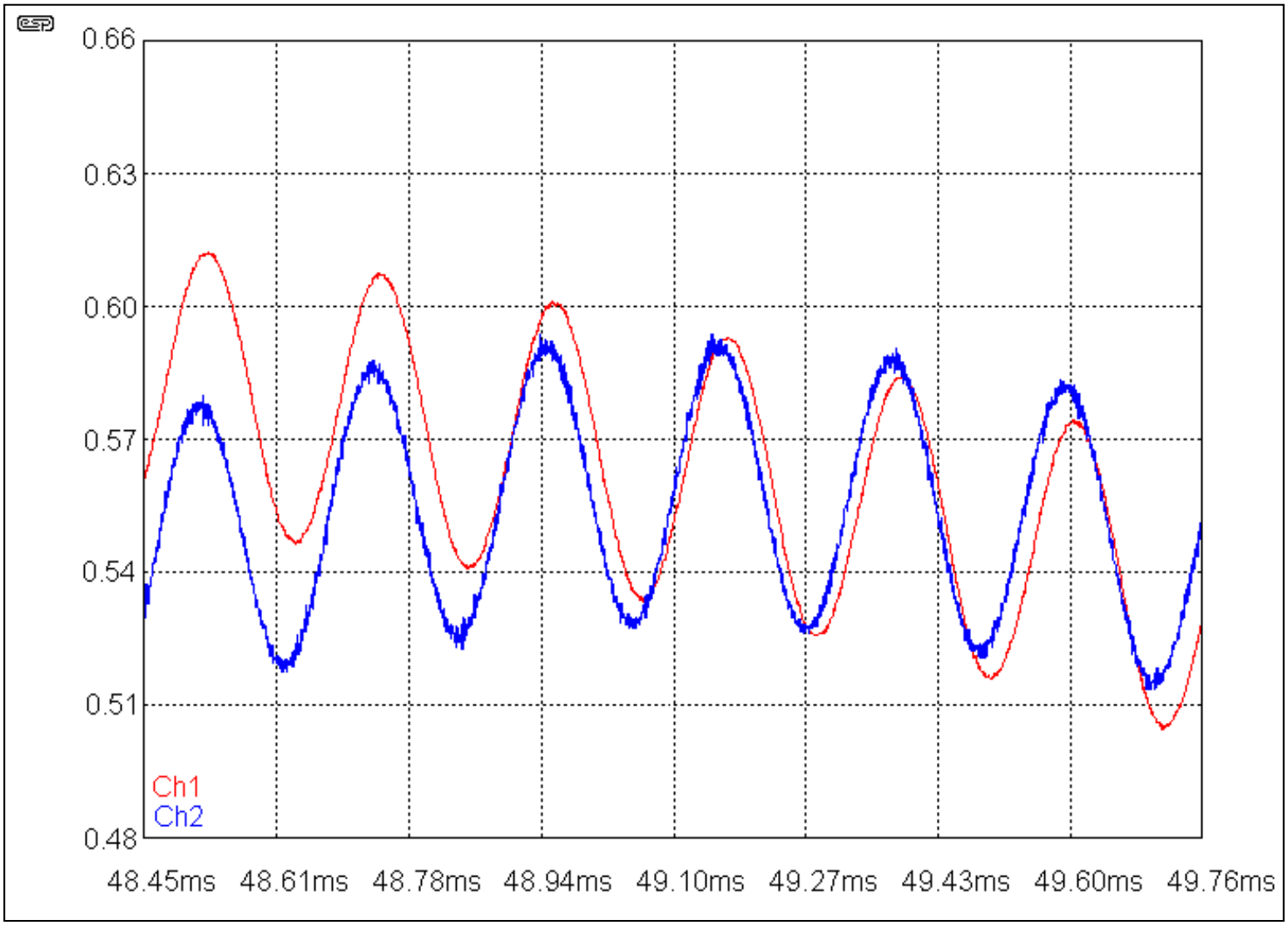


Figure 3.3 - Signals at LF Positive Peak

Here, we can see that the HF acoustic signal is arriving just before the electrical signal - the cone is closer to the microphone, so the signal arrives earlier. Look at the waveform peaks - not the zero-crossing points. The difference is discernible, but is still quite small. This is quite possibly the first time you have *ever* seen this effect, and it's all captured from a real test.

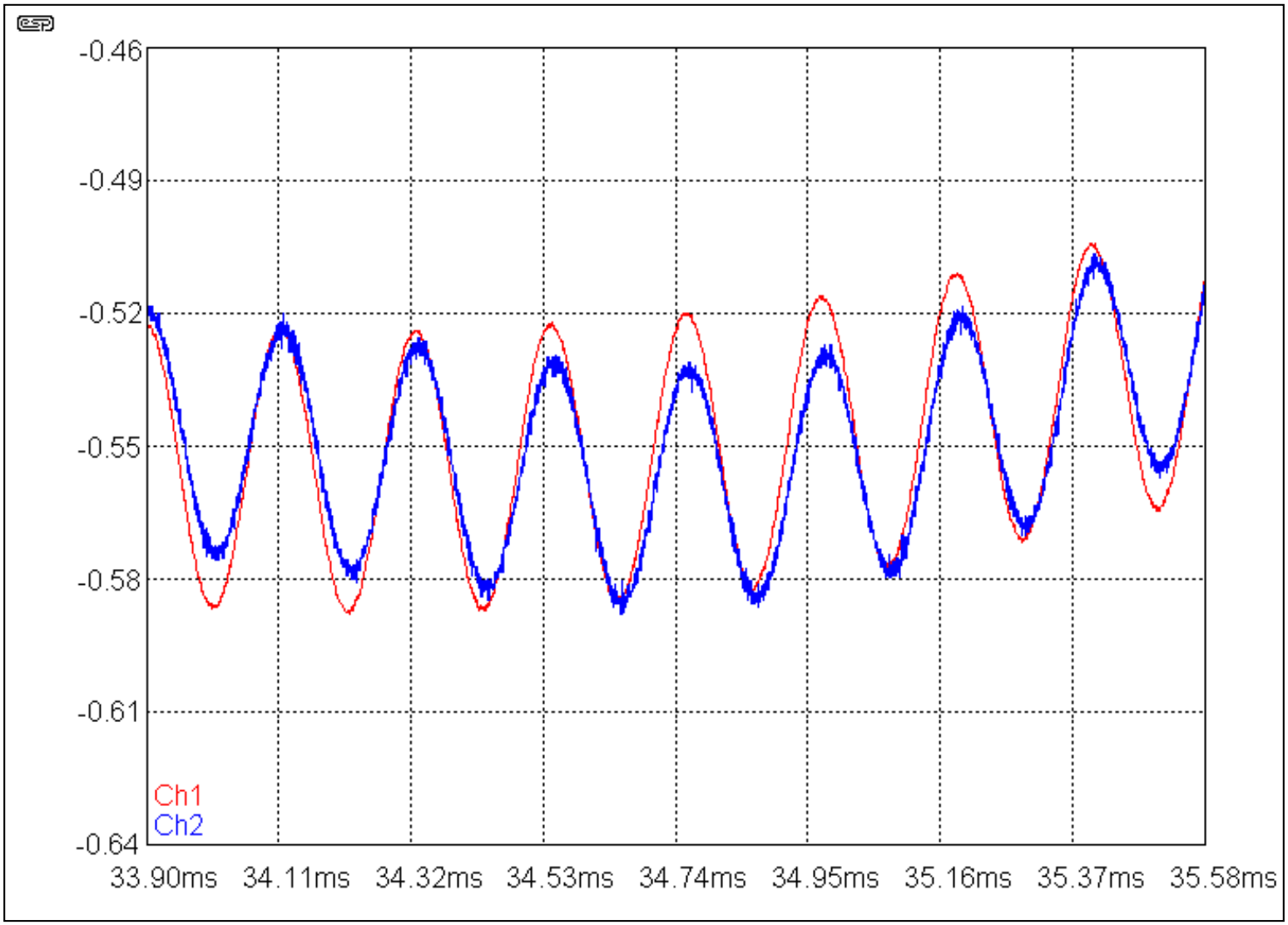


Figure 3.4 - Signals at LF Negative Peak

While not quite as pronounced as with the previous example, a small amount of shift is visible, since the cone is now further from the microphone than before. Even using a 4.65kHz signal, the wavelength is still just over 74mm, so the cone travel I was able to achieve without excessive distortion makes the variations smaller than desirable. The variation is there though, as predicted.

It is important to point out that the only manipulation of these graphs was to reverse them, because the microphone I used inverted the signal (so the positive and negative excursions were back to front). Otherwise, the graphs are as they were saved from the PC oscilloscope, but the text (inserted when the graph is saved) was removed to minimise file size.

4.0 - Conclusion

Upon testing, it is easily seen that the maximum phase shift of the HF signal occurs at the peak of the LF waveform. There is little or no discernible (phase) shift at the LF zero crossing, so the effect is phase shift, caused by the cone being closer (or further away) from the observer/listener at any point in time. Yes, there is an effective frequency shift, but (and although this is an apparently minor point, it very important), the shift is **not** caused by the Doppler effect per se (i.e. the 'conventional' velocity based interpretation), so cannot (or should not) be correctly called 'Doppler distortion'. Exactly the same phase shift can be seen simply by disconnecting the LF signal, and manually moving the cone or microphone (slowly) by the same amount, or by using DC or an

extremely low frequency to achieve the same result. This cannot cause the Doppler effect, as should be obvious. What it does is shift the phase, and if the phase shift is fast enough, then the phase modulation will 'create' frequency modulation - exactly as predicted.

It is fair to state that 'Doppler distortion' in loudspeakers exists, but the term is misleading.

It is equally fair to state that Phase Modulation in loudspeakers also exists, and is the correct definition of the effect.

Using the predictions that may now be obtained based on phase modulation, it can be demonstrated that the phase shift is independent of the LF modulation frequency, and depends only on the peak-to-peak amplitude. Nonetheless, the FM component of the phase modulation that *is* produced is a direct function of the LF modulation frequency. The two theories are not at odds with each other, but describe the same effect from a different perspective. The phase modulation model is more intuitive, and eliminates any of the arguments about the sound source moving through the medium etc. While it is unlikely that any of this will make the arguments go away, the procedures described here can be reproduced by anyone with the right equipment. Indeed, a complete test is not even needed to be able to see clearly that changing the mic position with respect to the sound source will change the phase - everything else can be calculated from there.

Further analysis reveals that the traditional predictive formulae that have been used work - they predict exactly the same frequency shift as is obtained by analysis of the phase modulation. What was not taken into consideration was the fact that the observed FM was really PM all along (a subtle difference). Since the peak cone travel was always used as part of the equation - as it must be to determine instantaneous velocity - the analysis turns out to be identical regardless of the method used. When the rate of change of phase is considered (where maximum rate of change implies maximum frequency shift), the peak frequency variation coincides with the maximum rate of change of phase, so the maximum frequency shift is also seen at this point.

> **Note:** By examining the phase model described herein, and upon realising that it is identical to the conventionally explained Doppler effect, it is now obvious that there is no requirement for the sound source to move *through* the medium at all. While this is a common argument by the non-believers, it fails to stand up to scrutiny. All that is required is that the sound source moves with respect to the listener - whether (or not) it moves through the medium is immaterial. While this may not be apparent from the traditional descriptions and explanations of the effect, the phase model makes the situation quite clear.

Very much unlike many (most?) of the other examinations of the subject, this test procedure was performed on a real speaker, and the results of these real-life tests are shown above. There is sufficient information in this paper to allow anyone else to duplicate the results, provided the necessary equipment is available. This is encouraged - the more people who understand the physics that cause the effect the better, and I would be delighted to hear from others who have applied this (or any other) test method to demonstrate the mechanism involved.

This is *not* merely a theoretical discussion of the effect, but contains the unadulterated results of the test procedure described. At the risk of offending those who believe that there are things in audio that no instrument other than well trained ears can detect, this is proof that a properly designed test method can be devised, and a theory thus proven or disproven.

The effect is very small (to the point of being virtually inaudible by itself), and is usually swamped (or masked if you prefer) by amplitude modulation and intermodulation distortion, so could be considered immaterial in any typical loudspeaker system. If people really want to describe the effect as a distortion - not an unreasonable assumption, since it does exist - then it should be re-named. The correct term is (in my opinion) Phase Modulation Distortion (PMD), and I suggest that the term 'Doppler distortion' be dropped from usage, since it is too easy for people to misinterpret.

Even PMD is possibly technically incorrect, since the effect is perfectly linear, and will occur in a perfect

loudspeaker. Regardless, it is probable that people will want to call it distortion. Linear or not, it still adds something to the signal that was not there in the first place, and it is not unreasonable to call that distortion.

Many readers will be glad to see that there is no mathematical proof of PMD, nor have I attempted to devise a formula to allow you to calculate the shift of any given frequency with a known LF cone displacement. So many others have already done this that there is no reason for me to add more on the topic.

Siegfried Linkwitz is one who has done exactly that - the mathematical proof (and equivalency) may be seen on his site, and his page now covers the topic from both the traditional and 'new' perspectives.

The methods that may be used to minimise PMD are exactly the same as those used to minimise intermodulation distortion, primarily, reduce the excursion of the mid-bass driver. This may be done by using a crossover and a subwoofer, or by choosing an alignment that reduces the low frequency excursion to the absolute minimum. Naturally, a 3-way system will outperform a 2-way in this respect, since the midrange driver's excursion will be minimal with no bass content.

Finally, it should be understood that this is a purely physical phenomenon, and short of extensive DSP (Digital Signal Processing) nothing can be done to prevent it if a driver is expected to handle bass and high frequencies simultaneously. While DSP is quite capable of applying a delay correction to remove the effect, the frequency shift is so small that the benefits are of dubious value.

If you *really* want to hear Doppler shift in a loudspeaker, get hold of a Leslie rotating speaker cabinet [Ref 9]. The high frequency rotating horn (in particular) satisfies all the criteria for Doppler effect - the sound source moves through the medium, and has sufficient diameter (and speed) to create a spectacular sound effect. There is also considerable amplitude modulation, and interesting phase effects as well. That you cannot expect to get this from a conventional loudspeaker is readily apparent, since Don Leslie would have not had to go to all that trouble if it were otherwise.

5.0 - References

1	Paul Klipsch	CE Hall of Fame
2	Issues in loudspeaker design - 1 (section 'J')	Siegfried Linkwitz
3	Doppler shifts in loudspeaker. Fact or fiction?	John Kreskovsky
4	HyperPhysics (Georgia State University)	Very useful and detailed explanation of the Doppler effect
5	HyperPhysics	Effect of Loudness Changes on Perceived Pitch
6	http://www.tmeg.com/esp/e_modulation/modulation.htm Types of Modulation (website gone)	By Dennis J. Ramsey
6a	Types of Modulation	ICT Technologies
7	Phase Modulation	Integrated Publishing
8	Simetrix	One of the best simulators I have ever used.
9	Unearthing ... The Leslie Cabinet	Clifford A. Henricksen - Community Light & Sound (1981)
10	Piston Vibrating in a Tube	Art Ludwig

Other useful reading material

1	Doppler Distortion	Piston Vibrating in a Tube Including the Effect of Excursion - Art Ludwig
2	Sound	Clint Sprott - Physics Department, University of Wisconsin
3	Hearing, physiological & psychological aspects	Sampo Syreeni, Finland
4	Christian Andreas Doppler	Biography of the man himself

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 20 August 2004

The Audio Pages

Elliott Sound Products **Effects Of Source Impedance on Loudspeakers**

Copyright © 2001 - Rod Elliott (ESP)
Page Created 22 Jul 2001

Articles Index

Main Index

Contents

- Introduction
- Conventional Voltage Drive
- Current Drive (+3 Ohms)
- Current Drive (+32 Ohms)
- Negative Impedance (-3 Ohms)
- Conclusion

Introduction

This is not the final word on this topic, and I am sure that others with more (and better) equipment than I have could do much more. An hour or so of testing is hardly a fully scientific study, but the results agree with expectations for voltage and current drive - what came as a surprise was just how nasty negative impedance really is.

This article was prompted by a reader who did some simulations of a loudspeaker having no voice coil resistance. Of course, this is not possible, but driving with negative impedance can theoretically accomplish the same thing.

Negative impedance has been around for quite some time (probably about 50 years, maybe more), but I suspect that I have one of the few amplifiers around that can be preset to any impedance between -16 Ohms and +32 ohms. To this end, it was reasonably easy to test, and I used 42Hz as the test frequency - being representative of a typical bass signal (open "E" on a bass guitar). Just to make matters as bad as they could be, this is also very close to the resonant frequency of the test loudspeaker and enclosure.

I used a sound level meter at close range (100mm) connected to my oscilloscope to see the actual loudspeaker waveform. The speaker is a 4 Ohm 300mm Cerwin Vega subwoofer, in a 28 litre sealed box. This box has been used many times as a test system for the Electronically Assisted Subwoofer (see EAS project), described on these pages.

During these tests I made no real attempt to use zero crossing switching, since this was difficult for a quick test (my tone burst generator is on the blink and needs repair - one day). The lower trace is the output from the signal generator, and switching could - and did - take place whenever I switched the output.

The upper trace of each oscilloscope screen photo is the signal taken directly from my sound level meter, and the lower trace is the input signal. I made no real attempt to maintain any specific level, since it was immediately apparent that this made little difference.

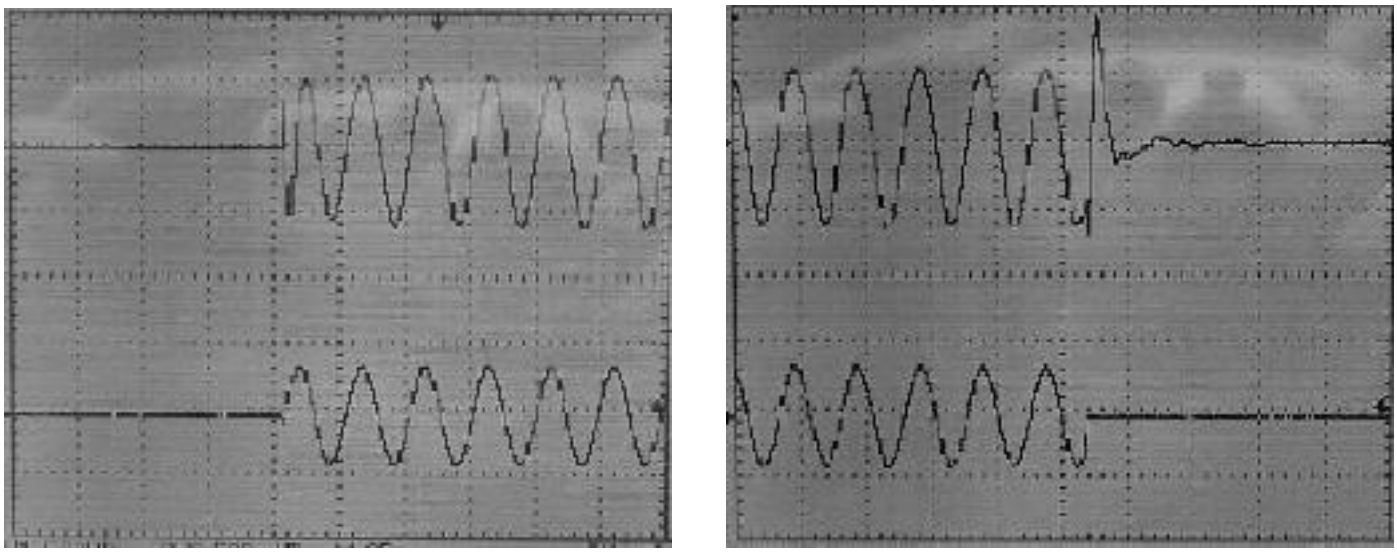
The waveforms displayed are exactly as measured, and the phase and timings were not tampered with in any way. The input signal was not bandwidth limited, nor was the amplifier output - the test results are therefore "worst case".

Transients are generated at the moment of switching - it is predominantly these that are of interest - a photo of an oscilloscope trace of a 42Hz waveform is fundamentally boring, whatever the source. The oscilloscope traces shown were captured with my digital camera - less than ideal, but it works (other than the ghostly image of the camera reflected in the front of the display).

Conventional Voltage Drive

Voltage drive is the most common way to drive loudspeakers. In theory, a "perfect" voltage amplifier has zero ohms output impedance, but this is never achieved in practice. Output impedances of 0.1 ohm or less are not uncommon at low frequencies,

representing a "damping factor" of 80 or more.

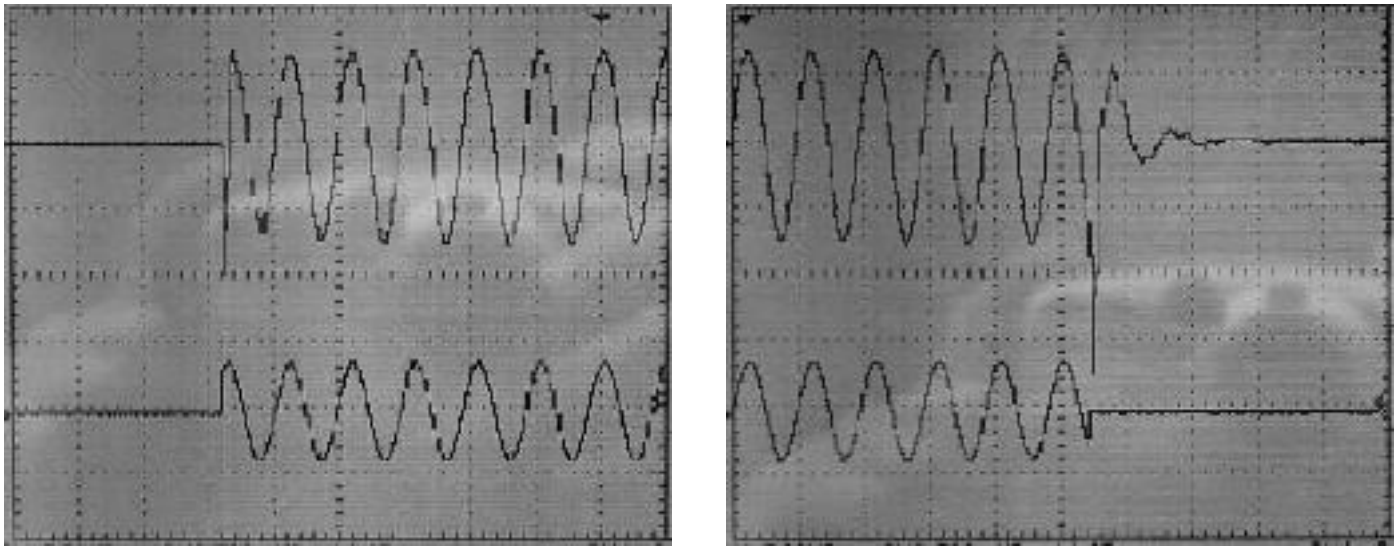


Voltage Drive - Z = <0.5 Ohms

One might say that this is text book performance. There is evidence of the switching transient, but overall this combination is well behaved. This is the most common connection for loudspeakers, and seems to be about as good as it gets. Overhang (ringing after the signal has gone) is almost non-existent.

Current Drive at +3 Ohms

With a pure current drive signal, the input determines the current that flows through a loudspeaker (or other) load. This is rarely used in practice, as it is unsuitable for driving speakers. Using a defined impedance means that the amplifier's output is dependent on the load impedance at any given frequency, but in a way that is predictable. This can be used to advantage, but is most commonly (and IMHO incorrectly) frowned upon. Generally, I would suggest that the amplifier's output impedance should be no greater than the speaker impedance (and preferably less).

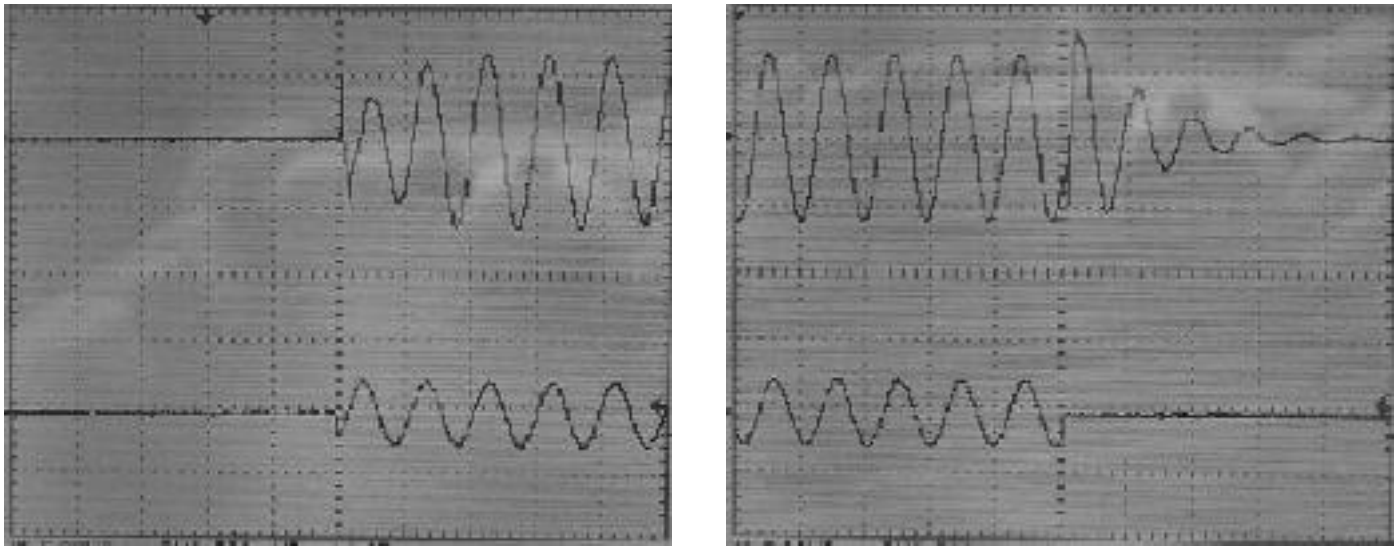


Current Drive - Z = +3 Ohms

The switching transient is still in evidence, as is to be expected - the big spike is a result of the switching point, and is not unique to current drive. The attack is almost perfect, but a slight ringing is apparent when signal is removed. Although this is quite visible, it is inaudible (at least in my workshop), and since bass signals don't normally stop partway through a cycle, it is unlikely to be audible in a typical room, whose response will be far worse than the speaker.

Current Drive at +32 Ohms

This is the same as the previous definition, but in this example, the output impedance was increased to 32 ohms - this is 8 times the quoted impedance of the speaker I used.



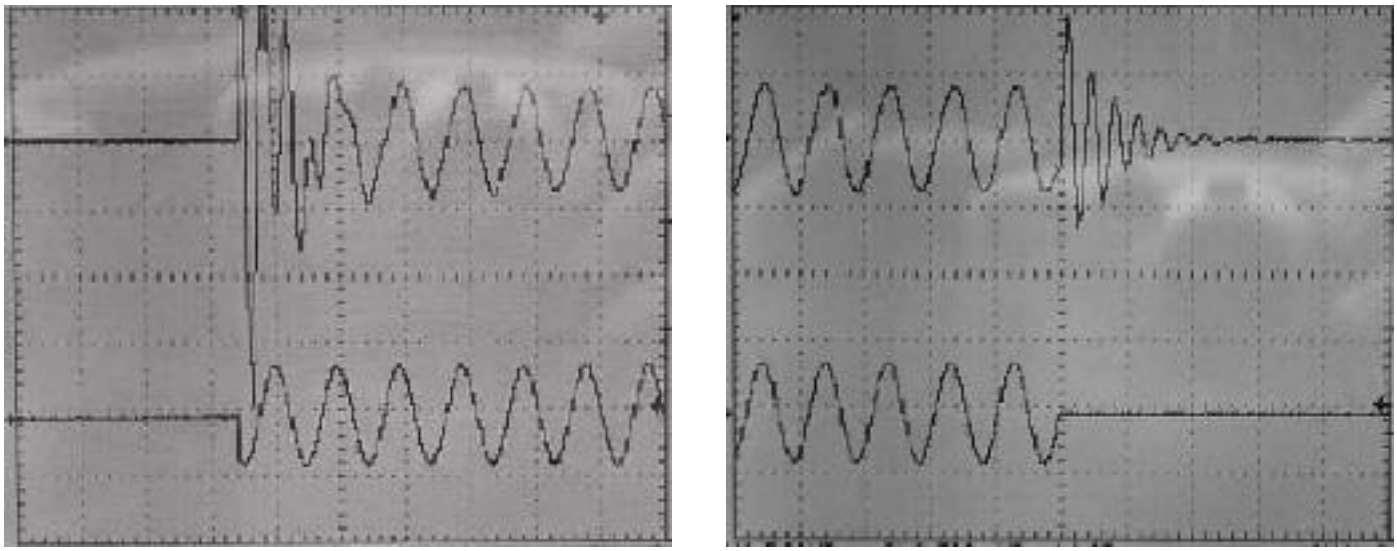
Current Drive - Z = +32 Ohms

This is an extreme impedance, and would never be used in practice. There is a slow attack, with the signal building up to the steady state level. Considerable ringing is apparent when the signal is removed. This is both visible and (in my workshop, only just) audible. This is not a good combination for hi-fi, but may be quite acceptable for a guitar or bass amp (for example), as the colouration simply adds timbre to the sound of the instrument. The switching transient is almost non-existent, as the energy is converted into underdamped cone movement instead.

This is "slow" bass, with a woolliness that may be attributed to some valve amplifiers - especially those with no (or very little) feedback. There are many people who like the sound like this, and this is perfectly OK. Each to his own, and all that. In general, the impedance is too high in my opinion.

Negative Impedance Drive at -3 Ohms

Negative impedance does exactly what it implies - when the load is increased (with a lower impedance), the signal applied to the load increases (i.e. the exact opposite of what normally happens). This results in an intrinsically unstable system, and great care must be taken to prevent the creation of an oscillator.



Negative Impedance Drive - Z = -3 Ohms

The performance of this combination is completely unacceptable in every sense of the term. This combination could never be used in practice, for any reason. There is a vicious attack, with the signal doing something at a frequency completely unrelated to the input signal - unrelated in any way that I can determine, at least. Considerable ringing is apparent (again at an unrelated frequency) when the signal is removed. This is highly visible and audible, and the sound of the attack and decay is grossly inferior to

any other combination. The others (using positive impedance) have some character, but it is related to the signal, and makes some sort of musical sense.

To add insult to injury, the amp decided to become an oscillator (at about 15Hz) on a few occasions, this being one of the risks of negative impedance. It subsided, but was a little disconcerting. I doubt that this is high on the "desirable" list for most audiophiles.

Reducing the negative impedance to -1.5 Ohms stopped the oscillation entirely, but the effect on the attack and decay performance was merely a minor reduction of the many artefacts created, and was still quite unacceptable.

Conclusion

As I mentioned in one or more articles on these pages, I have never found a cone speaker that "likes" negative impedance. Horn compression drivers seem quite happy and this is a test that I may now be forced to perform to find out what it does, and how it affects the sound (other than an apparent improvement in bandwidth).

The net result of these tests indicate what has long been known - that voltage drive remains the best choice for most loudspeakers, regardless of some desirable characteristics of defined current drive. With additional internal damping, moderate current drive can modify the Qts of the loudspeaker driver, and allow an enclosure design that is (typically) larger than optimal for flattest amplitude (Butterworth) response.

There was recently an article in Electronics World, where it was claimed that current drive helps to compensate for the resonant peak and voice coil inductance in a driver. While this is undoubtedly true (I've been doing it for many, many years), the designer must be very careful to ensure that the enclosure is particularly well damped, to avoid "overhang" when the signal is removed. This will typically manifest itself as poorly defined bass, particularly on transients, since the speaker must rely primarily on its own devices for damping.

Overall, this exercise has simply proven that which has already been proven by so many before me ...

- Voltage drive is the best choice in the vast majority of cases
- Moderate values of positive impedance can be useful - if the enclosure is particularly well damped
- Negative impedance is loathsome, and should be avoided

For further reading on the subject of modifying amplifier impedance, see the Impedance article, and Project 56 (Variable Impedance) shows you how to accomplish this.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 22 Jul 2001

The Audio Pages

Elliott Sound Products **Linkwitz Transform Circuit**

Copyright © 2001 - Jeremy Wolf
Additional Material by Gareth Abrey
(Edited by Rod Elliott - ESP)
Updated 20 Jan 2002 *

Return to Articles Index

Back to Home Page

Contents

- Introduction
- How The Linkwitz Transform Circuit Works
- An Example System
- Compromises
  - Power Compression
- Woofer Selection Guidelines
- Power Distribution
- Editor's Notes
- Special Thanks
- Update

Introduction

From the editor ...

Project 71 has been a very popular project, and with good reason. Unfortunately, most people who have built it, don't actually know how it works. Jeremy Wolf wrote this article with some considerable consultation from Siegfried Linkwitz and a small amount from

me. Since the Linkwitz transform circuit seems so mysterious (which it is), it is assumed by many readers of The Audio Pages (and others) that it must be complex. The mathematics certainly are, but the principle is not, as Jeremy explains.

There are caveats (of course) and compromises (which are a base requirement in every speaker ever made), but this article explains the overall benefit. I know the benefits well - a sealed equalised subwoofer will simply wipe the floor with anything else, provided that you have done your homework. Now, you don't even have to do that, since Jeremy has done it for you.

How The Linkwitz Transform Circuit Works

Q What is a Linkwitz Transform?

A The Linkwitz transform was developed by Siegfried Linkwitz. It allows you to take a driver in a sealed enclosure that has an Fc and Qtc for that box and lets you "transform" or simulate a new Fc and Qtc for that driver in that box.

Q How does it work?

A Magic

Q No seriously, how does it work?

A It works by creating a precise equalisation curve to compensate for any peaking or rolling off that the driver is encountering. By doing this, a new Fc and Qtc can be assigned to that system.

Actually it is a simulation because you cannot physically change the Fc and Qtc of a closed system without actually making the box bigger or smaller or putting stuffing in the box and making it appear bigger. The following is an example of the above explanation with graphs to help clarify what is being said.

An Example System

Ok, let's assume that you have a driver with the following specifications.

Fs = 33.5 Hz
Qts = 0.75
Vas = 46.1 litres (1.647 ft³)

Now we are going to put this driver in a 38 litre (1.378 ft³) box. This will net us a Qtc of 1.0 This speaker in this box will yield the following results (these data can all be obtained from the Linkwitz Transform spreadsheet on the downloads page, courtesy of True Audio) ..

Fc = 49.6 Hz
F3 = 39.0 Hz
Qtc = 1.000
Vb = 38 litres

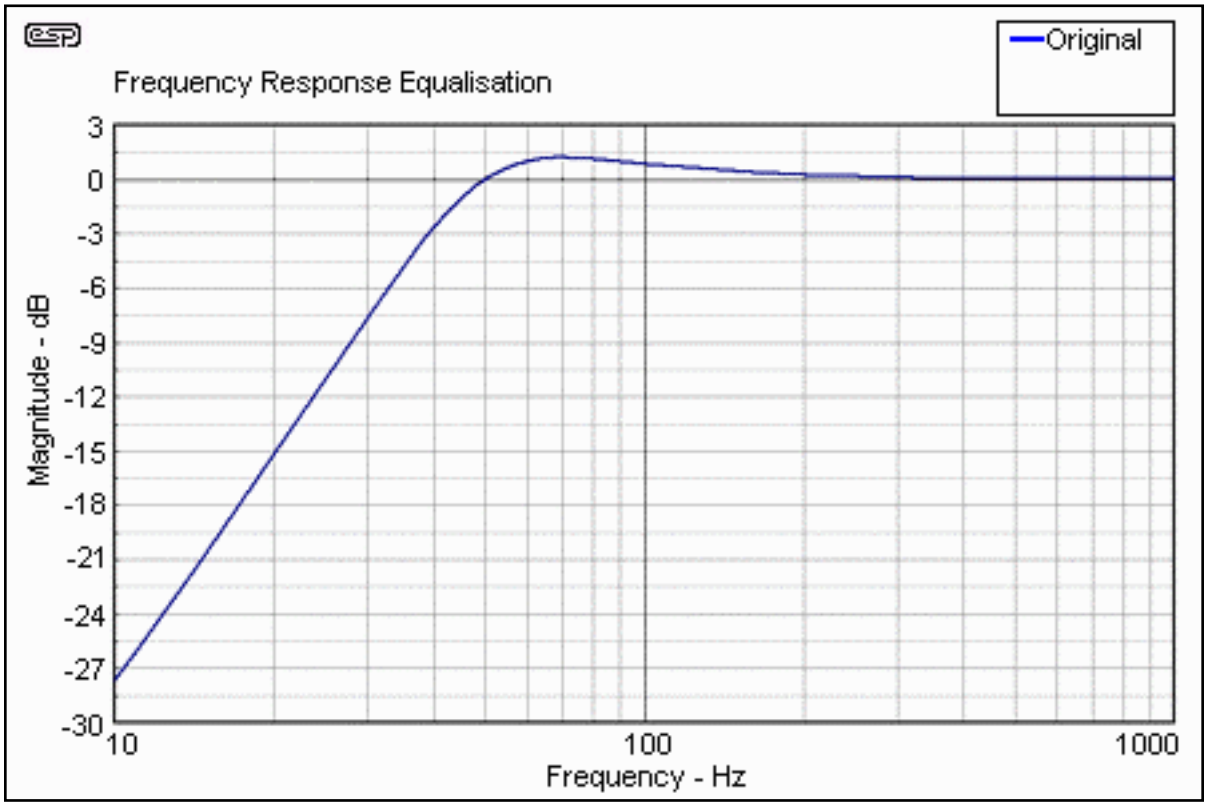


Figure 1 - Unequalised Speaker in 38 Litre Box

If you notice in this graph, there is a 1.2 dB hump before the roll off occurs. While this response is ok for your average listener, we are not satisfied with it and want to change the response. Wouldn't it be nice if we could simply move the Fc down to say 20 Hz and have the Qtc = 0.707 to get a response that looked more like this ...

Fc = 20.0 Hz
F3 = 20.0 Hz
Qtc = 0.707

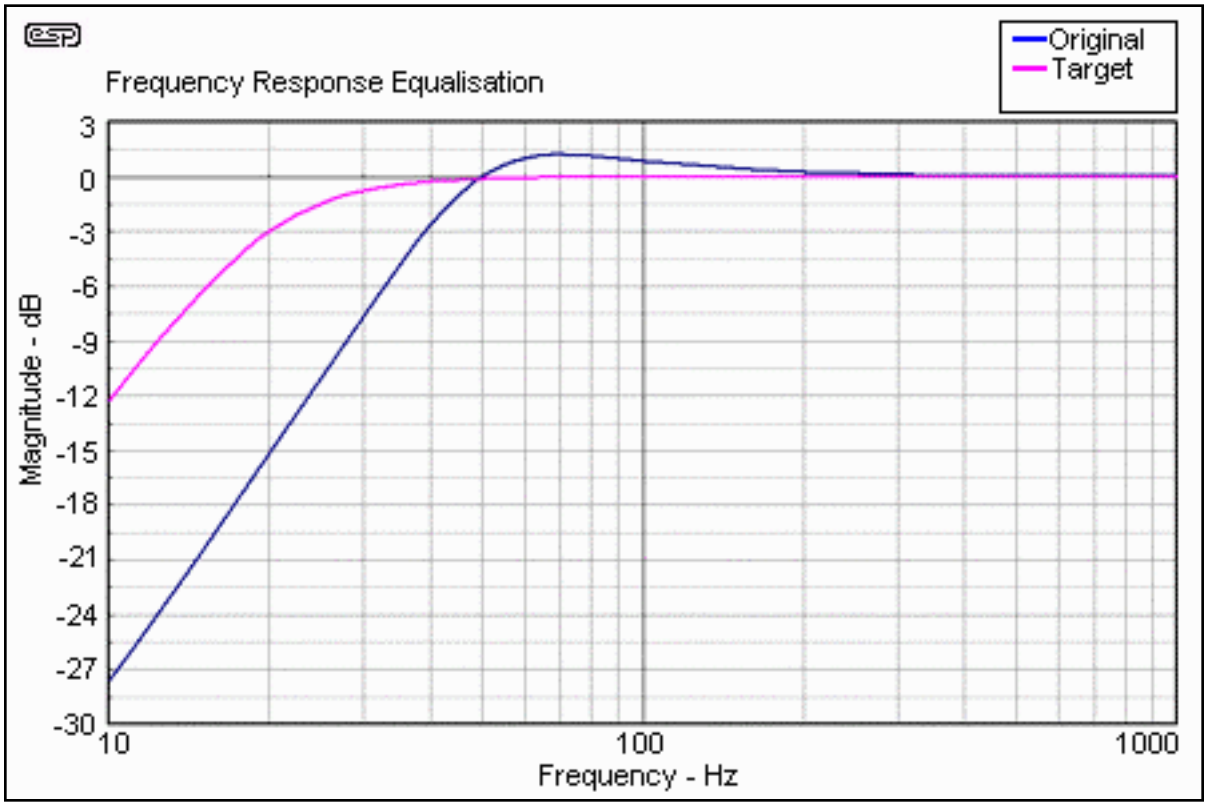


Figure 2 - Equalisation (Transformation) to Qtc = 0.707

As you can see, the magenta line is the newly transformed response. It is much more suited for low frequency reproduction after being transformed. There will now be a much greater output in the lower octaves.

And to show that the Fc is actually down at 20 Hz, here is a response of the same driver in that same box with after being transformed to a Qtc of 1.00 and Fc of 20 Hz. The slight peak causes the response to be 0 dB at 20 Hz, and this response is easily created by the Linkwitz transform spreadsheet if desired.

Fc = 20.0 Hz
F3 = 20.0 Hz
Qtc = 1.000

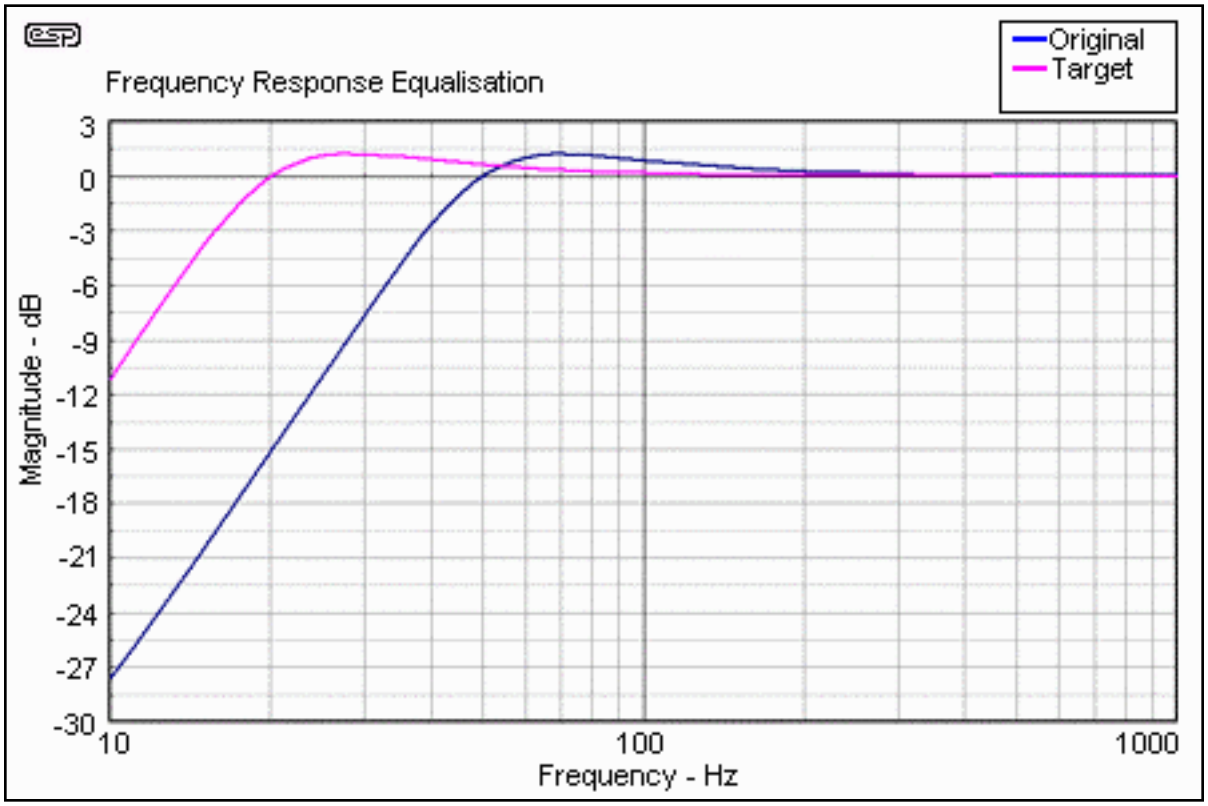


Figure 3 - Equalisation (Transformation) to Qtc = 1.0

Getting back to our desired response of Qtc .707 and Fc of 20 Hz. The graph in Figure 4 shows the compensation that the Linkwitz transform is using in order to flatten out the response curve of our driver.

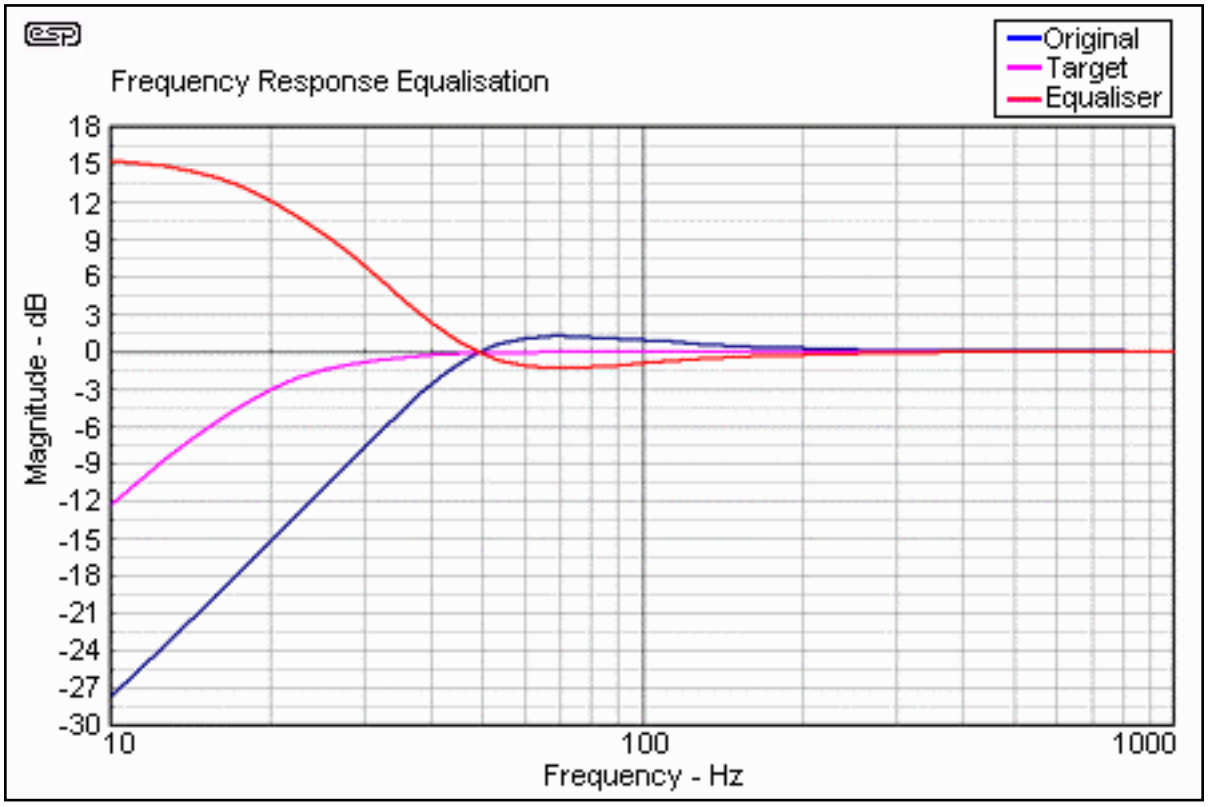


Figure 4 - Equalisation Applied by Linkwitz Transform Circuit

The blue line represents the original driver with a Qtc of 1.0 and an Fc 49.6 Hz, F3 of 39 Hz. The red line represents the equalisation curve that the Linkwitz transform is supplying to the amplifier in order to compensate for the new Qtc and Fc. The transform is cutting out 1.2 dB in order to compensate for the Qtc of 1.0, which is causing a 1.2 dB, boost around 70 Hz. It is then providing boost at the lower frequencies at a rate that is equal to the natural roll off of a sealed enclosure. Instead of rolling off at 12 dB an octave, the speaker is being forced to maintain a flatter response due to the amplifier giving it a lot more power at the lower frequencies. For example at 20 Hz, the amplifier is giving the speaker an additional 12 dB of gain, or 16 times more power than at frequencies above 50 Hz.

Getting this kind of response out of a sealed enclosure setup requires some tradeoffs. You will be giving up some of the overall SPL producing capability of the driver because of the excursion overhead needed at the lower frequencies.

Compromises

If your speaker is flat down to 50 Hz and you want to extend one octave below that to 25 Hz then you will lose 12 dB of overall output capability when producing sound at 25 Hz because you are using up all of the driver's excursion and most likely power handling capability too.

We'll say that we have an imaginary speaker that has these basic parameters ...

Efficiency	88 dB/m/W
Xmax	4 mm
Max Power	250W

The driver is in a sealed box with a Qtc of 0.707 (optimally flat). We'll also say that it is flat down to 50 Hz, and at 50 Hz it is capable of producing (just under) 112dB SPL at full rated power. Let's say that the speaker is in a sealed enclosure, and an equaliser will be used to obtain the last octave (down to 25 Hz).

Below the Fc of 50 Hz the speaker will roll off at 12 dB / octave. That means at 25 Hz, the output of the speaker will only be 100 dB. To obtain the same 112dB output as before, you will need 4 times the excursion and 16 times the power (i.e.4,000 Watts!) as

at 50Hz. This is because for every octave lower that you want the speaker to produce you need 4 times the excursion, and in order to obtain the excursion in an equalised system, you must have 16 times the power. To put this another way, excursion is equal to the inverse square of frequency. Half the frequency, four times the excursion, one quarter the frequency, eight times the excursion (etc.). The power requirement is the square of the excursion.

So, now that our speaker does not roll of at 12 dB / octave (because of the equaliser) and maintains a flat response to 25 Hz, it will need to use 4 times more excursion and 16 times the power to produce 25 Hz at 112 dB.

Now comes that tradeoff part that I was talking about before. The driver is using all of its power rating and 1 mm of its 4 mm of Xmax to produce 112dB at 50Hz. Now we want to achieve a flat response down to 25Hz by using the Linkwitz transform. That means we will be trading off some of our 112 dB SPL to gain some low frequency flat response. The speaker will be using 4 times the 1 mm Xmax or 4 mm of excursion to produce this 25 Hz frequency. The power needed is well in excess of the speaker ratings, so must be limited to 250W.

If we want to use this speaker in the transform, we now need to trade off 12 dB total maximum output for a total maximum output of 100 dB that is flat from 25 Hz up. That's not too bad of a trade in my opinion. But remember, this is only an imaginary driver that I made up, not a real world example, although in reality, most "real" speaker drivers will not be all that far off. Every 3 dB increase in SPL requires double the previous amount of power. And in our case we needed 12 dB of gain which is 16x more power.

The two things to remember from this example are ...

- every 6 dB increase in loudness requires 2x the excursion and 4 times the power
- every octave lower you go requires 4x the excursion and 16 times the power

---

Power Compression

Power compression is another aspect that should be considered. This occurs whan any loudspeaker is driven with a significant amount of power. The voice coil heats up, and the available power is reduced accordingly. Depending on the program material, you may easily lose 6 dB of SPL because of power compression (assuming that the system is being pushed to its limits). Power compression cannot be compensated for by using more power, as it is dynamic in nature. The effects are (perhaps surprisingly) not as

noticable as one might expect, since sustained high power at extremely low frequencies is rare in virtually all normal program material.

Fortunately, this is not as big a problem as may be imagined, since typical low frequency energy levels are actually surprisingly low most of the time (see Power Distribution below). Home theatre systems will be called upon to reproduce large amounts of relatively deep bass, but only for short periods at a time.

With most music, there is very little energy below 40 Hz, so power and excursion are not normally a problem. Pipe organ music is an exception - the 64' pipe on a full pipe organ is 16 Hz, but it is not used a great deal - in some cases because of the structural damage it does to the building housing the organ! If you are an aficionado of such music, I suggest that you use the largest box you can, with a very large driver. It may be wise to reinforce your home while you are at it (and no, I'm not joking).

Guidelines

With the explanation and examples out of the way, you might be wondering what kind of specs to look for when choosing a driver. Here are some guidelines that should help you.

- Look for a driver with a BIG linear Xmax. The driver should have a one-way Xmax of over 12 mm (0.5").
- Look for a driver that is 300 mm (12 inches) or bigger. Remember, that producing low frequencies is all about displacing large quantities of air.
- The driver should have a high power handling capability, in my opinion at least 300 watts RMS. The driver will need a lot of power to hit those low frequencies.
- The driver should also have a low Fs. It should be the lowest you can find. The reason for this is because you want the transform to use as little gain as possible to reach the lower frequencies. The lower the Fs of your driver, the lower the Fc of the closed box system will be and the lower overall gain the circuit needs to apply.
- And finally the driver should have a high sensitivity rating, unless you have a really big amp to power it. The higher the efficiency rating of the driver is, the less power it will take to reach those insanely low frequencies. If you have a driver that is 89 dB sensitivity and a driver that is 92 dB sensitivity, the 92 dB driver will require half as much power as the 89 dB unit, for the same sound pressure level.

These guidelines are just that, only guidelines. One thing that you need to be careful of is the excursion and the power the driver can handle. I say this because these are the two that ***will*** damage your precious driver if you exceed them by too much. When I first

did my setup, I thought that a single 300 mm (12") driver would be more than adequate based on my equations for excursion and amount of surface area they had. I was wrong because of my seating location and the room that I was placing them in. Don't get me wrong, a single 300 mm sounds awesome, but 3 x 300s is absolutely unreal because I can hit 105 dB at 25 Hz from my listening position. My listening position also happens to be the place with the best bass response in my room.

Power Distribution (by Gareth Abrey)

I was in the process of building a Linkwitz Transform cct for my 305mm (12") sub. ESP said in the articles about ELF and EAS that low frequency content in music has much lower power levels than the power calculations would suggest. I decided to investigate this, and digitised some tracks off various CDs with various styles of music.

My sound editor gives the WAVE graph amplitude in 16bits (-32000 to 32000). I then performed a low pass filter of 40hz on the track and found the highest peaks, then compared them to the highest peaks of the full range signal, and did a dB calculation ...

20 log (V) (where V is the 16bit amplitude)

Typically, I got peaks of +/-30 000 for the full range signal, and only +/-4000 for the < 40hz signal. This is a 17.5dB difference. The results are tabulated below.

Music Type	**Relative Level at <40Hz**
Rock music	- 13dB
Maria Carey Song	- 15dB
Rap Music	- 14dB
R&B song	- 12dB
Rave track	- 12dB
Second Rave Track	- 21dB
Vinyl Bass Track	- 11dB
Rave track with bass sweep	- 9 dB
Average	**- 11.875 dB (12dB)**

These figures would suggest that boosts of around 10-12dB are possible with the Linkwitz circuit, before any extra amplifier power is needed above that which is required for the frequencies above 40hz.

Editor's Notes

1. There is a strong case for applying a highpass filter at between 5 and 15 Hz. This prevents excessive excursions at sub-audible frequencies, and offers a measure of driver protection. Ideally, this filter would have a steep slope (12 dB / octave minimum), but a simple 6 dB (first order) filter can still be used. The filter may be before or after the Linkwitz transform circuit, having the same effect regardless of physical position.

Use of any filter will have an effect on the actual response of the completed system, however this is generally small, causing perhaps a 2 dB error at 20 Hz. It is probable that virtually ***any*** room will create errors many times this figure, so it can generally be discounted.

2. There is a recommendation in the spreadsheet (see downloads page) that great care is needed with a maximum boost over 20 dB. I think that this is understatement, and care is needed with any boost above about 10 dB. The increase in power and cone excursion becomes extreme, although with most music, the actual energy level of signals below 40 Hz is relatively low.

There are some exceptions to this, and it should never be assumed that you won't need the power or excursion - someone ***will*** eventually prove you wrong.

3. You must remember that the box is sealed. The pressure exerted by a 380 mm (15") cone with an Xmax of 10 mm will literally split the seams of a box that is not sturdy enough (it apparently happens quite regularly with a certain well known subwoofer using a similar principle). The box must be as strong as you can make it - screwed, glued, and substantial internal cleats at all joins are essential. There is no such thing as a box that is too strong, but make sure that you account for the volume occupied by the strength members when you do the calculations). Bracing is usually not needed, since the frequencies are so low that panel resonance is unlikely if the unit is a self contained subwoofer. All panels should be of 18 mm (minimum) sturdy ply or medium density fibreboard (MDF) - do not use chipboard, the box will not hold together!

4. The use of a small (say 5 mm) vent stoppered with felt to present a significant resistance to airflow is also a good idea - especially if the woofer does not use a vented polepiece (via the dustcap). This allows air pressure to equalise slowly, since you will have to expend considerable effort to make sure that the box has no air leaks. If present, they will whistle or make some other equally undesirable noise when the subwoofer is in

use. It is unlikely that you will be able to blame the dog for these noises (in case you thought you might get away with that excuse :-)

5. The Linkwitz transform circuit is available as a PCB with full construction details. The board incorporates a 15 Hz filter (this can be changed) and uses one dual opamp. To have a look, see Project 71.

6. My thanks to Jeremy for putting this article together. His efforts have saved me an enormous amount of time, and the article is written directly for the beginner or relatively non-technical reader. As many of you may have noticed, this is something I often have trouble with :-)

7. I would also like to thank Gareth for his contribution, which is a useful addition. The power calculations he did are somewhat more scientific that the "gut feel" method I had applied - even though the net result is much the same.

Rod Elliott

Special Thanks

A special thanks goes to Siegfried Linkwitz for verifying this document and helping me explain in simpler terms what his circuit is doing.

Jeremy Wolf

Update Information

Jeremy sent me an e-mail from a reader, who pointed out a couple of errors in the calculations for excursion and power. I have amended the "offending" section, which is now (hopefully) correct.

14 Sept 2002 - Added Gareth's power calculation information.

Rod Elliott

Return to Articles Index

Back to Home Page

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Jeremy Wolf, Gareth Abrey and Rod Elliott, and is Copyright © 2001 /2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The authors (Jeremy Wolf, Gareth Abrey) and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allow that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 06 Jun 2001 - Updated 16 Sept 2002, added Gareth's power measurement details

The Audio Pages

Elliott Sound Products **NTM™ Crossovers**

Copyright © 2005 - Rod Elliott (ESP)
Page Created 13 September 2005

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Description
  - 2.1 Circuit Diagram
  - 2.2 Circuit Explanation
- 3.0 Conclusion
- 4.0 References

1.0 - Introduction

The Neville Thiele Method™ * (or simply NTM™ *) crossover network has been described in AES papers and elsewhere, but there is scant information available to most of us about exactly what it is and how it works. There have been claims that it is anything from 48dB/octave to 100dB/octave, and that it may replace the Linkwitz-Riley crossover for general use.

What is not generally available is a description of the filter type, or any information about how it works. This article hopes to rectify that, and de-mystify the hype that inevitably builds up when something new and exciting is first introduced, but with little supporting data to allow an informed choice.

I must confess that I was mightily perplexed when I saw my first NTM crossover - it was claimed to be 100dB/octave, the published frequency response I saw first showed 24dB/octave, and the circuit had far too few opamps and caps to approach any conventional filter greater than 24dB/octave. There was no opportunity to try to analyse the circuit or run any tests, since it didn't belong to me, wasn't in my workshop, and I only had the opportunity to have a brief look inside the case.

It was only when I saw the real frequency response graph of an NTM crossover that the penny dropped, and I recognised that it was probably an elliptical filter.

* NTM™ and Neville Thiele Method™ are trademarks of Precision Audio Pty Ltd (2 Seismic Drv, Rowville VIC 3178, Australia)

For further information, licensing, etc., please contact Wal Dower at Precision Audio Pty Ltd on (613) 9763 9111

2.0 - Description

The descriptions that one usually finds describe a filter network and a notch filter. These are combined to produce a filter with a greatly accelerated rolloff slope. Part of the description from a brochure by BSS audio [1] states the following ...

> A Neville Thiele Method™ Crossover Filter (NTM™) is a new type of electrical/acoustical filter offering significant performance advantages over all previous crossover filter types in audio applications.

The article continues with a description of 'how it works' - while not actually giving any figures whatsoever - just the diagram referred to ...

> The NTM crossover uses a unique notched response to achieve a very steep roll-off rate outside the pass-band. The 4th order Thiele crossover amplitude response looks like the diagram overleaf. You will see that notches in the responses speed-up the rate of roll-off. Beyond the notch, the response rises again, but remains respectably attenuated.

I object (a bit) to the term 'unique' in the above. The filter type is described in 'The Active Filter Cookbook [2], and is commonly known as an elliptical or Cauer filter. There is no denying that this filter type is rather obscure (not too many will have heard of it), but it is neither new nor unique. It is however, an *extremely* clever application of an old technique, with some lateral thinking and necessary adaptation to maintain a *flat* summed response. Its use in a crossover network is certainly new, and the application is unique (if not the filter type itself).

The frequency response is shown below, and this is an almost perfect match to that shown in the BSS brochure. The frequency and amplitude scales are different (from the BSS graph), but the response is virtually identical. For comparison, the response of a Linkwitz-Riley filter is also shown. Not shown is the summed response, which is completely flat for both filters.

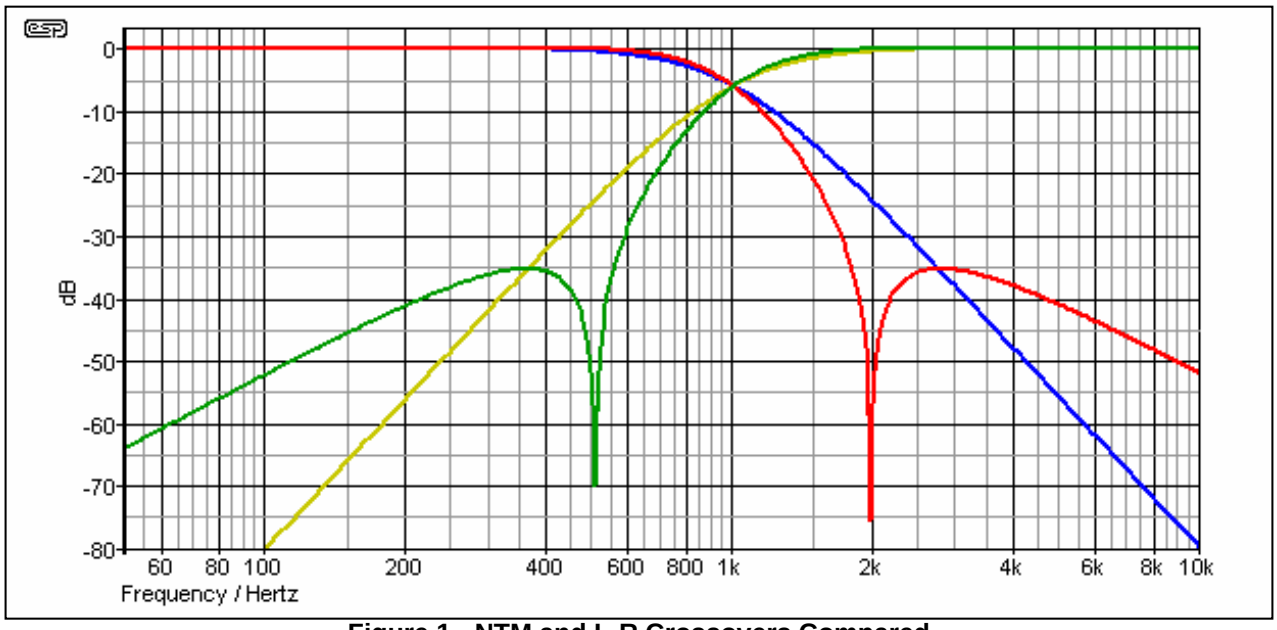


Figure 1 - NTM and L-R Crossovers Compared

The NTM crossover response is shown in green (high pass) and red (low pass), while the L-R equivalent is in a yellowish colour (high pass) and blue (low pass). It is undeniable that at one octave either side of the 1kHz crossover frequency shown, the response is better than 60dB down from the nominal output level. Looking at the L-R filter by comparison, at the same frequency it is only 30dB down. Unfortunately, the response of an elliptical filter rises again after the notch - again readily visible.

About ½ an octave above and below each notch, response is back up to about -36dB (equal to the L-R), and beyond

that performance is inferior to the Linkwitz-Riley implementation. Ultimate rolloff for a fourth order elliptical filter is 12dB/octave. None of this is shabby by any means, but it does show that some of the descriptions used in advertising literature are rather misleading once the full details are known.

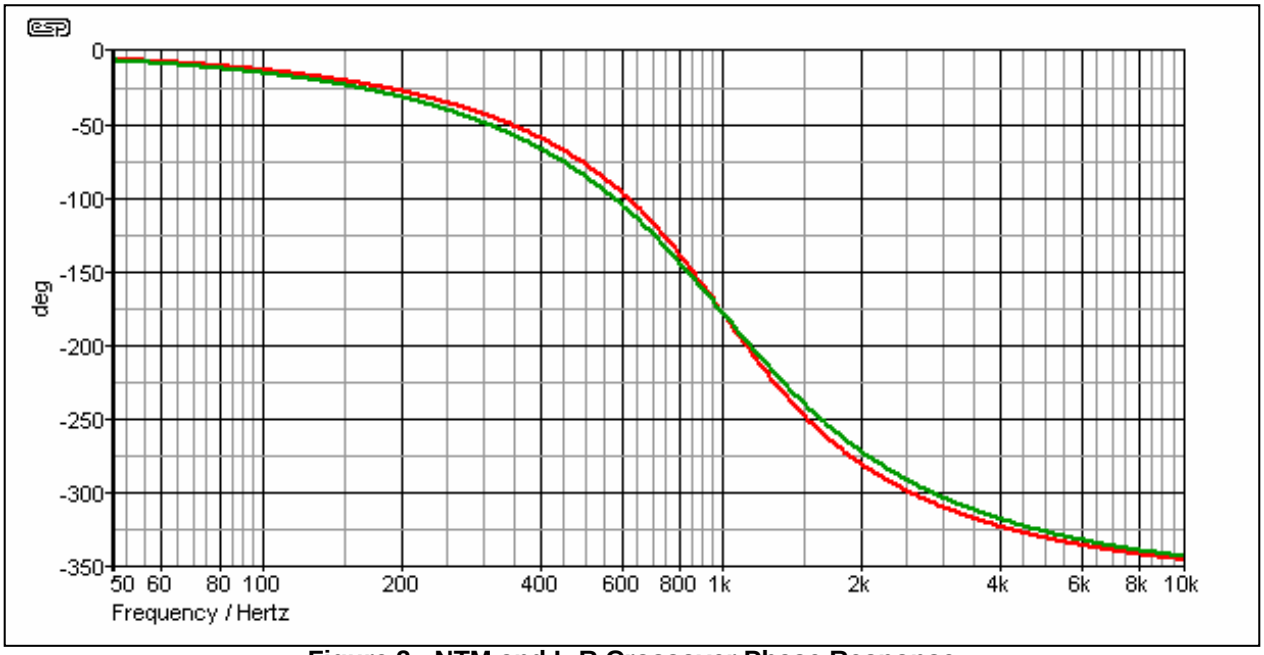


Figure 2 - NTM and L-R Crossover Phase Response

There is not much phase difference between the two - they are (for all intents and purposes) the same in this respect. The red graph is the NTM filter, and the L-R is shown in green. Both filters have a 360 degree shift across the band, with the NTM filter being *very* slightly worse in this respect than the L-R filter.

In real terms, the difference is marginal only, and should not be audible. Despite the apparently radical phase shift, both drivers remain in phase with both filter types, and the phase shift in itself is normally inaudible (despite some claims to the contrary). While there are circumstances that can make phase shift audible, such a discussion is outside the scope of this article.

2.1 - The Circuit Diagram

Please note that the diagram shown below is taken directly from my simulation. It is not (and does not purport to be) the actual circuit of an NTM filter, but I strongly suspect that it will be rather similar. I have not seen the actual circuit, nor has it been traced from an actual working NTM crossover, so the 'real thing' could possibly be completely different.

The general principle of an elliptical filter is/should be pretty well known in engineering circles where filters are used extensively, and it consists of a conventional second order filter, followed by a second order state-variable filter.

The high pass and low pass outputs of each state-variable filter are summed to give the response shown above. The values shown are those used in the simulation. The hardest part of implementing a filter such as this is component 'sensitivity' - the requirement for close tolerance is increased compared to (say) a more conventional Linkwitz-Riley crossover network.

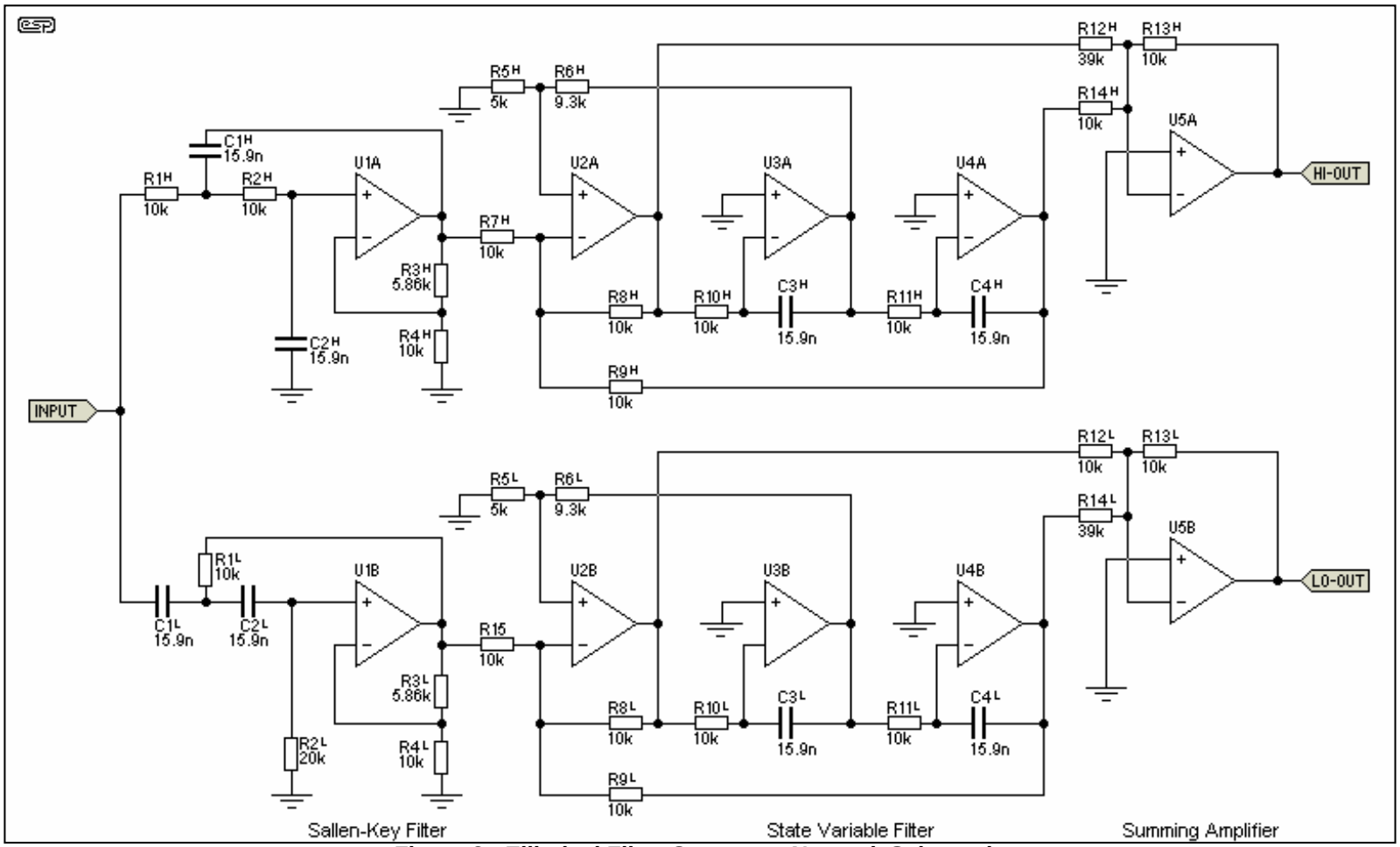


Figure 3 - Elliptical Filter Crossover Network Schematic

As you can see, the capacitor values are non-standard, but that was done to obtain a nominal 1kHz crossover frequency. This was selected because it is a defacto standard when showing general filter responses. It is not a useful crossover frequency for real use. Phase response is shown above, and is almost identical for both filters (NTM and L-R), and the ripple in the summed outputs is less than 0.2dB.

2.2 - Circuit Explanation

The circuit itself is relatively conventional for this filter type, but there are some important variations and points that need some explanation. The first stage (around U1) is what is known as an 'equal component value' Sallen-Key filter. Unlike the standard circuit such as that shown in P09, the Q of the circuit is determined by the gain of the opamp, rather than the filter component values. This allows the circuit to use the same capacitor values as the next stage.

The second filter is a state variable type, using U2, 3 and 4. The filter frequency is set by R10, R11, C3 and C4, and the Q is adjusted by R6. The Q needs to be somewhat higher than for a standard filter, because the summing amplifier (U5) adds a selected amount of (out of phase) high pass to the low pass filter (and vice versa). This creates the notch, and R12 determines the notch frequency.

For such a complex filter, it is remarkably tolerant to component variations, but predictably less so than a fourth order L-R filter of the type used for Project 09.

In case you were wondering, the circuit shown will work, but the frequency determining components will have to be changed to get the frequencies needed rather than the 1kHz shown. In answer to the question (which *will* get asked at some stage) "How do I change the frequency?", the answer is that you will have to figure that out for yourself. This is not (and is definitely not intended to be) a construction project ... it is an explanation of how the circuit works. No

more, no less.

3.0 - Conclusion

This article has hopefully removed some of the mystery behind the NTM crossover network, and shows what can be achieved using some lateral thinking. The rise in amplitude beyond the notch is rather unfortunate, but at -35dB represents an effective power that is over 3,000 times (3,162 to be exact) less than the maximum applied.

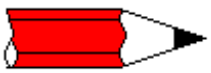
This means that for an applied 100W input power to a loudspeaker driver (via the crossover of course), the worst case out-of-band power level (> 1 octave) is only 31mW. This is further attenuated at a rate of 12dB/octave. At 3 octaves from the crossover frequency, (125Hz and 3kHz), the out-of-band power level is down by 48dB - 1.58mW for 100W input. This is insignificant.

By comparison, the L-R crossover is at -24dB one octave from crossover frequency (400mW), at two octaves it is at -48dB (as above - 1.58mW), and at three octaves it has an output of -72dB (6.3uW) - again, assuming 100W input power to the loudspeaker drivers. For all practical purposes, this is also insignificant.

The end result is that loudspeaker drivers can be pushed closer to their limits, because the out-of-band power is reduced. There is usually no good reason to push any driver that hard in a domestic system, but it can result in a useful improvement for high powered professional applications.

The greatest benefit is obtained at between ½ octave to 1 octave either side of the crossover frequency, with an improvement of around 10dB at the ½ octave frequency, increasing dramatically to the 1 octave point. This represents a significant improvement, but only where drivers are being pushed to their limits. In a domestic system, all drivers will (or *should*) generally have sufficient 'spare' bandwidth to be able to cope with the out-of-band power levels with no stress whatsoever.

Overall, the circuit is very impressive though - not so much because of the cunning application of elliptical filters, but more because of a complete re-think about the way such filters are normally designed and tuned. The Neville Thiele Method™ certainly delivers a very worthwhile improvement in overall crossover network performance.



Please note that the NTM crossover network is patented, so commercial use of the information presented here will infringe patent rights and may result in a law suit or other potentially expensive unpleasantness.

Also, as pointed out above, the circuit shown is **not** taken from any literature, service manual, physical crossover or anywhere else. It is my interpretation of a circuit that will achieve the same result as an NTM crossover produced by a licensed manufacturer.

Based on extensive searches, it would appear that this is the first published circuit (for general viewing) that achieves the results claimed for the NTM crossover network. It remains to be seen if I am asked to remove the circuit (since I haven't used any proprietary information, no laws have been broken - the details published here are all my own work).

4.0 - References

1 Neville Thiele Method™ Crossover Filters, BSS Audio, www.bss.co.uk

2 Active Filter Cookbook, Don Lancaster, ISBN 0-672-21168-8, Howard W Sams & Co., Inc (1979 Edition)

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 13 September 2005

The Audio Pages

Elliott Sound Products **New Speaker Box Project - Part 1**

Copyright © 2001 - Rod Elliott (ESP)
Page Updated 18 Nov 2001

Articles Index

Main Index

Part 2 - The Electronics

Contents

- Introduction
- Initial Assembly
- Final Assembly
- Drawings
- Driver Calculations
- First Listen, + Measurements

Note: Click on any photo to enlarge image.

Introduction

The speaker system shown here is part of my own system, and is intended for tri-amping, so there will be no pasive crossover details. I am using the Linkwitz-Riley crossover (Project 09), my existing amps for the bass and midrange, and (probably) Project 19 LM3876 amp for the tweeters (this is still to be decided at the time of writing).

The enclosures are designed as mirror image pairs for the best imaging. Midrange and tweeter drivers have been carefully located so that diffraction effects are minimised, by ensuring that the distances are different from the centre to each edge of the baffle.

The drivers are as described below, with each having been selected for response, linearity and power handling. Each is excellent in it own right - not necessarily the most expensive, but all have very good performance.

	Woofer	**Midrange**	**Tweeter**
Make / Model	Visaton / GF-250	Focal / 5K4411	Audax / TW025M0
Z (Ohms)	8 (2 x 4)	8	8
fs (Hz)	24	70.7	900
Qms	4.35	-	-
Qes	0.29	-	-
Qts	0.28	0.32	-
Vas (litres)	134	7.7	-
Xmax (mm)	32	1.55	0.30 mm
Power (W)	150	150	55W
Efficiency (dB/W/m)	90	91	92

Table 1 - Driver Details

The Focal 5K4411 is described as a 5¼" Polykevlar Midrange with phase plug, and although I am normally not a great fan of Kevlar, the cone is well treated to prevent the normally nasty breakup effects at high frequencies.

Initial Assembly

The cabinets are made from 18mm Medium Density Fibreboard (MDF), with a laminated construction for the baffle. Figure 1 shows the laminating in progress - I used every clamp I could find to keep them together.

Figures 1 and 2 - Laminating the Baffles ... Enclosure, Half Assembled

The boxes are very solidly braced, and the various stages of the construction are shown below. The midrange is in its own sealed enclosure of approximately 14 litres, and the total volume for the woofer is 35 litres. The tweeter is in its own tiny enclosure, fully isolated from any back pressure from the midrange.

Midrange and tweeter are to be surrounded by a felt-filled cutout, and the front edges of the enclosure will be rounded to reduce refraction. All drivers are mounted as close to each other as possible. To this end, the bottom of the tweeter has been cut off so it sits closer to the midrange than would otherwise be possible. This necessitated the separate tweeter enclosure, since the mounting flange no longer seals the tweeter properly. No grille cloth will be used, as the frame would cause refractions I would much rather be without.

All drivers are secured with metal thread screws and tee nuts. The latter are very firmly attached and glued in position to ensure that they don't come off while I am assembling

the system. Those for the woofer are visible in the photo. The brace above the woofer and the base and side of the midrange enclosure are also seen. The small section beside the midrange enclosure forms part of the main cabinet, and will be stuffed full of fibreglass before final assembly.

The enclosures are quite intricate, and represent a significant amount of work (not to mention sawdust!). As always, the initial preparation takes the longest in terms of assembly. Once all the panels are cut, holes drilled and rebates rebated, the actual assembly is fairly fast, and the boxes were brought from basic bits of MDF to their current status in less than one day. The preparation took the best part of a full weekend.

Figures 3 and 4 - Rear View of Cabinets ... Front / Rear Views

In Figure 3, you can see the rear view of the two boxes (upside down). The block of plywood below the midrange cutout forms the tweeter enclosure, and again, you can see the tee nuts glued in position. At this stage, no corner reinforcements or "minor" internal braces have been added - these are being prepared from various offcuts, and will be added next.

The front view of the left speaker shows the cutout for the felt around the midrange and tweeter. The right box is a mirror image. As you may be able to tell from the photos, these shots were taken shortly after assembly - the glue is not dry yet. All panels are glued and screwed or nailed - the selection of screws or nails was based on the stage of assembly. A nail gun is very fast, and does a great job, but is fairly useless in the first stages of the assembly where initial alignment is critical.

Final Assembly

Now that the boxes are basically assembled, the corner braces, panel braces and rear panel support trim can be added. Figure 5 shows one box from the front, with the drivers mounted to check that everything lines up. Once the backs go on, there is nothing that can be done about a misaligned brace or inadequate clearance, so everything has to be right before this happens.

The colour of the Focal driver's cone is a little unfortunate, but the system is being put together for the sound, with appearance taking second place. Despite the foregoing, the screw heads securing the drivers *will* be painted black before the boxes are finished. The recess for the felt can be seen in Figure 5, as can the rounded edges (only at the front - the others are "conventional", and just lightly sanded to take off the sharp edge that would be too easily damaged).

Figure 5, 6 and 7 - Front and Rear View of Box Showing Drivers and Braces

The internal braces were installed using wood glue, but this was reinforced after the photos were taken, using automotive body filler. This is very strong, and sticks to wood extremely well, providing a very well braced and extremely sturdy box. Don't underestimate the usefulness of "bog" for cabinet making - it is a better adhesive and better gap filler than almost anything I have ever used. Highly recommended :-) Unfortunately, it is no good with solid timber that is to be stained and varnished, but it can be veneered over perfectly well.

Note that the braces are all at an angle to help break up any internal standing waves. Considering the amount of fibreglass that will be used, this is probably unnecessary, but

it is worth the very minor effort. Note also the carpet that is glued and stapled to the bottom of the midrange enclosure. There is also some carpet wrapped around the tweeter enclosure "block" to prevent diffraction within the box itself, although this cannot be seen in the photos.

At the bottom of the box, the tee nuts for mounting the stand can be seen. As with all tee nuts in this project, they are glued into position so they cannot come loose.

Drawings

Not a great deal here - the boxes were designed pretty much "on the fly", but knowing exactly what I wanted. The drawing came last - basically after the boxes were completed to the degree shown above. This is not a complete drawing - braces and the side of the midrange are not shown, but these are not overly critical (perhaps surprisingly, or perhaps not). All dimensions shown are in millimetres, and the drawing is not to scale.

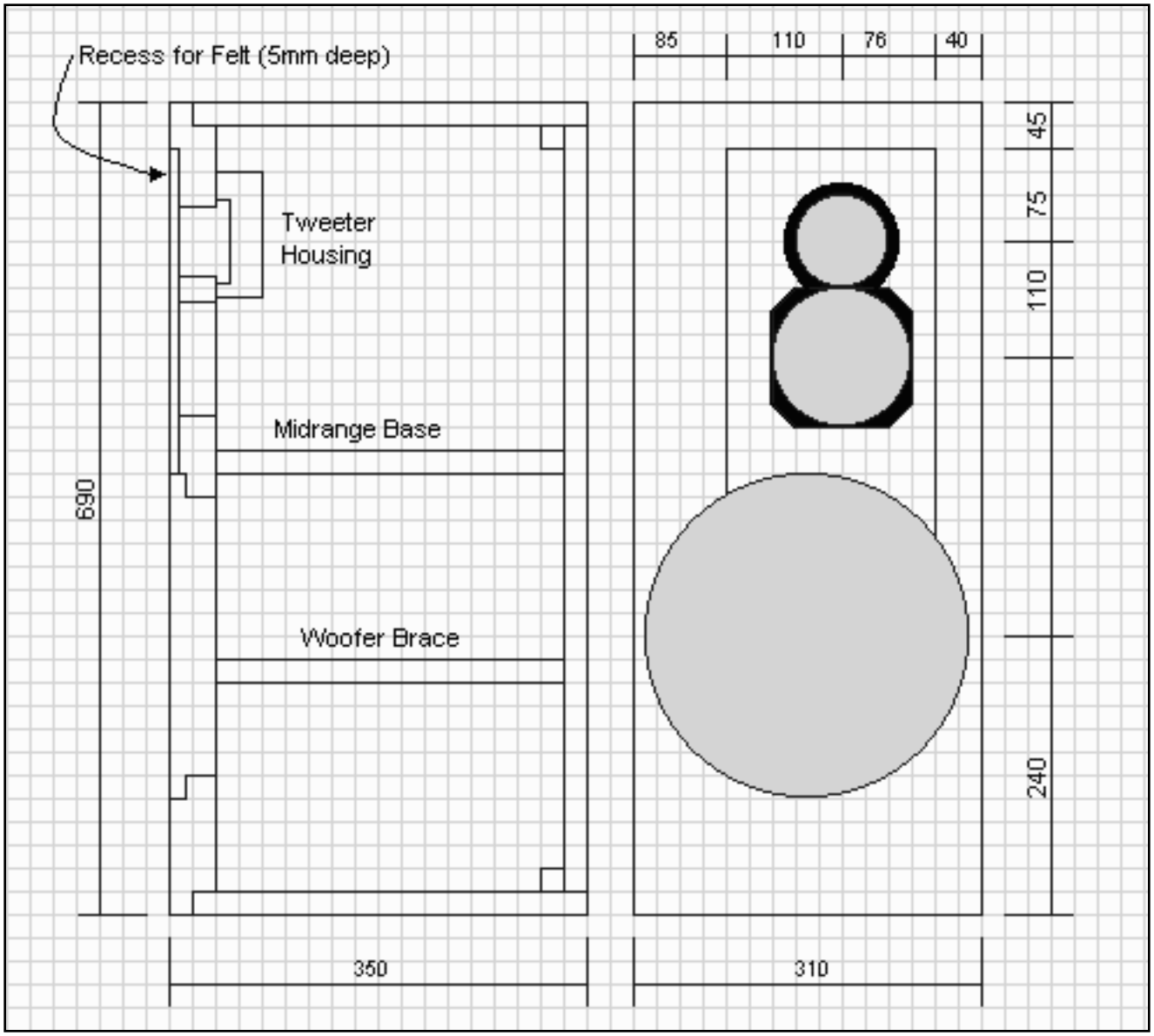


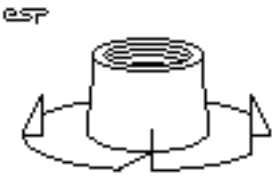
Figure 8 - Side and Front Elevations

As noted above, the midrange housing is about 14 litres, so it is quite easy to work out where the side panel goes - simply determine the distance based on the volume of the enclosure.

The actual volume becomes slightly less than the calculated value, because of the braces and the volume of the speakers themselves. The response will be plotted (then measured) in the next episode. In the meantime, the photos below show the boxes with all the fibreglass in place, both with and without the back installed. The enclosures are now ready for final sanding and finishing (after a listening test, of course :-)

Figures 9 and 10 - All Fibreglass Installed ... One Back In Place

As can be seen, the packing is extensive. The panels are all very acoustically dead, and a full test for airtight sections will be performed before the drivers are installed. I am not entirely sure how to do this properly yet, but I'm sure I will think of something before next weekend :-)



The cutout for the connector panel can be seen in Figure 10, and again, this is held in place by tee nuts that are glued in place. It is a source of considerable annoyance to many reviewers that so many speaker manufacturers use particle board screws to attach the drivers to the baffle - and this includes many "high end" models. While my speakers will not get "reviewed" (other than by friends in the industry), I do not think that anything less than a metal thread screw is adequate for fastening drivers - or anything else for that matter. The combination of the metal thread screws and tee nuts is almost indestructible - the box will rip apart before the screws fall out! The drawing of a tee nut may help those who don't know about these wonderous little fasteners. Drill a hole and hammer it in (and use a little glue to ensure it *never* falls out).

Driver Maths

As some readers may be aware, the amp I am using at present has an output impedance of about 2 ohms. This has the effect of increasing the Qts (total Q) of the drivers, which was useful in my current speakers, but is equally valuable with the new drivers. I have not determined the exact impedance I will need yet, but the basic results that I will need are shown below.

Visaton GS-250 Woofer

Qts - 0.28 Standard, 0.34 Desired

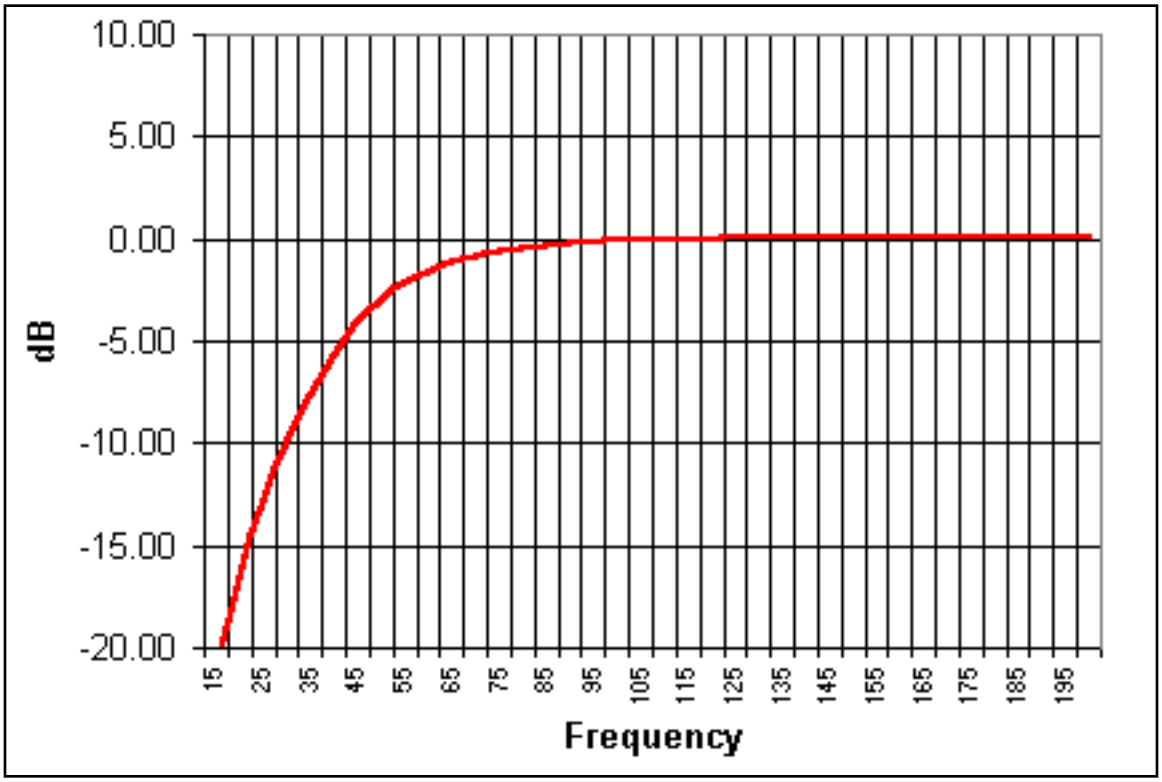


Figure 11 - Desired Woofer Response (Qts = 0.34)

This is about as good as it gets, the response is much more desirable than it would be if driven from a "conventional" amplifier with close to zero ohms output impedance. Resonance is at 53Hz, and F3 is 50Hz.

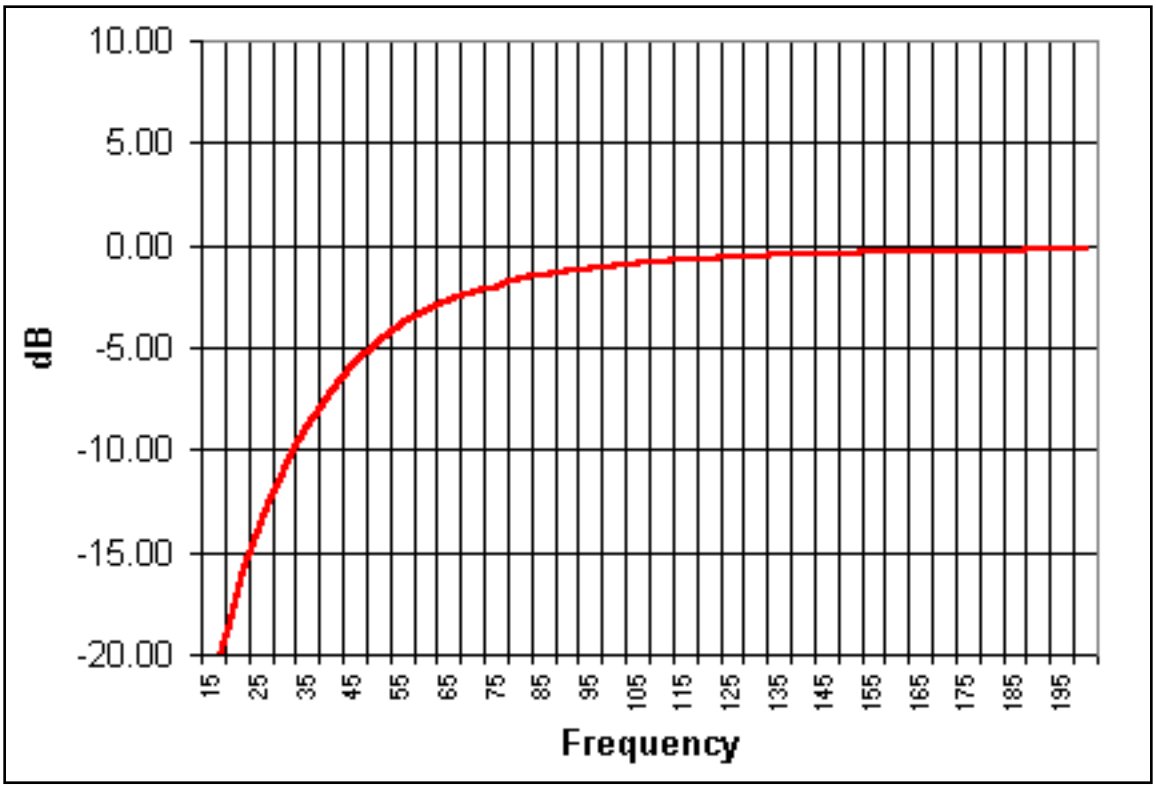


Figure 12 - Existing Woofer Response (Qts = 0.28)

Resonance remains at 53Hz and F3 is 61Hz - combined with the slow droop, this is not as good as it should (or could) be. The suggested size for the box is 22 litres, giving F3 of 63Hz - somewhat higher than I want.

The slow droop in response is quite typical of any speaker in a box that's too big, but the downside of making the box smaller is that the low frequency -3dB frequency increases (as it must). A Linkwitz transform circuit could be used to achieve the same result, but with added electronics and a lot more work than simply increasing the amp's output impedance.

Focal 5K4411 Midrange
Qts - 0.32 Standard, 0.60 Desired

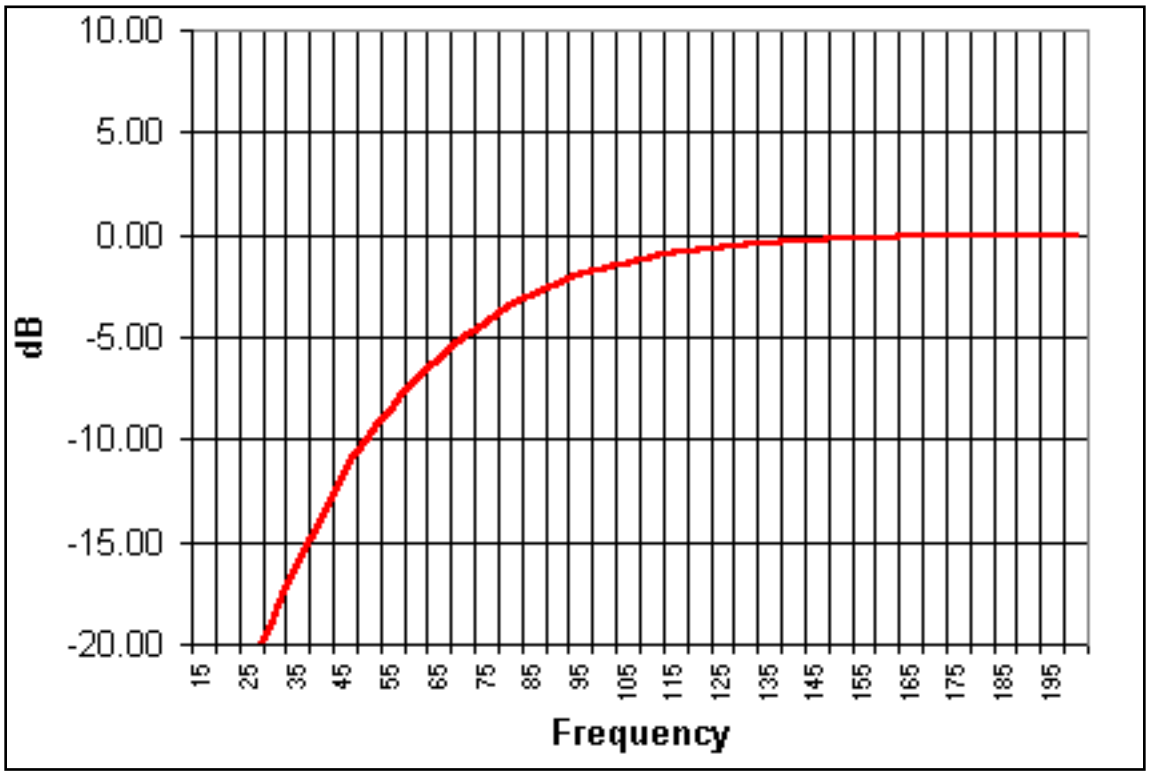


Figure 13 - Desired Midrange Response (Qts = 0.60)

Resonance is 88Hz, and F3 is 83Hz, which is just fine, since it will be crossed over at 300Hz. This is an excellent result - all I have to do now is determine the exact output impedance I need to obtain this result.

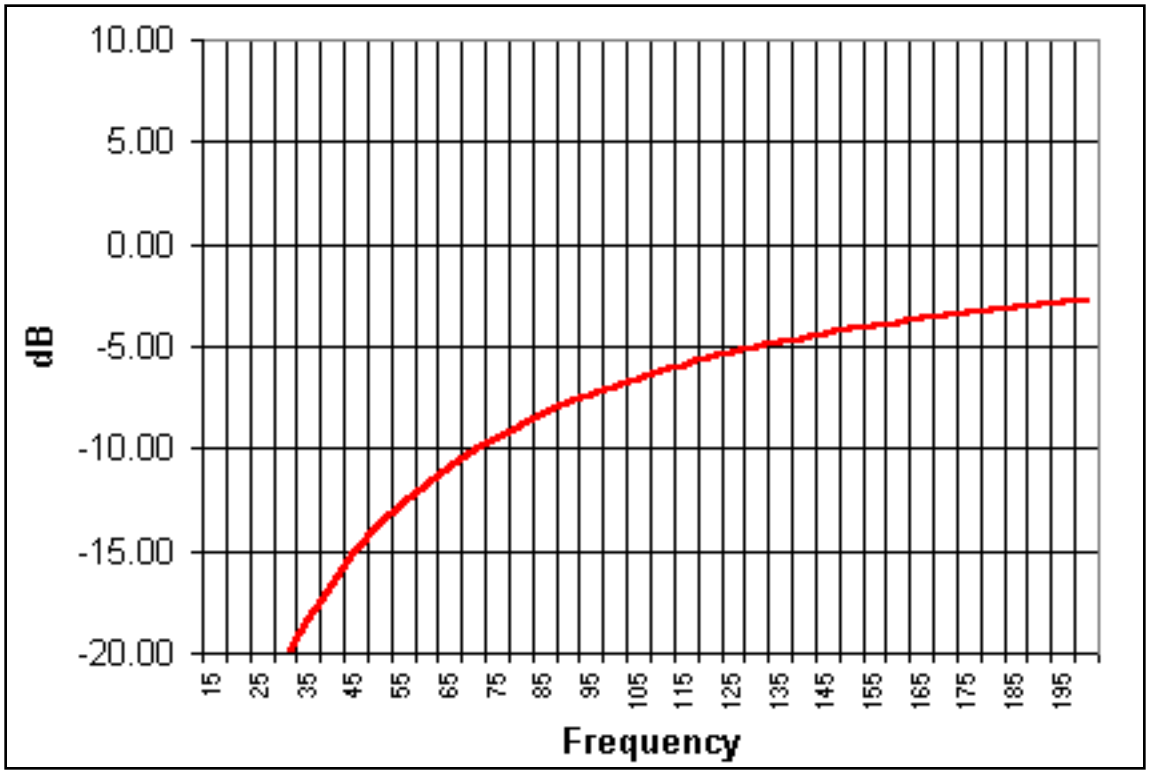


Figure 14 - Existing Midrange Response (Qts = 0.32)

Using a conventional drive amplifier gives resonance is still 88Hz (as it should be), but F3 is 187 Hz - way too high. Again, equalisation could be used, but adds additional electronics to the equation. The slow droop extends all the way to the crossover frequency, a less than desirable outcome. Again, using a higher than "normal" amplifier impedance can correct for parameters that are not exactly what you want.

First Listen, + Measurements

The first listen has shown that these loudspeakers are exceptional. Hooked up to my triamped test bench amp, all I had available for the inaugural listen was an FM tuner, but the results sounded very impressive. Bass, mid and treble are well balanced, there is virtually no panel resonance, and the sound quality remains extremely good above, below and each side of the axes. Listening from a separate section of my workshop showed that the overall balance is very good indeed.

It never sounded like a speaker around the corner, but with speech I was almost tempted to rush back to see who was there - very lifelike indeed. Bass is actually much better than expected, and with an amp output impedance of 4 ohms, there was very good extension to 40Hz, and even 20Hz was reproduced (but somewhat subdued, as is to be expected).

The resonance of the Visaton driver in the cabinet is 52Hz - quite close to the calculated value. The Focal was a different matter, with resonance at 69Hz - considerably lower than the specifications indicate. This is of little consequence for a triamped system, but shows the importance of verifying the important driver parameters which will have a profound (and undesirable) effect on any passive crossover.

A near field scan of frequency response of each driver (bass and midrange) indicates that the response is almost dead flat across the designed frequency ranges. There are the usual minor peaks and dips, but absolutely no major "suck-outs" or resonant peaks were to be seen - or heard. The response across the crossover frequency is harder to measure near field (actually it is almost impossible), but there was no audible variation - so much so, that I had to check the frequency and feel the drivers to find out which one was reproducing at the time.

The system is very well behaved with any crossover frequency between 100 Hz and 400 Hz (bass to midrange), and is also extremely tolerant of anything between 2 kHz to 4 kHz between midrange and tweeter. I will be using 250 Hz and 3 kHz crossover frequencies, and a power analysis is yet to be performed to verify the relative power needs of each driver. Bass and midrange are expected to be about equal because of a slightly lower than my "normal" crossover frequency, and different driver sensitivities. I don't expect that more than about 10W will be needed for the tweeter to balance the 70W available for bass and mid.

The drivers have now been removed from the enclosure, and final finishing is in progress. This is destined to be time consuming, but I used four coats of black spray enamel, followed by four coats of clear. The final finish will be to rub the cabinets down with Scandanavian finishing oil and fine steel wool. This imparts a lovely smooth satin finish, and is similar to the final treatment of the subwoofer project described in the Projects Pages.

Finishing and Assembling

The final finish is rather nice, but cannot really be shown to any great advantage on a web page. You will get some idea from the photos that follow. Figures 15 and 16 show front and rear views respectively. The felt surrounding the midrange and tweeter can be seen, and the connection panel is visible in Figure 16.

Figures 15 / 16 - Front and Rear of Completed Boxes

Figure 17 shows a close up of the midrange and tweeter - not exactly exciting in hindsight, but it does give you a better idea of how the drivers are close-coupled, the felt surround itself, etc.

Figure 17 - Mid and Tweeter

The full length photo in Figure 18 lets you see the finish (to some degree at least), and what the final enclosure looks like. The stand is mounted using metal thread screws and a felt spacer to prevent vibration. This is shown in a little more detail later.

Figure 18 - Finished Enclosure

The photo in Figure 19 shows what had to be done to get the height right. The only stands I could get were 300mm high, and this was too much. I cut off the top section, then re-attached it to the shortened stand with screws and nuts. Welding was out of the question, since it would have damaged the finish badly.

Figures 19 / 20 - Speaker Stand and Connection Panel

The connection panel shows how the binding posts were glued in position to prevent them from coming loose. The colour coding of the connectors should have had a violet or blue binding post for treble, but my supplier didn't have any :-(

Figure 21 - A 1.5V Cell ??

Figure 21 seems somewhat incongruous at first sight, but I used it to ensure that all speakers were properly phased. When a positive potential is applied to the plus (+) terminal of a speaker, it is meant to produce a positive pressure wave, so the cone should move outwards. This is readily seen even with 1.5V on any speaker, and the current is low enough that it will not damage the most sensitive tweeter.

Listening

The speakers are in operation, and my old faithfuls have been relegated to a spare room. Since I am by nature somewhat impatient (OK, *very* impatient), I wanted the speakers operational before I'd built the electronic crossover and tweeter amplifier. Since all speakers simply terminate on the rear panel, it was a simple matter to use an

external passive crossover network between midrange and tweeter as a temporary measure.

As regular readers well know, I am not a fan of passive crossovers, however, listening to my new speakers was more important than waiting until I could triamp them properly. I used a commercial 12dB/octave crossover, which would normally mean that it would be Butterworth alignment (See Passive Crossover Design for more info). This I did not want, so with a bit of an educated guess I decided that 12 Ohm resistors in parallel with each driver would do quite nicely, converting the filter alignment to a sub-Bessel (Linkwitz-Riley) having a Q of 0.5. A quick bench test confirmed that this was almost perfect, but in reality, there is a small discrepancy because I didn't equalise the midrange driver's inductance. This was acceptable, since it is temporary only.

The sound is wonderful ! Even before careful equalisation and the full electronic crossover (including baffle step compensation), I am extremely satisfied with the performance. There is a hint of over-brightness, but otherwise colouration is extremely low, and imaging is superb. The definition is extremely good, and voices sound so natural that you'd almost swear that the person was in the room.

The second installment covers the electronics - See Part 2

Articles Index

Main Index

Part 2 - The Electronics

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 26 Jun 2001

The Audio Pages

Elliott Sound Products **New Speaker Box Project - Part 2**

Copyright © 2001 - Rod Elliott (ESP)
Page Created 18 Nov 2001

Articles Index

Main Index

Part 1 - The Loudspeaker Boxes

Contents

- Introduction
- Initial Assembly
- Final Assembly
- Listening Tests

Introduction

The speakers are complete, as described in Part 1 of this article. The next phase (actually completed some time ago now) was to finalise the Linkwitz Riley crossover, tweeter amp and re-establish my phono preamp. The idea was to make the final design a complement to the VP103 valve preamp, both in looks and function.

To this end, one power switch now brings my preamp and all seven power amplifiers on-line (six for the stereo tri-amped speakers, and the sub-woofer amp). There are four amplifiers in my original power amp unit, a further two in the tweeter amp, and the separate sub amplifier.

All mains switching is performed in the tweeter amp unit, which has a 20A relay. This is powered from the 12.6V heater supply from the preamp - I added an extra output for the supply, and there is a corresponding input on the tweeter amp unit. There is zero power usage when the system is off.

Initial Assembly

Construction was not without difficulty. For example, neither the toroidal power transformer nor the heatsink would fit in a 1RU (1 "rack unit" is 44.45 mm, or 1 3/4 inches) rack case, so some modifications were performed so that they would. I was able to utilise many of my workshop "toys" to bring this project to fruition - always a good thing, since it justifies their purchase to SWMBO (She Who Must Be Obeyed :-)

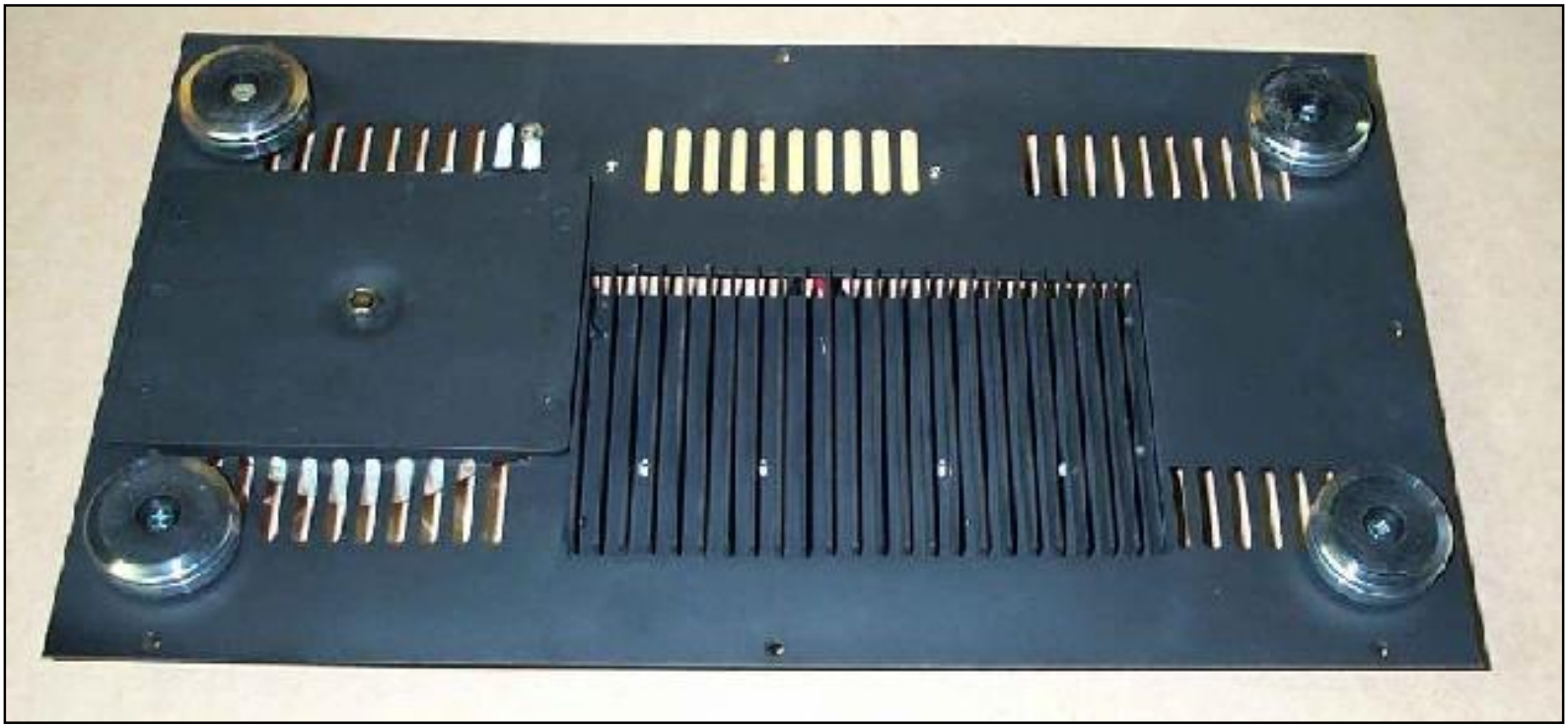


Figure 1 - Base of the Rack Case

Figure 1 shows that there is a cutout for both heatsink and transformer, and both project about 10mm below the bottom of the case. The bottom plate was cut and bent to provide mounting flanges for the heatsink, and a separate panel was attached for the transformer.

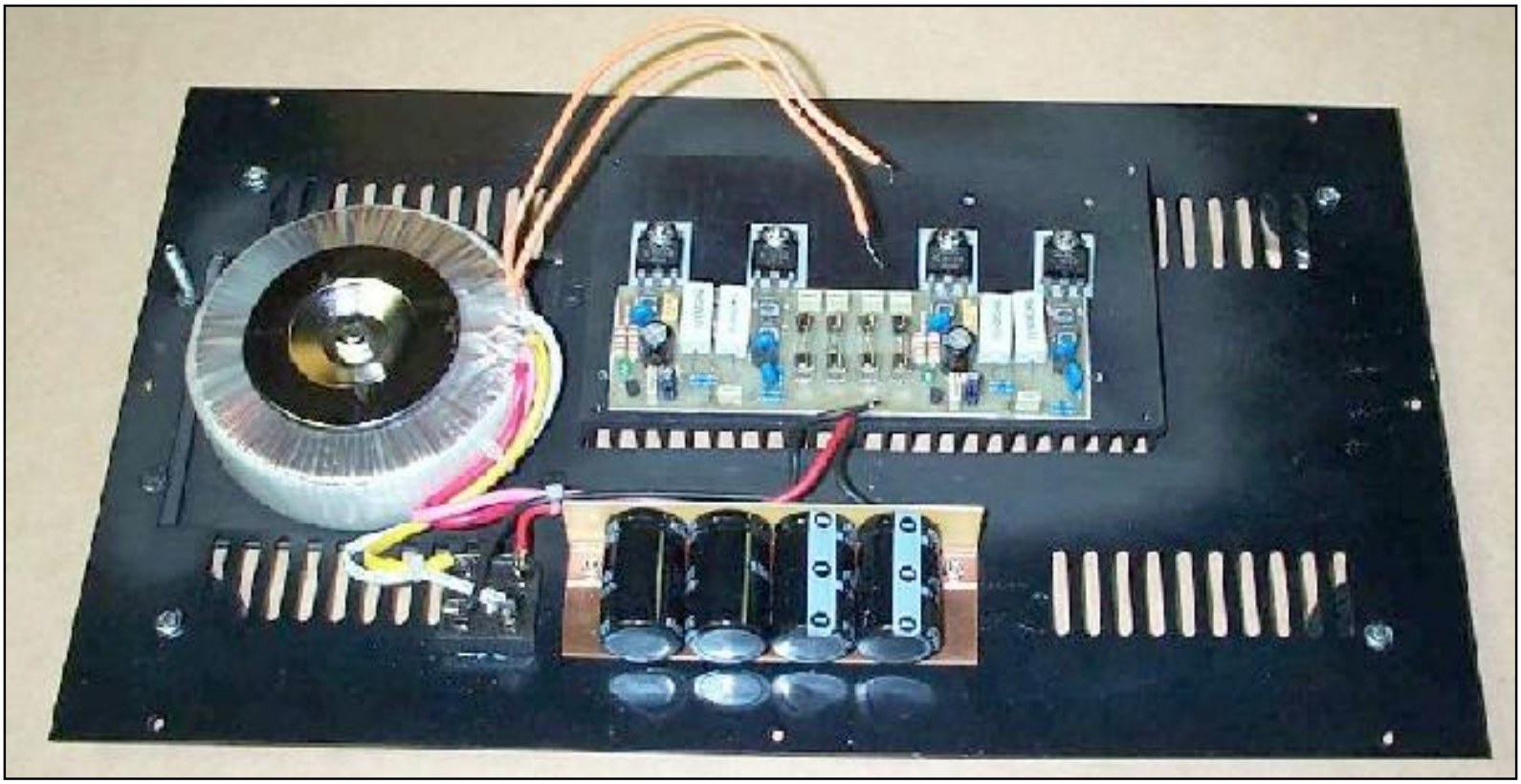


Figure 2 - Top View of Baseplate

In the view in Figure 2, the transformer can be seen, now recessed into its little cavity. The large screw to the left is for earthing. The power supply has been wired at this stage, and the filter caps were mounted on a piece of blank PCB material that I mechanically etched using a hand-held engraver unit. The power amplifier is actually the prototype of the Project 3A amplifier board - before I decided to extend the PCB for the output transistors.

The supply voltage is +/-25V from a 120VA toroid, and the amp is used only above 3,500Hz, so loading is light, even at very high volume. As you shall see, the main supply is also used for the preamp power supply (Project 05), which powers all the smaller boards.

Final Assembly

The next photos shows the unit in a very advanced stage - it is almost complete. All the circuit boards are wired and installed, as well as the switching module (bottom right) and muting relays (bottom centre). Standard IEC mains connectors are used for mains input and switched output, with all mains connections properly shrouded to prevent contact.

Figure 3 - Rear View

The inscrutable box on the left side houses the phono preamp. Being high sensitivity, the last thing I wanted was hum or other noise, so complete shielding was the answer. It is also made with adjustable gain, since I use a (relatively) high output moving coil pickup most of the time.

The earth connection for the turntable is on the extreme left of the rear panel. That is followed by phono inputs and outputs, and the gain switch. The next set of connectors is the input from the preamp, followed by midrange and low frequency outputs. The "ordinary" phono socket is for the DC switching voltage. The set of binding posts / banana sockets is for the tweeters - not surprisingly.

Figure 4 - Front View

From the front, you can see the Linkwitz-Riley crossover - two boards stacked on top of each other on the extreme right. Just to the left is the power supply unit (Project 05), and you get a clearer look at the muting relay board (just right of centre at the top of the picture).

Although it all looks neat and tidy, there were a few stages where it looked like a rat's nest, while I moved wiring about to get the lowest noise levels. Because there is so much packed into a small case, this was not a simple task, and even moving a cable by a few millimetres made a substantial difference - especially the wiring to the tweeter amp inputs.

In this view, the power supply regulator is again visible, and no - the heatsinks do not touch each other, it just looks that way in the photo. They are actually held apart with a small dab of hot-melt glue, to make sure that they cannot short out.

As you can see, there is no power switch, just one green LED on the front panel (plus the obligatory ESP logo, of course).

Listening Tests

The final listening tests with this system have been a true revelation. It is quite possibly the most revealing system I have heard, with nothing - ever - disappearing into the mix. Even at quite astonishing sound levels (about 110dB peak!), a violin does not tear out your ear drums and leave them bleeding on the floor. It just sounds like an incredibly loud violin. Other instruments fare equally well, as do vocals, both male and female. Normally though, I have

found that my preferred (serious) listening level is most usually about 90dB (average unweighted), and often less.

> After some measurements, I can give the following (completely useless :-) information. For an average listening level of 85dB SPL, I measure just under 3V at the midrange and woofer terminals - i.e. about 1W for each. The tweeter voltage is around 1V, or 125mW. From this, I deduce that the total power per speaker is about 2W on average - the tweeter power can be ignored as insignificant. With speaker efficiency of 90dB/W/m, this means that I have a room loss of about 10dB at the listening position, based on applied power versus measured SPL. (Listening distance is about 2.5 metres.)

One thing that this system has highlighted is just how bad some recordings really are, and the artefacts of compressors and limiters (for example) are immediately audible. Although I prefer to listen to the music, rather than the recording technique or the system, sometimes it is just not possible when every flaw is revealed so clearly. As a result, I am re-assessing my CD and vinyl collection to some degree - in many cases to hear details that I have never noticed before. As always, there is a good and a bad side to having such a transparent system.

In fact, the sound is so realistic, that from another room voices sound as if there are people in my lounge room. On more than one occasion, I have had to get up and see who was there when I am working in my study, only to find that it is just the TV, which (naturally!) is connected through the hi-fi as well.

Overall, I am enormously satisfied with the speakers, crossover and indeed the complete system. It has not been a cheap exercise, but compared to the cost of purchasing a system that could come anywhere near it, I feel that I have a real bargain.

Articles Index

Main Index

Part 1 - The Loudspeaker Boxes

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 18 Nov 2001

The Audio Pages

Elliott Sound Products **Design of Passive Crossovers**

Copyright © 2001 - Rod Elliott (ESP)
Page Updated 09 Aug 2004

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Filter Types
- 3.0 Speaker Effects on Filter Response
  - 3.1 Impedance
    - 3.1.1 Woofers and Midrange
    - 3.1.2 Midrange and Tweeters
  - 3.2 Amplifier Impedance
  - 3.3 Temperature
  - 3.4 Atmospheric Changes
- 4.0 Selecting the Filter Slope, Alignment & Components
  - 4.1 Filter Slope
  - 4.2 Filter Alignment
  - 4.3 World's Worst Passive Component
- 5.0 The Maths Behind the Filters
- 6.0 Attenuation Networks
- 7.0 Power Losses / Resistor Power
- 8.0 Winding the Coils
- 9.0 Conclusion
- References

1.0 Introduction

While there are many articles elsewhere that discuss passive crossover design, not all follow a scientific approach. There are several 'off-the-wall' designs scattered throughout the Internet that are a case in point, and unless there is *real* science described in any article you see, it is best avoided.

Highly recommended reading is Project 82 - Loudspeaker Test Box. This will make quite a bit of the testing described below redundant, and you will be able to see at a glance what network is best suited to your drivers. It does not do notch filters though, so this is something that you may still need to determine mathematically (or just cheat and use the spreadsheet)

Most people who have read my pages will know by now that I am not a fan of passive crossovers. However, sometimes it is the only sensible approach, or is necessary because of financial considerations or just for simplicity. Before deciding on the use of a passive rather than active crossover, the following article will surprise you - perhaps even enough to make you decide to go active after all.

Where passives are to be used, I prefer a simple 6dB/octave unit, but in many cases this is not possible - most commonly because of the low rate of roll-off, which can cause excessive power to be delivered within the stop band. This can excite the resonant frequency of tweeters, and causes a honkiness in the upper midrange - usually subtle, but audible nonetheless.

A conventional Butterworth 12dB/octave filter is still by far the most common crossover, but is now under threat from the Linkwitz-Riley alignment. The latter has a crossover frequency where the output of each filter is 6dB down, and this has the advantage of a zero rise in output at the crossover frequency. The 'conventional' crossover filter is 3dB down at crossover, and the summed output shows a slight peak of 3dB at the crossover frequency. This phenomenon occurs with both electronic and passive crossovers using the Butterworth alignment. While it is possible to modify the frequency or slope of either the high or low pass section to compensate (usually by manipulation of both amplitude and phase response), this is uncommon in budget designs, and non-existent for off-the-shelf crossover networks.

All crossover networks have problems. Some have more than others. The ideal loudspeaker is a single point source (i.e. small compared to all wavelengths of interest) that reproduces all frequencies. Such a driver is not possible with any technology currently available, so (as always) we must compromise.

The most common compromise is to use two (or more) loudspeakers, each optimised for the frequency band it must cover. Since it is highly undesirable that drivers intended for high frequencies be subjected to low frequencies (and vice versa), the audio signal is separated into bands by a crossover network - either electronic or passive, or a combination of the two.

In this article, I have concentrated on a two-way crossover network. Three-way (and above) will become a nightmare if you must use passive crossovers all the way through. My recommendation is that all low frequency crossovers should be active, and if you must use a passive network, then it should be for the mid to high frequency section only. The basic principles apply to all drivers regardless of frequency range, so it is not hard to extrapolate the examples given to low frequency networks.

The purpose of this article is to explain how to obtain the best possible performance from a passive crossover network, and avoid the major pitfalls that await us in our endeavours. Crossovers are not simple. Electronic units require us to have a multiplicity of amplifiers (one for each loudspeaker driver), and passive units impose other constraints and limitations - not all of which are satisfactorily addressed by loudspeaker manufacturers (including some 'high end' components).

There is information presented here that I have not seen in any other material on the subject of crossovers, with the exception of my article, "Biamping - Not Quite Magic, but Close". Specifically, this is to do with the shift of filter frequency and alignment that occurs when a voice coil is hot. Since a stable operating temperature will never be achieved with music signals, there will be a constantly shifting crossover frequency, and a peak in the frequency response where the amplitude is dependent on the power at any instant in time. This cannot be considered a satisfactory situation, and is quite possibly one of the most compelling reasons to use active crossovers whenever possible.

The article that follows requires that you are very familiar with the use of a spreadsheet or scientific calculator. There are many calculations and measurements to be made to get it right - but the end result will be well worth the effort.

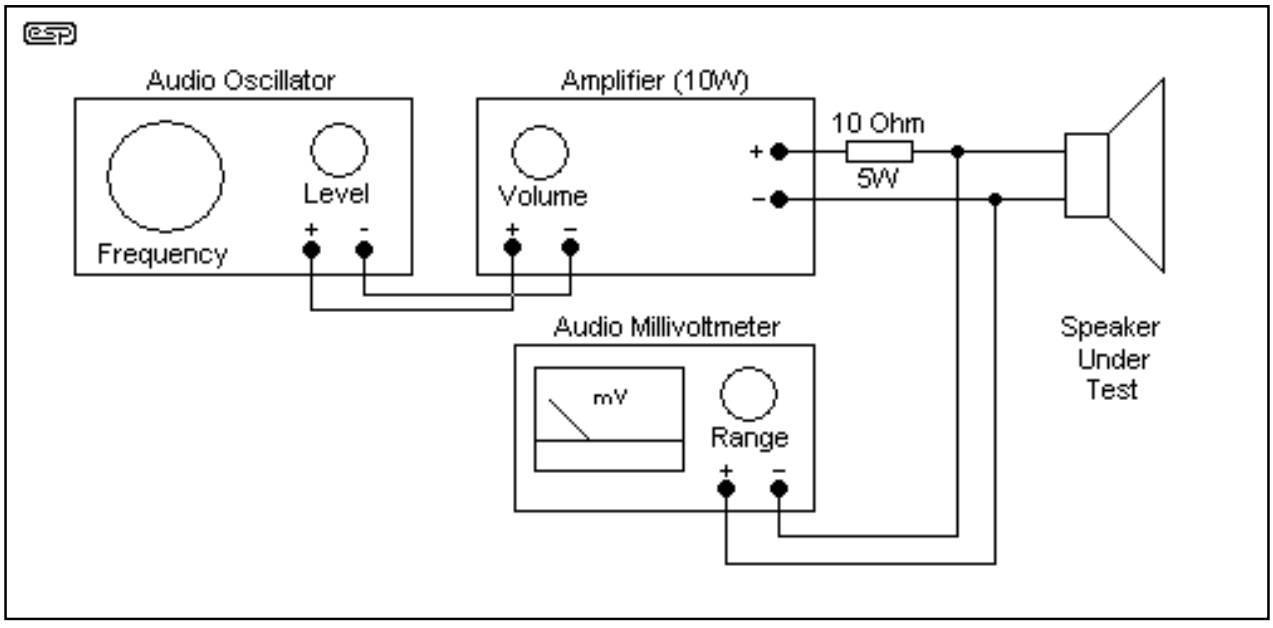


Figure 1.1 - Speaker Test Setup

You will need to set up the equipment shown in Figure 1.1 to carry out any tests on the speakers. The millivoltmeter may be a digital, if you are sure that it has a good enough frequency response. A frequency meter is also very helpful, but is not essential if the oscillator is well calibrated. If an amp larger than 10W is used, make sure that you keep the volume right down - most tests will require a maximum of 1V RMS, so even a 1W amp will be enough.

A spreadsheet is now available to perform the maths for you. You still need to measure the drivers, but once measured, you can simply insert the values into the spreadsheet for a complete starting point. The spreadsheet cannot compensate for all possibilities, and some experimentation will almost always be needed to arrive at the optimum solution.

Passive Crossover Design Spreadsheet

Ideally, when you use the spreadsheet, you should use the *actual* measured impedance of the loudspeakers at the frequency of interest, rather than the quoted nominal impedance. Using the test set shown above makes this quite easy.

Remember that once a Zobel or notch filter has been determined for a driver, that becomes *part of the driver*. The network and driver must be treated as one, since the network's purpose is to remove some objectionable characteristic of the attached driver - most commonly unwanted impedance variations.

To use the test set, follow these steps ...

- Disconnect the speaker, and set the voltage to a convenient level (say 1V)
- Connect the speaker, and sweep the frequency to locate the centre of the "flat" range, where the voltage does not change appreciably
- Measure the voltage

Impedance may be calculated, based on the known input voltage and the voltage measured across the loudspeaker. Using the long way (because it is easier) ...

Vr = Vin - Vs

Where Vr is voltage across resistor, Vin is unloaded voltage, Vs is voltage across speaker

I = Vr / R

Where I is current, and R is the value of the resistor (10 ohms is suggested)

Z = Vs / I

Assume an input of 1V, R = 10 ohms and Vs = 400 mV

Vr = Vin - Vs = 1 - 0.4 = 0.6 V
I = Vr / R = 0.6 / 10 = 0.06 A
Z = Vs / I = 0.4 / 0.06 = 6.67 Ohms

This calculation is also done by the spreadsheet.

2.0 Filter Types

A comparison of filter alignments is in order, so that the reader not experienced in such matters will know what I am on about. There are three primary filter alignments that can be used, and they differ only in the damping (or "Q") factor. Q, or "quality factor" is an abstract term that is applied to many passive components in many applications, and is effectively the inverse of damping. Thus, Q=1/d or d=1/Q. Fascinating stuff (if you happen to be a mathematician :-)

Filter	Main Characteristic	Other Characteristics	Q
Butterworth	Maximally flat amplitude	-	0.707
Bessel	Maximally flat phase	Fastest settling time	0.5 to 0.7 (typ) *
Chebyshev **	Fastest rolloff	Slight peaks / dips	0.8 to 1.2 (typ)

Table 1 - Filter Characteristics

* A true Bessel filter has a Q of 0.57, however there are many 'almost Bessel' alignments in common use
** also spelled Tchebychev in some texts

One of the "magical" number in electronics is √2, or 1.414, and its inverse, 0.707 and the latter can be seen in the table as the figure that provides "maximally flat" frequency response. This means that the response in the pass band is as flat as it can possibly be, until the cutoff (-3dB) frequency is reached. This forms the classic Butterworth filter that has been the mainstay of nearly all crossover systems in common use.

A Bessel filter has a slower and "sloppier" response, that starts to droop well before the cutoff frequency, but has the minimum phase shift (and best transient response), and one that is comparatively gentle. 6dB/octave filters are neither fish nor fowl - i.e. they may be thought of

as a poor Bessel, Butterworth **and** Chebyshev all rolled into one, and have a Q of 0.5 - this cannot be changed by any topology, regardless of whether electronic of passive crossovers are used.

The Chebyshev filter is characterised by peaks and/or dips in its response, and usually has a (slight) rise in amplitude just before the cutoff frequency, the magnitude of which is determined by the Q. The higher the Q, the greater the peak in the response. Depending on the order of the Chebyshev filter, it may have dips as well as peaks. Many vented subwoofers use a Chebyshev response for the port tuning, as do quite a few sealed enclosures. The Q will typically be about 0.8, so the rise in amplitude is less than 1dB, but some will use a Q as high as 1.0 to make the loudspeaker sound as if it has more bass.

Chebyshev filters are rarely used in crossovers - some electronic crossovers have used them, but these are most uncommon. This alignment will not be discussed further (except where it happens by accident due to impedance variations).

3.0 Speaker Effects on Filter Response

The amplitude and phase of a filter is easily plotted by means of circuit simulation. Amplitude response may be measured using a simple signal generator, small power amplifier (for loudspeaker crossover filters) and an AC voltmeter, which must have a bandwidth that covers the audio spectrum - not all do, especially cheap analogue and digital meters. Phase shift is very difficult to measure without an oscilloscope. Digital oscilloscopes have the added advantage of cursors that can be used to obtain accurate readings of time delay at any given frequency, and phase shift can be calculated from this.

Fortunately, once armed with enough information, phase shift measurements are not generally needed. This simplifies the design process considerably, since the phase shift of any given filter type will be known in advance.

In this section, we shall examine some of the possible influences that have (or may have) an adverse effect on the performance of a crossover network. Some of these are well known and are catered for in many (but by no means all) commercial and hobbyist (DIY) loudspeakers. Others are less well known, and are ignored completely by virtually all loudspeaker builders - perhaps with good reason, perhaps not.

3.1 Speaker Impedance

One area where measurement is essential when designing passive crossovers, is the loudspeaker driver itself. There is usually very little information in the makers' data that will prepare you for the behaviour of a loudspeaker / crossover network combination, and these data are usually derived empirically. In some cases the voice coil inductance will be quoted,

and if so, this may be a bonus, as will be shown shortly.

3.1.1 Woofers and Midrange Drivers

In almost every case, the crossover frequency selected for the woofer and midrange driver will be at a frequency where the voice coil inductance is significant. As frequency increases, the effect of the voice coil inductance is to increase the driver's impedance, and this plays havoc with the crossover network's performance.

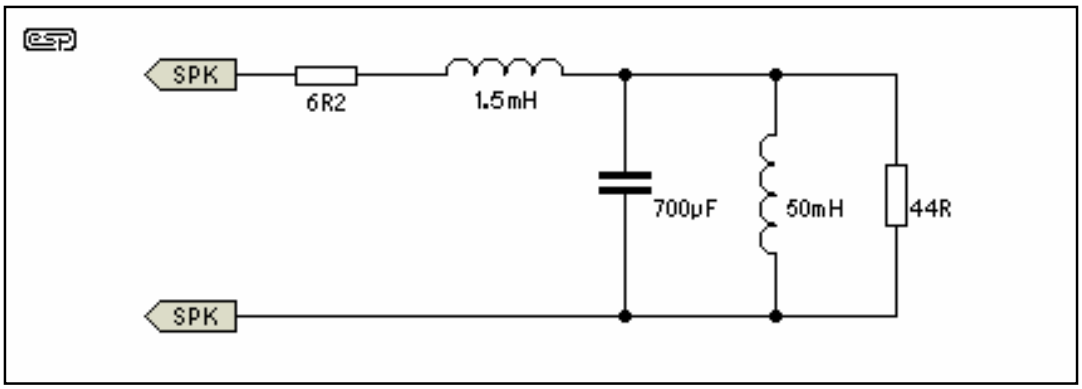


Figure 3.1 - Equivalent Circuit of a Loudspeaker

Examination of a simulated low frequency driver depicted in Figure 1, shows a large peak at resonance, and a relatively small section where the impedance is flat. This equates to the nominal impedance of the speaker, but as shown below, this covers a limited frequency range. The biggest problem for the crossover is not resonance (for a woofer, at least), but the rise in impedance as the voice coil inductive reactance starts to become significant relative to the nominal impedance.

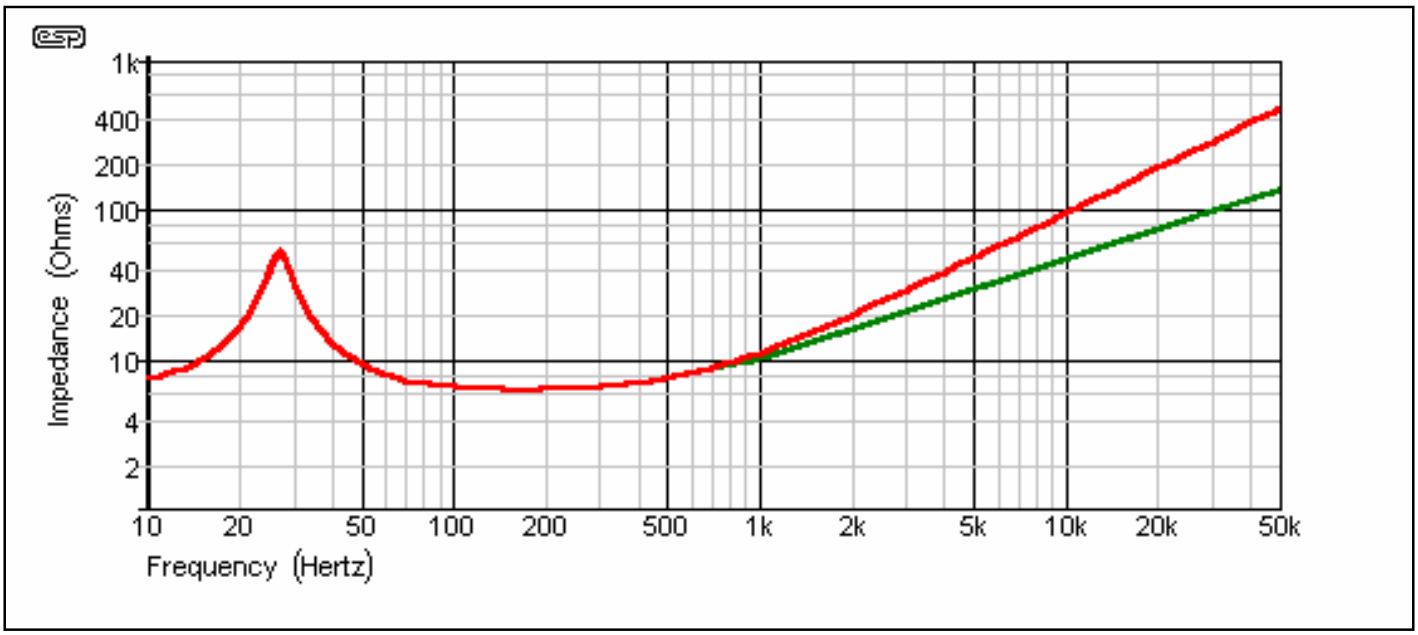


Figure 3.2 - Impedance Curve of "Simulated" Loudspeaker

The green line on the above graph represents the 'semi-inductance' of the voicecoil. Because

the coil is surrounded by conductive metal (steel), its inductance is not linear with frequency. This creates a lossy inductor, and as such it cannot actually achieve the 6dB/ octave (20dB/ decade) impedance rise that one would normally expect. The actual rise varies from one driver to the next, and cannot be simulated in general terms - it must be measured. The real impedance rise due to voicecoil inductance (or semi-inductance) is usually between 3 and 5dB/ octave. For the range of frequencies that we need to be concerned with for a crossover network, the error introduced into the formulae that follow is minimal, and can be ignored.

This is most commonly corrected with a Zobel network (Figure 3.3), connected in parallel with the speaker. As the inductive reactance rises, the capacitive reactance falls, and the resistance is typically made equal to the DC resistance of the voice coil. The net result is a flat impedance curve, as shown in Figure 3.4. This is absolutely essential for proper behaviour of the crossover network, but sadly is not used in a great many designs. The result is a shift in the crossover frequency, and phase response that is not exactly ideal. The crossover may be designed to work with the impedance actually presented by the driver, in which case it will be asymmetrical, having different inductance and capacitance from the theoretical values that one might expect.

In the case of the speaker shown above, at a typical crossover frequency of perhaps 3kHz, the impedance (as simulated) is a little over 28 Ohms, or about 22 Ohms for semi-inductance - imagine the error if a crossover designed for 8 Ohms were used without correction!

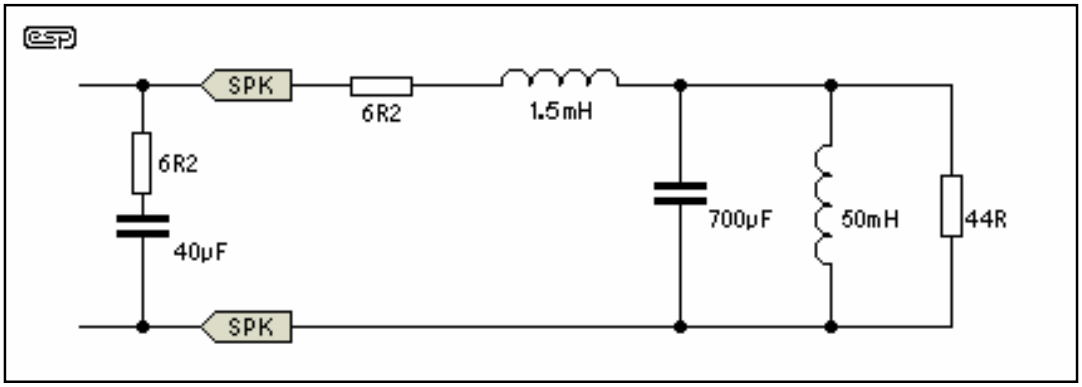


Figure 3.3 - Addition of an Impedance Correction Zobel Network

The Zobel network will flatten the impedance of the speaker, but at the cost of power dissipation, and a slightly lower than expected overall impedance. Naturally, the power dissipated by the resistor is turned into heat, not sound, reducing effective efficiency. The lower impedance may cause some stress to certain amplifiers, but most should be able to cope with the slight extra loading. It must be understood from the outset that the flattened impedance curve does not make the speaker perform any better at the higher frequencies - the sole purpose of the Zobel network is to ensure that the impedance presented to the crossover network remains essentially constant over the frequency range where variations would cause an unacceptable frequency response variation in the filter network. The determination of the required values for the Zobel network is most easily done by

measurement and experimentation.

Once determined, the Zobel network must be treated as part of the loudspeaker. All measurements or calculations for the crossover network *must* include the Zobel network and loudspeaker driver combined. If correctly done, the combination of the two will give an acceptably flat and very usable impedance across the entire crossover region. This will result in a crossover filter with the minimum error possible.

If the voice coil inductance is known, then a suitable value of capacitance may be calculated quite readily. The first thing to determine is that frequency where the inductive reactance is equal to the DC resistance of the voice coil ...

$$f = Rvc / (2 \pi Lvc)$$

Where ...

f = frequency
Rvc = Resistance of voice coil
Lvc = Inductance of voice coil

Once this figure is found, it is a simple matter to calculate the capacitance for the Zobel network ...

$$C = 1 / (2 \pi f Rvc)$$

Using the simulated speaker above as an example, we already know that Rvc is 6.2 ohms, so ...

$$f = 6.2 / (2 * \pi * 1.5 * 10^{-3}) = 658Hz$$
$$C = 1 / (2 * \pi * 658 * 6.2) = 39 * 10^{-6}F = 39uF$$

It should come as no surprise that this is almost exactly the value found by simulation, so we can safely assume that the formula works, and is easy enough to use. The resistance will nearly always be approximately equal to the voice coil resistance - in some cases it may be found that a small variation is needed, but this is unlikely to be significant.

Although the capacitor does not have to be "audiophile" quality, and a bipolar electrolytic could be used, the main problem with bipolars is that they are not stable over time. I recommend that polyester, polypropylene or oil filled paper cap be used, and suggest that you will be faced with a not inconsiderable expense to implement this scheme properly. Is it worth it? ***Absolutely!***

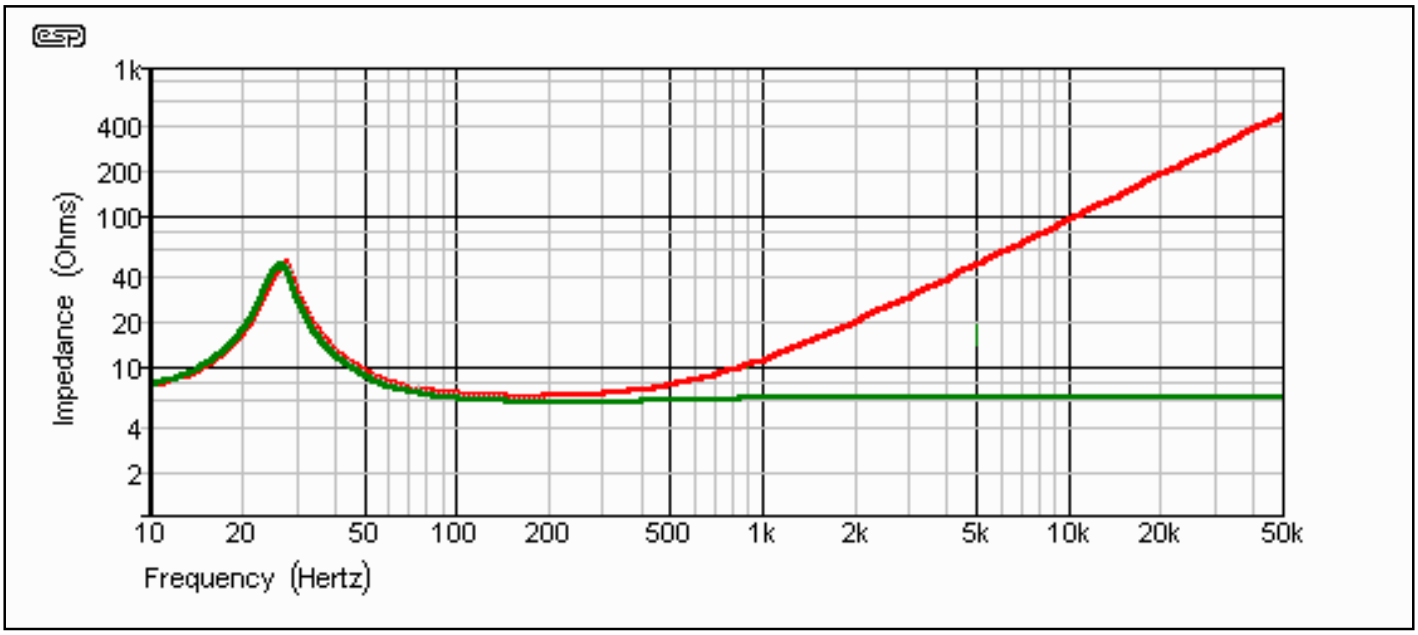


Figure 3.4 - Resulting Loudspeaker / Zobel Impedance Curve

The red trace shows the uncorrected impedance, and the black trace shows the impedance with the correction network in place. Naturally, it is important that the crossover is designed for the actual (rather than the nominal) impedance presented by the driver at the crossover frequency. An 8 ohm speaker will rarely be 8 ohms in reality, and this is especially true when an impedance correction circuit is used.

As noted above, one could also design the crossover for the impedance actually presented at the crossover frequency, but this varies with frequency. Unless the impedance remains reasonably constant for at least 2.5 octaves (and preferably more) above and below the crossover frequency, the network cannot be expected to provide a predictable response. As a direct result of this, a 6dB/octave passive crossover at (say) 300Hz is barely acceptable, because of the impedance peak of the woofer. Although it is possible to equalise the woofer's impedance peak, the components needed will be very large (electrically and physically), and will be very expensive.

If the loudspeaker is installed in a vented enclosure, then there will be two impedance peaks to equalise out - this can become very tedious, and will be costly to implement. There is usually little to be gained by equalising the woofer impedance peaks, and if an electronic crossover network is used, there is absolutely no reason to do so.

When you have completed the network, connect it (and the speaker) to the test setup shown in Figure 1.1 and measure the response. It should be quite flat (within 1dB) up to the highest frequency of interest.

3.1.2 Midrange and Tweeter Drivers

Most tweeters and midrange drivers can benefit from using a compensation circuit at their resonant frequency when a passive crossover is used. This is especially true if you are using

a crossover network with a slow rolloff, or the frequency is too close to the resonant frequency of the driver. With a 6dB/octave filter, I suggest an absolute minimum of about 1.5 octaves between the driver resonance and crossover frequency. A tweeter with a 900Hz resonance should therefore be crossed over at a minimum of 2,500Hz, but preferably higher. If you use the minimum possible frequency separation, there will be a small peak at tweeter resonance - this is a combination of the tweeter's resonance itself, and the fact that the crossover cannot maintain the correct rolloff if the load impedance changes.

Although there are allegedly formulae to calculate the values needed to make a network whose impedance is the exact opposite of the resonance peak, I suggest that they are of minimal use in practice. One I have seen requires the Thiele/Small parameters, and these are rarely available for tweeters in particular. One could measure the parameters, but the effort of doing so is equal to (or greater than) the effort needed to experiment with a few selected values. With experience, this will become quite easy - albeit a little tedious.

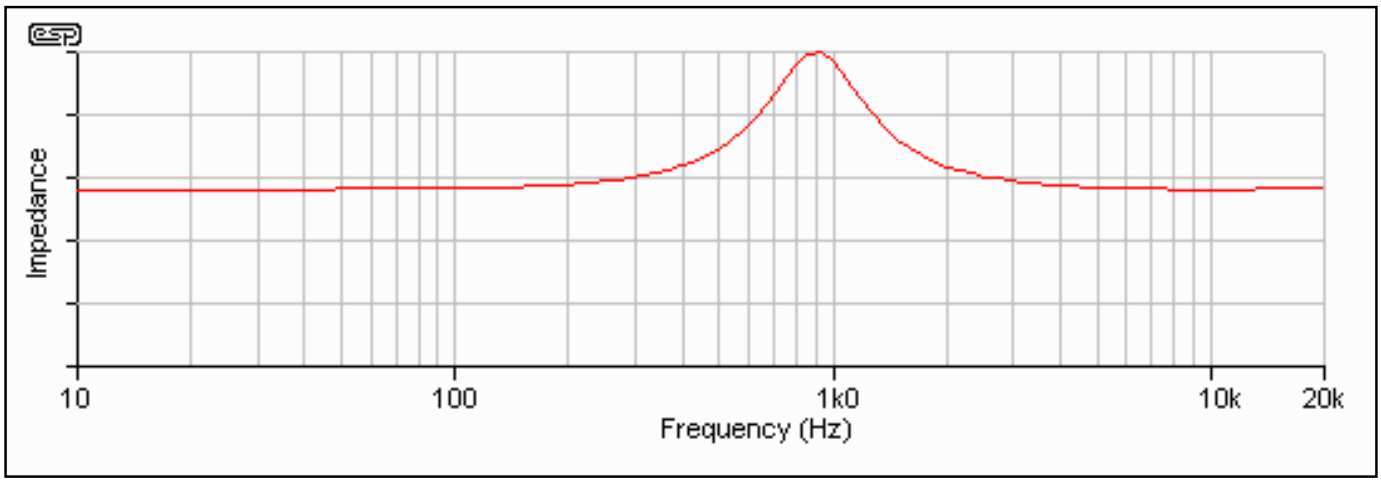


Figure 3.5 - Tweeter Resonance

Experimentation is made easier with the following procedure. The first requirement is to plot the impedance of the tweeter or midrange driver, which should be done with reasonable accuracy. The resistance used in the network will be equal to the DC resistance of the voice coil - that much at least is quite straightforward. With the procedure explained below, I eventually arrived at the network shown in Figure 3.6 - this is hardly a trivial circuit to implement, especially with a relatively large capacitance. Again, the capacitor does not necessarily have to be "audiophile" quality, so a bipolar electrolytic could be used in this circuit as well - but the same caveats apply as with the woofer inductance compensation circuit. A bipolar electrolytic will change value over time, and the compensation circuit's performance will deteriorate as capacitance is reduced with age.

There is a strong case for manufacturers of midrange and tweeter drivers to offer a compensated version of their drivers, which would simplify the process considerably. At the very least, the needed parameters should be supplied to allow us to calculate the values needed. Regrettably, I have never seen the parameters or a suggested circuit with the

specifications for any tweeter - IMHO this is the very least the manufacturers could do to help us all out on this. Instead, we are left to our own devices to determine the network by trial and error. I find this to be somewhat irritating.

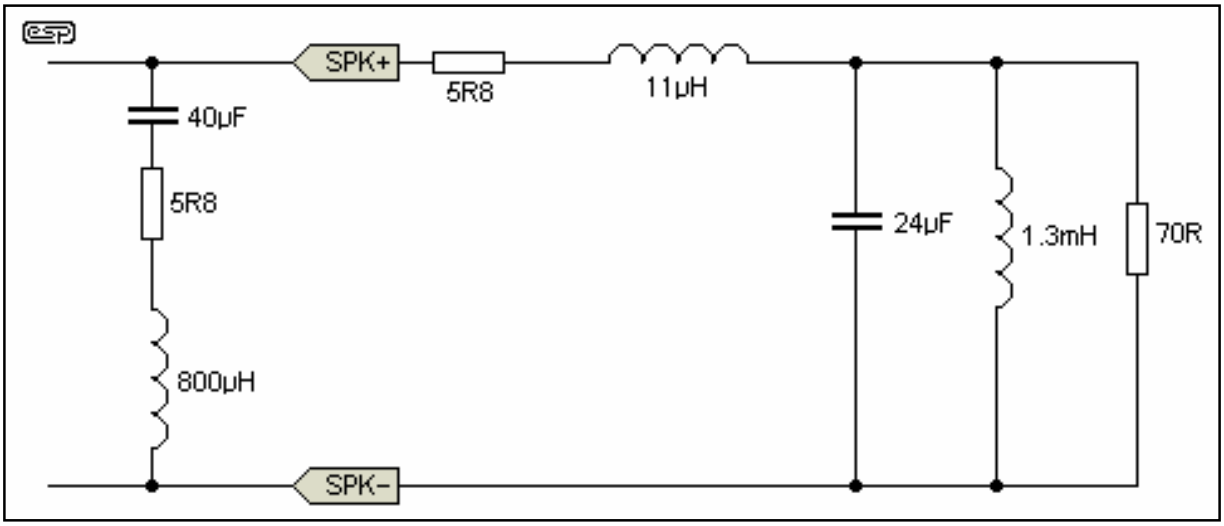


Figure 3.6 - Compensation Circuit, and Equivalent Circuit of Tweeter

Unfortunately, it is extremely difficult to get the required details for most tweeters. The resonance is almost always quoted, but most of the time it is difficult to find the voice coil inductance, let alone the Qts and Vas, which would enable one to calculate the required network. The values I used in the simulations are assumed, and gave a resonance at about the right frequency - the reality will (of course) be different, and will differ even further from one tweeter to another. Some basic measurements on a selection of tweeters confirmed that the simulation is not too far off, although it may be a little broader than some tweeters.

The Q of the compensation circuit must be the same as the resonance Q, or it will simply form a sharp notch in the middle of the resonance peak (Q too high) or a broad notch that spans the resonance (Q too low). Unfortunately, this is not as easy as it may first appear, but it is not actually difficult once you know what to do.

As a first approximation, find the (actual) resonant frequency of the tweeter. You will need a small amplifier, and a resistance of about 10 ohms. Wire the 10 ohm resistor in series with the amplifier output and the tweeter. Keep the output voltage as low as possible (less than 1V RMS). Change the frequency of the audio oscillator slowly and note the frequency where the voltage directly across the tweeter terminals is at its maximum. Reduce the frequency slowly, until the level has decreased by 3dB (i.e. 0.707 of the previous level). From this, you can calculate the required capacitance that will null the inductive component of the resonant peak.

Measure the DC resistance of the voice coil, and note the resonant (Fo) and -3dB (F3) frequencies - you will need both for the next steps.

Capacitance and inductance are calculated from the formulae ...

C = 1 / (2 * π * Rvc * F3)
L = 1 / (4 * π² * Fo² * C)

The details of the simulated tweeter are ...

Rvc = 5.8 Ohms
Fo = 907 Hz
F3 = 635 Hz

Substituting our tweeter, we will obtain the following ...

C = 1 / (2 * π * 5.8 * 635) = 43uF
L = 1 / (4 * π² * 907² * 43 * 10^{-6}) = 716 uH

These values will be difficult to obtain, and substituting 40uF and 800uH caused such a small error that it is of little consequence (it actually improved matters very slightly).

Note that this procedure is intended as a starting point only, and you will almost certainly need to experiment if you want the flattest possible impedance. Small variations will not cause significant errors, so it is not beneficial to go to extremes.

Again, when you have completed the network, connect it (and the tweeter) to the test setup shown in Figure 1.1 and measure the response. It should be quite flat (within 1dB) across the resonance frequency.

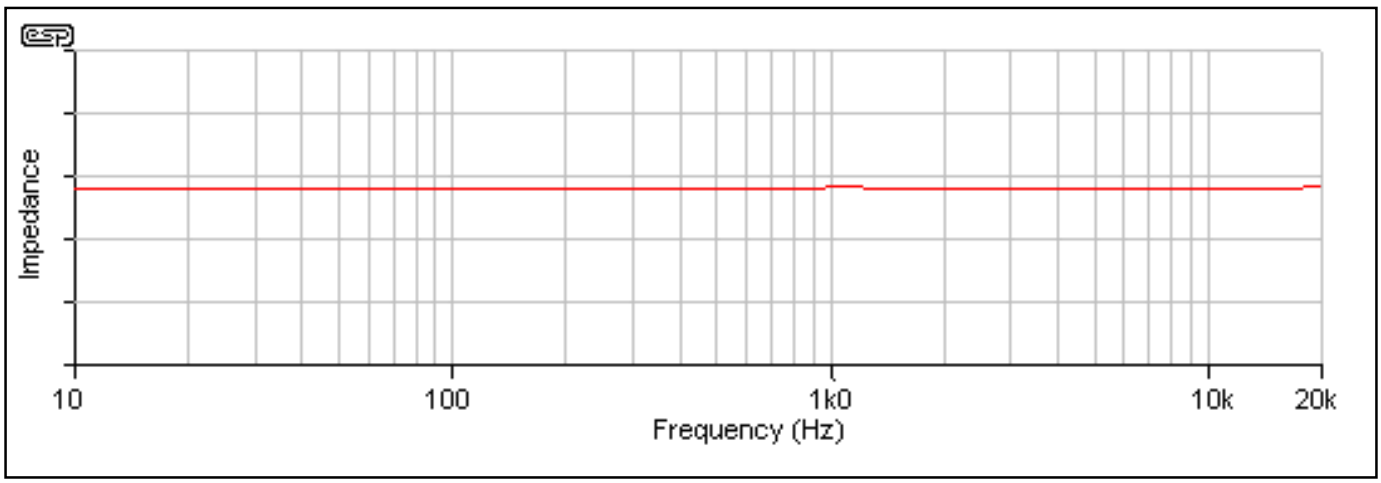


Figure 3.7 - Tweeter Impedance With Correction Circuit

The impedance is now commendably flat, but at the expense of overall impedance, which is reduced from the nominal 8 ohms to the real impedance of the tweeter - in this case, about 6 ohms. This figure must be used when the crossover network is designed - ***not*** the nominal 8 ohms impedance. This is a very common mistake made by the novice, and from what I have seen, quite a few professionals as well. The nominal impedance is just that - nominal. When

designing the crossover, the actual measured impedance must be used - ***always!***

Figure 3.8 shows the result of the crossover with and without the correction network. This was simulated with a 12dB Bessel filter, and as you can see, there is a profound difference. This is the AC amplitude response only - the actual frequency response (dB SPL) will be quite different in most cases, but will reflect (to a modest degree) the problems that are immediately apparent from the diagrams.

An uncompensated (or partially compensated) tweeter can actually be used in a 12dB filter with little audible difference, provided there is sufficient "distance" between the crossover frequency and resonance. However if you are going to go to all that trouble to build the "ultimate" speaker system, you might as well go the extra mile (kilometre?) and get it right.

Ferro-fluid damped tweeters will usually have an impedance that changes so little that compensation is not required. Regardless, the tweeter should be measured to make sure it will not upset the crossover network.

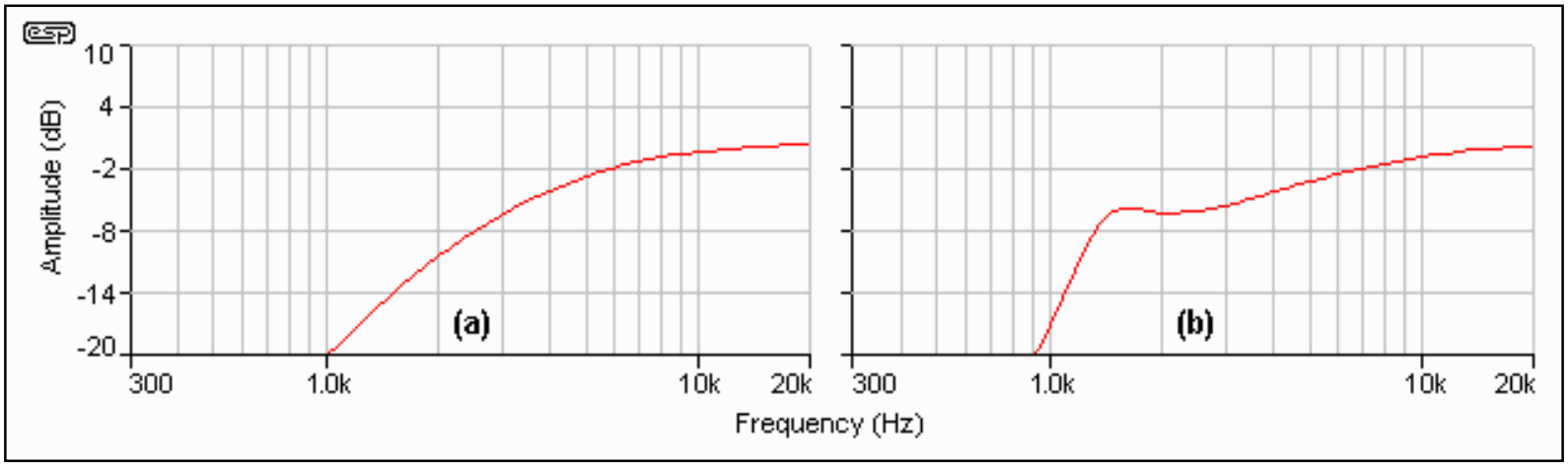


Figure 3.8 - 12dB Filter Response With (a) and Without (b) Compensation

The effect is much worse with a 6dB filter, and this is shown in Figure 3.9. Note the tweeter resonance - it is completely undamped, and the attenuation of a 6dB filter is obviously not sufficient to reduce power to a respectably low level, even though the crossover frequency is set to 3kHz.

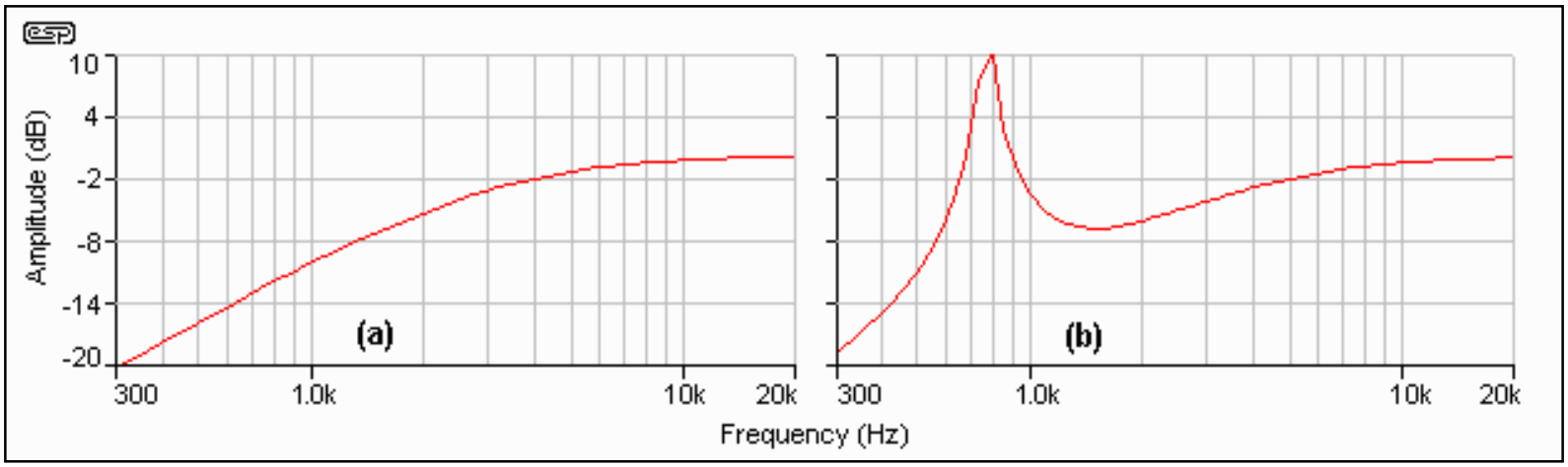


Figure 3.9 - 6dB Filter Response With (a) and Without (b) Compensation

As you can see in (a), the tweeter signal is down by only about 10dB at resonance, and the uncompensated version is very much worse. Somewhat surprisingly (somewhat ??), the audible effect is not as bad as these diagrams indicate, but there is no doubt that for a high quality system, the effects are there to be heard - you just have to know what to listen for.

3.2 Amplifier Impedance

If the speaker impedance has an effect on the crossover performance, then it follows that the amplifier's output impedance will also have an influence. With the majority of transistor amps, this is not an issue, but a valve amp is very different indeed.

Most valve amps have an output impedance that is at least a few ohms, and for the sake of the exercise, we will assume an impedance of 4 ohms.

In a crossover that was previously completely flat when driven from a zero ohm source (a 12dB 'sub-Bessel' filter with a Q of 0.5), there is a 1.5dB (approx.) peak at the crossover frequency when the source impedance is increased to 4 ohms. The effect is less pronounced than with a variation in loudspeaker impedance, but may be considered objectionable nonetheless.

Again, this is not allowed for in any speaker that I know of. If a speaker is to be able to be driven with either a valve or transistor amp, then a switch is needed to modify the crossover to suit the source impedance. Any claim as to "audiophile" performance is negated if the speaker cannot be accurately matched to the amplifier that will be driving it.

This also rather destroys the "audibility of cables" argument, since those who can "clearly hear a difference" between two equivalent quality cables, rarely seem to hear the peak in response that occurs when they use a valve vs. transistor amplifier. One of these effects is far more pronounced than the other - I shall leave it to the reader to decide which is likely to be more audible 😀.

*A word of warning is worthwhile here. **Never** operate an amplifier into a crossover network with the drivers disconnected. It may be tempting to look at the response, but at a frequency equal to the series resonant frequency of the inductor and capacitor, the network may present almost a dead short circuit to the amplifier (depending on the filter type - second order filters are the greatest risk).*

Current is limited only by series resistance, and dangerous voltages can be developed across the capacitor and inductor. These can be sufficient to damage the capacitor (due to over-voltage), and can give you a very nasty electric shock. The amplifier may not survive this abuse either, so it could be a very expensive temptation indeed.

With only 10V RMS applied at the resonant frequency of the 3kHz filter shown below (and assuming a total series resistance of 1 ohm), the amplifier will be supplying 8.3A RMS, and there will be 98V RMS across the inductor and capacitor. Provided the Zobel network is left in place, the resonance is damped so heavily that any risk is eliminated.

3.3 Temperature

In case you were wondering, the voice coil temperature used in the examples below (150°C) is not as outrageous as it may seem. Since loudspeakers have an efficiency of typically 1% or less, this means that 99% of all the power going to the speaker must be dissipated as heat. Although there is some air movement through the voice coil gap, it cannot keep the temperature down low enough to ensure that the effects described will not disturb the behaviour of the crossover network. An efficiency of 1% indicates just over 92dB/m/W, which is quite a respectable figure in the world of loudspeakers!

Copper has a thermal coefficient of resistance such that its resistance increases by 0.395% per °C (variously listed as anything from 0.39% to 0.43% on different websites). We can safely assume that the impedance is based on "room temperature", which will generally be in the order of 20°C. When power (in the form of music or test signals) is applied to a speaker, the voice coil temperature must rise. Given a typical 6.6 ohm (DC) voice coil for an 8 ohm nominal speaker, at 150°C, the resistive component alone will rise to 10 ohms - and naturally the impedance must be somewhat greater than this figure.

The loading on the crossover network is then radically different from the basic design figure of 8 ohms, or any corrected impedance obtained by using a Zobel network. The real issue here is that there is virtually *nothing* you can do about it, so the loading on the crossover network will vary depending on how loud the speakers are playing!

You may have read reviews where a loudspeaker system was described as becoming "edgy" or "brittle" at higher levels. This may be because the amplifier was clipping, but it could also be the result of what is called "power compression". Woofers are particularly susceptible to this phenomenon, since they are expected to handle more power than tweeters. Many

tweeters use ferro-fluid in the voice coil gap, which not only improves damping and reduces resonances, but also gives a higher power handling. This means that the voice coil temperature rise will often be much lower than for the woofer, so the relative efficiency (in dB/m/W) from the tweeters remains the same, and that of the woofer falls (as it must if the impedance increases).

Of course, some speakers may have been optimised for a higher than average listening level, and these are likely to sound somewhat dull at "normal" listening levels.

That the impedance varies with power level is not conjecture, it is a fact, and the effects can be proven in demonstrations and by measurement - this variation is as dictated by the laws of physics, and no speaker manufacturer has been able to break those laws.

> The question remains - what can be done about it?
> The answer (regrettably) remains - virtually nothing!

It would be possible to use a thermistor (a special resistor whose resistance varies with temperature), but matching its thermal characteristics to the voice coil would be a formidable task. Since a thermistor is by definition a non-linear device, it may also introduce distortion of its own - a less than desirable outcome.

It is not only the power compression of relative levels that has an effect in this case - the crossover frequency will shift with volume! It really has no choice, since the voice coil will change impedance, and the crossover frequency (and filter damping) is determined by the load - the loudspeaker, and any additional network you use to equalise its response.

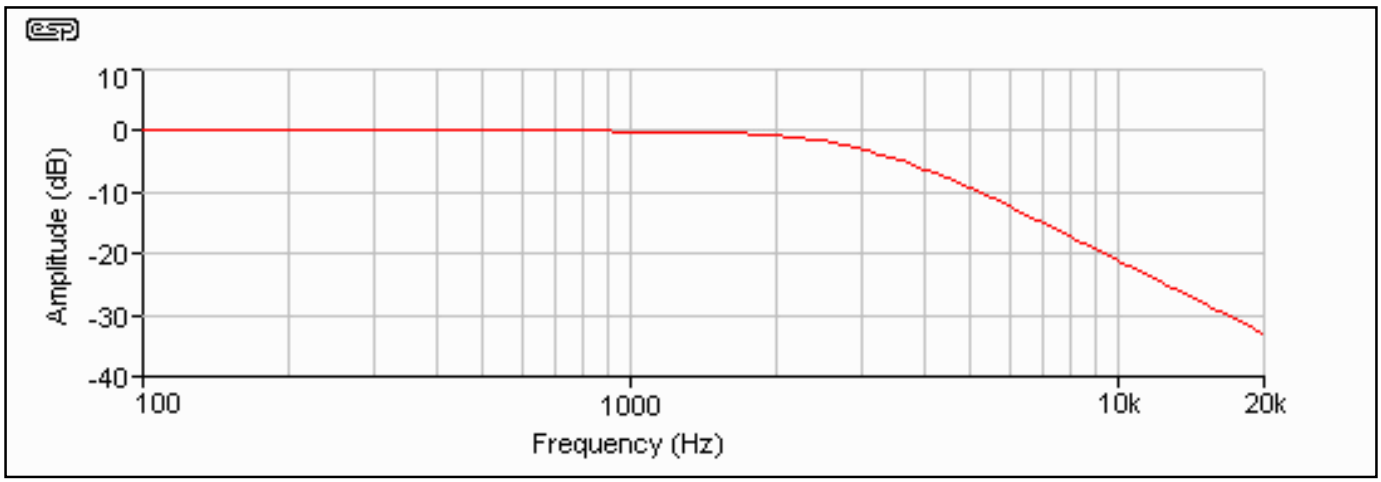


Figure 3.10 - Filter Performance at Ambient Temperature (Z = 8 Ohms)

As an example, let's assume that a loudspeaker has a flat impedance curve, and is exactly 8 ohms. A Butterworth crossover may be designed that will have a -3dB frequency of 3kHz (Figure 3.8). Should the impedance rise to 11 ohms (as in the above example of a 150°C voice coil temperature), the -3dB frequency will increase to over 3.8Hz, and the filter shape is

changed. A Bessel filter will become Butterworth, or a Butterworth filter will now be Chebyshev! This is shown in Figure 3.11, and the typical peak before rolloff of a Chebyshev filter response is easily seen - as is the frequency shift.

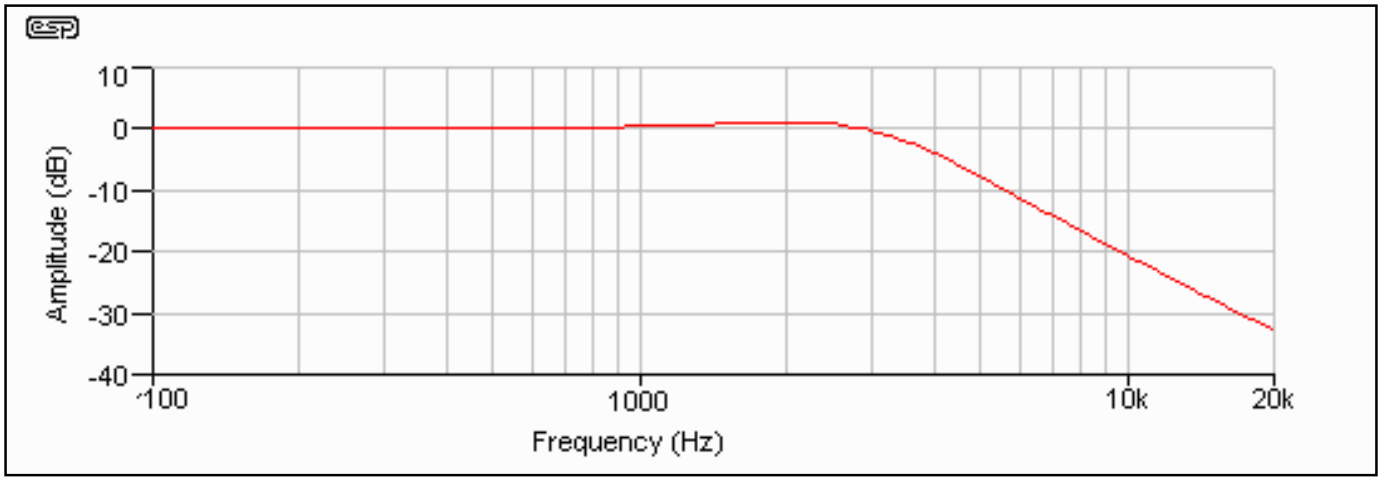


Figure 3.11 - Filter Performance at Elevated Temperature (Z = 11 Ohms)

Meanwhile, because of the relatively low power in the tweeter, and possibly due to the effects of ferro-fluid, its temperature rise may have caused the impedance to rise to only 9 ohms (example only). Even this is sufficient to cause the crossover frequency to fall to about 2.7kHz, and will also change the filter shape (but to a lesser degree).

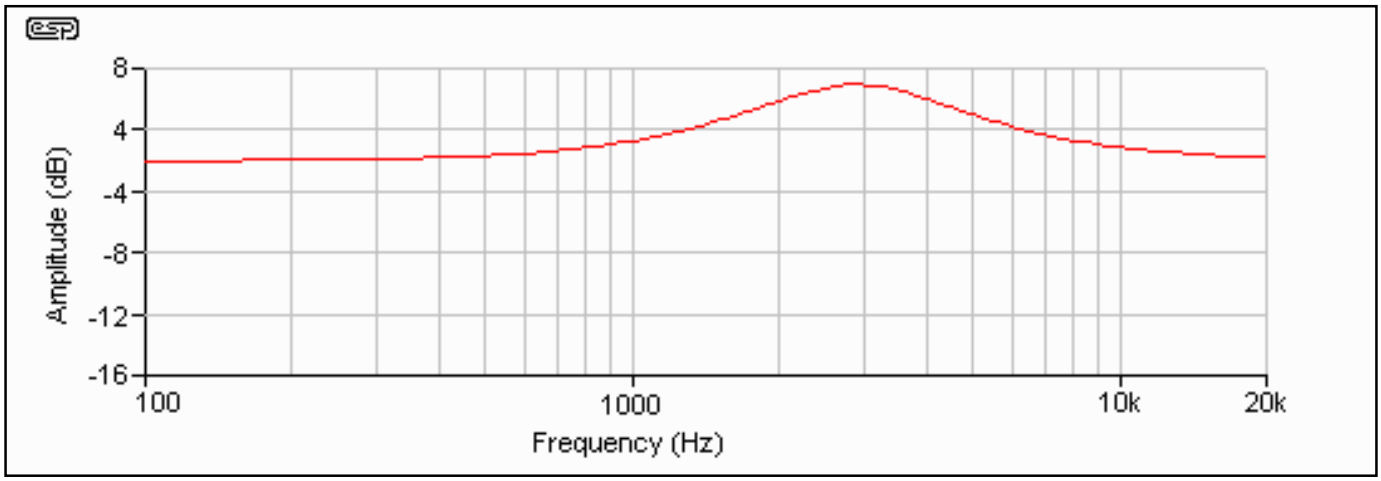


Figure 3.12 - Combined Crossover Response at Elevated Temperature

A crossover that at ambient temperatures had a nice stable (and well defined) crossover frequency of 3kHz, now has a 1.1kHz overlap! The audible effect is a disaster, since there will be quite a prominent peak (almost 5dB) at around 2.8kHz - the approximate centre frequency of the overlap region. Needless to say, if the tweeter had an equivalent temperature rise (due to the use of an aluminium voice coil, for example), then the overlap region becomes greater, and the peak is worsened - the amplitude will be relatively unaffected, but the peak will be wider as the two speakers are reproducing the same frequencies. The effects on relative phase and dispersion are less predictable, but we can safely assume that the outcome will be undesirable! (To put it mildly.)

A further consideration (although probably of minor consequence except in extreme circumstances) is that the tweeter will now absorb a little more power, causing its voice coil temperature to rise further, thus lowering the crossover frequency and allowing more power, thus raising the temperature even further (etc.). Whether this would reach destructive levels is doubtful - but the possibility exists! An amplifier pushed into distortion for extended periods could conceivably lead to this self destructive behaviour (assuming that the tweeter survives the high average power to start with).

This behaviour could be made to disappear completely when an electronic crossover is used. Although the details are not relevant to this article, a signal processor can be adapted to adjust the level to compensate for power compression in the drivers. To my knowledge this has never been done. The truth is, this is far less of an issue when an electronic crossover is used - the power compression still occurs, but there is no shift of the crossover slopes, so the effect is only on relative levels, and not the designed crossover frequencies. It is a simple matter to adjust the relative outputs of the electronic crossover to match the average power that will be used, so one could have a switch, marked "Soft", "Normal" and "Loud" if desired :-)

3.4 Atmospheric Changes

The loading on a loudspeaker cone, and therefore its Thiele/ Small parameters, will also vary with changes in the atmospheric conditions. High humidity, altitude or temperature make air less dense - such variations will cause changes to the loading on the cone, and thus the speaker's parameters.

In reality, these are relatively small, except at extremes. Even at the extremes, the physiological effect on the listener will probably be far greater than the atmospheric effects on the loudspeaker, but I know of no tests that have been performed to measure the changes experienced by any driver with differing atmospheric conditions.

I doubt that there are vast differences, but it is an additional consideration worthy of further investigation - preferably by someone with access to an environmental test chamber. Relying on the vagaries of the weather and quickly taking some measurements is unlikely to yield meaningful results.

Fortunately, it is unlikely that the performance of a passive crossover will be affected to any audible degree by normal variations in atmospheric conditions.

4.0 Selecting the Filter Slope, Alignment & Components

Having digested the above, you now bravely decide to go ahead regardless. The next task is to select the filter slope and alignment. Again, there are compromises that must be made, and

it is important to select the most appropriate crossover to suit the drivers you are using.

4.1 Slope

Selecting the best slope is important, both to protect the tweeter (in particular), and to ensure that the drivers are all operated within their optimum frequency and power handling ranges. A first order (6dB/octave) filter has the most predictable response, and is affected less by impedance variations than higher orders. On the negative side, the loudspeaker drivers will be producing sound at frequencies that are very likely outside their upper or lower limits. At low powers (less than 10W or so), this is usually not a major issue, but it becomes much more important when amplifier powers of 50W or more are considered.

Second order filters (12dB/octave) are better at keeping unwanted frequencies out of the individual speakers, but are more complex, and are affected by impedance variations to a much greater degree. The tolerance of the components used will also have a greater effect, so it is not uncommon for designers to make the inductors specifically for the job, rather than attempting to use "off the shelf" coils. The capacitance used must also remain predictable and constant over time and power, which specifically excludes the use of bipolar electrolytics (well apart from any other failings they may have - either real or imagined).

The design task becomes more complex and the tolerances more exacting as the order is increased. A third order (18dB/octave) filter requires closer tolerances than a second order, and is again even more susceptible to any impedance variations than the 12dB filter. Fourth order (24dB/octave) increases the complexity and tolerance requirements even further - a point must be reached where the requirements versus the complexity and sensitivity will balance out. With passive crossovers, I now believe that anything over 12dB is a waste of time, especially when the effects of voice coil temperature are considered.

Even with the second order filter, the possible variations (especially those caused by voice coil temperature) can totally ruin the sound - regardless of the quality of the components or care in making the crossover. I shall leave it to the reader to determine for him/her self where to draw the line.

4.2 Filter Alignment

The traditional passive crossover is (and for the most part always has been) the Butterworth - at least for second order filters and above. Although this may seem the ideal, it is not, since there is a 3dB peak at the crossover frequency when the outputs are summed. It is now commonly accepted that this peak is also present in almost all speaker systems when the loudspeaker outputs are summed acoustically - i.e. in normal operation. Many (reputable and/or higher priced 'audiophile') loudspeaker manufacturers will modify the crossover to compensate for this effect, but most of the formulae you find on the Net (and even in books and magazines) will simply use the nominal impedance of the drivers, and work out a pair of conventional Butterworth filters. IMO, this is not usable as a crossover for high quality reproduction, but it remains firmly entrenched regardless.

Again, this is a world of compromise. My preference is for a sub-Bessel alignment with a Q of 0.5, since it provides a close approximation to a Linkwitz-Riley alignment, and has zero peak or dip at the crossover frequency. Since the Q is lower, it is also marginally less sensitive to variations in loudspeaker driver impedance, but this is not something that should be relied upon.

As I mentioned above, a first order filter has a Q of 0.5 as a matter of course, and this cannot be changed.

Second order filters have an overall phase reversal, and this must be accounted for. In a two-way system, the tweeter is usually connected out of phase - the negative terminal is connected to the "positive" or hot output of the filter network. In the case of a three way system, the midrange driver is most commonly reversed in phase, with the woofer and tweeter connected normally. This maintains the overall phase integrity of the crossover for all drivers. A deep notch is created at the crossover frequency if the phase reversal(s) is/are not done properly - this is objectionably audible!

4.3 World's Worst Passive Component

Inductors

It is worth pointing out that inductors are, in general, the worst passive component imaginable. This is particularly true for use in crossover networks. Because it is not possible to use a core material without introducing audible distortion, the coils used for crossovers are nearly always air-cored. This means that many more turns than might otherwise be needed must be used to get the needed inductance, and that means either a very large, heavy and expensive coil, or a smaller and lighter coil with significant resistance.

Because inductor coils use magnetic coupling, they are sensitive to stray magnetic fields, or any source of variable magnetic flux. This includes cross-coupling from other inductors in the network, or even speaker magnet flux modulated by cabinet vibration. The latter is unlikely in a well constructed enclosure, but it is sensible to keep the coils well away from strong magnets.

Coils also have inter-winding capacitance, and this causes them to have a self resonant frequency that may fall within an amplifier's pass-band ... although rarely if ever within the audio band. The possibility cannot be discounted that an amplifier may be confronted by a very low impedance at some frequency determined by the inductors in the system.

The biggest problem is resistance. Some people will spend a huge amount on 'special' low resistance cables, and/ or an amplifier with a very high damping factor. A typical loudspeaker

crossover inductor will undo all of that instantly, adding perhaps 0.1 to 0.5 Ohm resistance in series with the bass driver (which is the very one that supposedly needs maximum damping). Resistance also causes power loss, and heat. For a crossover inside the cabinet, the last thing needed is another heat source!

So, the world's worst component is the inductor, with those used in crossover networks generally being the worst of the worst. Adding a core to reduce losses simply increases distortion (usually dramatically), so there is no easy way out (other than to use a fully active system, of course ... hint, hint).

Because of the losses, inductors should be wound with the highest gauge wire that you can - within reason of course. To prevent mechanical noise (rattles, buzzes), it is worth impregnating the finished coil in varnish. Soak the coils in a suitable varnish for an hour or so before draining and drying. The varnish doesn't need to be electrical grade, because the voltages are small and the final temperatures should never get high enough to cause thermal damage.

To avoid unwanted interactions, inductors should be mounted at right angles to each other (see below for more on this), and should also be located at a 'safe' distance from ferrous (iron/steel) materials. What is a 'safe' distance? That depends on the size of the ferrous material, and large pieces (such as loudspeaker driver motor assemblies) should be considered potentially hostile. A (measured) 460uH air-cored coil I tested increased to 480uH at 10mm from a smallish sheet of thin (0.8mm) steel plate, so I would expect a minimum distance of 50mm would probably be acceptable in most cases. This inductor has a DC resistance of 0.38Ω, and this gave a dissipation factor of 0.24 according to my meter.

Capacitors

By comparison to inductors, capacitors are positively benign. Their series (internal) resistance is usually extremely low, and self-resonance is influenced more by lead length than by the component itself. Even high loss capacitors will dissipate far less power than the best low loss inductors. There are some capacitors that should be avoided though, most notably bipolar (non-polarised) electrolytics. When they are new, they work very well, but if they handle appreciable current they will lose capacitance (and gain distortion) over time.

In nearly all domestic systems, all film caps will be quite alright - there is no real need to insist on film and foil types unless there are very high currents involved. Metallised film caps will usually have more than sufficient current carrying ability for systems rated at up to a couple of hundred Watts, and often a lot more. Much has been made of the dielectric material, but this is generally wishful thinking, perpetuated by a fringe area of audio that insists that no reasonably priced component can ever sound any good. In many cases, you may *pay* the top price for polypropylene, but get polyester anyway, and few people know how to tell the difference between them.

Capacitor Type	Typical Temperature Coefficient
Polyester (Mylar®)	600 to 900 ppm/°C
Polypropylene	-200 ppm/°C
Polystyrene	-125 ppm/°C
Polycarbonate	+100 ppm/°C

Table 2 - Capacitor Dielectric Characteristics

From the above table, you can see that only polypropylene and polystyrene dielectrics have a negative temperature coefficient, so when heated (a hot air gun is one easy way), their capacitance will fall slightly. Polyester and polycarbonate will show an increase in capacitance when heated. All capacitors used in crossovers should be located away from any extreme heat source.

A perfectly ordinary 2.2uF polyester cap measured a dissipation factor or 0.02 - vastly better than the inductor. Although this can be improved upon, the change is unlikely to be audible in the vast majority of cases.

Because capacitors have a small variation with temperature, it is sensible to ensure that they are well separated from any resistor expected to get hot in normal use.

Resistors

Resistors are again benign, although they will always contribute heat if dissipating any power. While non-inductive resistors are available and are recommended, the error introduced by a normal (slightly inductive) resistor will typically be far smaller than the normal production differences between supposedly identical loudspeaker drivers. Any errors introduced will generally not be apparent within the audio band. The inductance of most power resistors is such that the wiring may introduce greater errors than the resistors themselves, given that each 10mm of (straight) wire adds about 5nH of inductance to the circuit.

It is important to ensure that the power rating for all resistors is well above (preferably double) the expected power or which they will be subjected. This is much lower than a full power steady state (sinewave) analysis might indicate, but it may be necessary to experiment a little during the final tweaking phase.

Naturally, any resistor that gets hot cannot be glued to the crossover board with hot-melt adhesive, and ideally should be clamped with a metal bracket to help dissipate heat and ensure that vibration cannot move the part - this may cause the lead(s) to eventually fracture.

5.0 And Now ... Some (more) Maths

The formulae for calculating the various filter component values are not at all complex, although they may appear so at first glance. There are quite a few variations, depending mainly on the slope and alignment. I have included those for 6dB and 12dB variations only, as I don't feel that there is anything useful to be gained by using higher order passive filters - especially in light of the discussions above.

Since it has been shown that the speaker impedance will rarely be 8 ohms - particularly when impedance correction has been applied - I will use 6 ohms for all calculations to follow. You will need to determine the exact impedance of your impedance corrected drivers yourself. It is unlikely that they will exactly correspond to my examples, but you might be lucky .

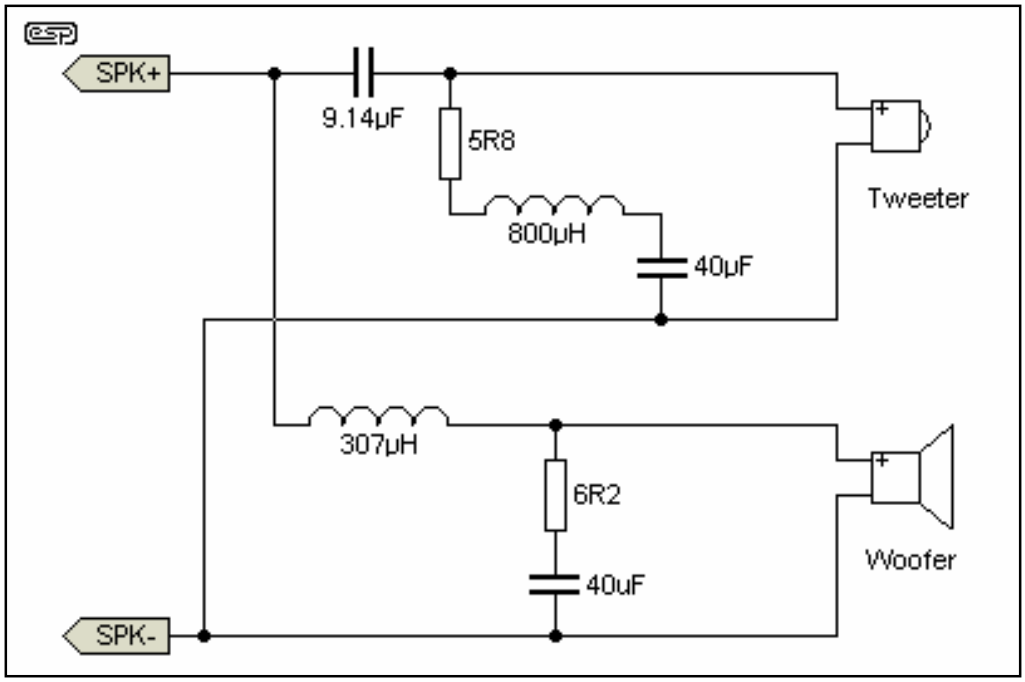


Figure 5.1 - 6dB/Octave 2-Way Passive Crossover

The first filter must be a 6dB two way crossover. This is about as simple as a circuit can get (except that it is actually quite complex when all the parameters are considered). Impedance correction components have been included for reference. I designed this filter using the same drivers used as examples above, and the crossover frequency is 3,000Hz.

$$C = 1 / (2 \pi Z f)$$
$$L = Z / (2 \pi f)$$

Where:

C = capacitance in farads
L = inductance of the coil in Henrys
f = frequency in hertz

Z = (actual) impedance of the speaker in ohms

These can be "simplified", and reduce to the following ...

C = 0.159 / (Z f)
L = (0.159 Z) / f

Thus, for a crossover frequency of 3,000Hz at 6 ohms (a standard I shall use throughout these examples) ...

C = 0.159 / (6 * 3,000) = 8.83 uF
L = (0.159 * 6) / 3,000 = 318 uH

The crossover frequency is the -3B point on the response curve, but since this is a sub-Bessel filter (having a Q of 0.5, or a damping of 2), the response is completely flat across the crossover point.

A schematic for a 12dB crossover using the simulated drivers as used above is shown in Figure 5.2, and again includes the impedance correction circuits. The schematic for nearly all conventional parallel crossovers is the same, only the component values change. Note that the component values have been calculated for the simulated drivers, and should not be used as shown - this also applies to Figure 5.1

Note especially that the tweeter is wired with its phase reversed - this is important, and must not be forgotten. This only applies to the 12dB example.

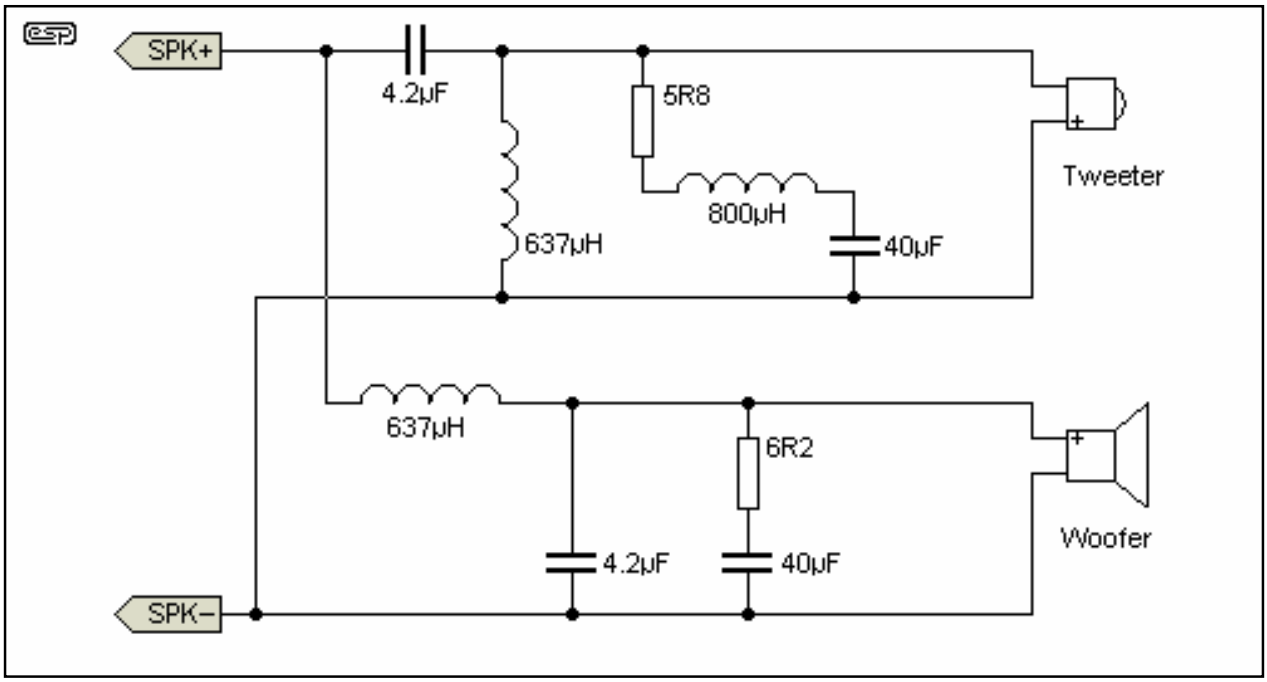


Figure 5.2 - 12dB/Octave 2-Way Passive Crossover

The same circuit is used for all calculations for 12dB filters. Using a 12dB filter network with a Q of 0.5 (which gives an approximation of a Linkwitz-Riley alignment) the following (simplified) formulae will determine the component values ...

C = 0.0796 / (Z * f)
L = (0.3183 * Z) / f

The derivation of these is marginally interesting, and will help you to understand the Butterworth and Bessel alignments a little better. The full original formulae are ...

C = 1 / (2 * π * Z * d * f)
L = (Z * d) / (2 * π * f)

Where ...

d = 1/ Q = 1 / 0.707 = 1.414 (Butterworth)
or ...
d = 1 / Q = 1 / 0.5 = 2 (sub-Bessel)
d = 1 / Q = 1 / 0.57 = 1.75 (Bessel)

A sub-Bessel filter (for our purposes) has a Q of 0.5 or a damping of 2. Note that this has nothing to do with amplifier 'damping factor' which is completely different, and as long as it exceeds about 10, has no influence on crossover performance (although woofer performance may require the amplifier damping factor to be higher than this).

6.0 Attenuation Networks

It is rare that the woofer and tweeter (or midrange driver) will have the same sensitivity (i.e. efficiency). The woofer should have the lowest efficiency, since it will require the most power, and any network that reduces the level to the woofer will absorb a disproportionately high power, and will adversely affect the damping factor.

The driver selection is very important - ideally, all drivers will have the same efficiency, and no attenuation will be needed. In the real world, this will rarely be the case. Attenuator networks are a necessary evil - it is immeasurably better not to use them at all, but they cannot be avoided unless the drivers have *exactly* the same sensitivity.

For the purpose of the exercise, assume that the tweeter has an efficiency 2.8 dB greater than the woofer/midrange. This means that the level must be reduced by 2.8dB, or the speaker system will sound too bright. Remember that this network can alter the impedance presented to the crossover network, so you either must design for the impedance with the attenuator in circuit, or ensure that the attenuator presents exactly the same impedance as the speaker.

The very first exercise is to determine the resistive drop caused by the low pass inductor (this step is almost always forgotten!). A typical coil of this value, using 0.8mm wire, will have a resistance (Rl) of about 0.53 Ohm. We can calculate the low frequency loss in dB with the formula ...

dB = 20 log ((Rl / Z) + 1)

For our example, this gives ...

dB = 20 log ((0.53 / 6) +1) = 20 log (1.088) = 0.73 dB

We have now found that the woofer's sensitivity is slightly lower than before, so we need an attenuation of 2.8+0.73=3.53dB, which we can safely round down to 3.5dB. The tweeter must therefore be reduced in level by 3.5dB so that it matches the sensitivity of the woofer.

The attenuator must be placed either before the filter (basically not a good idea), or between the crossover filter and the driver - including any impedance compensation. The driver and its associated impedance correction network should be considered as one, and they should not be separated (unless you feel like re-calculating the entire crossover and compensation networks). I have seen a number of design examples that state that the attenuator should be before the crossover - *wrong, wrong, wrong!*

This practice increases power dissipation needlessly, since the attenuator must work over the entire frequency range. If attenuation is after the crossover, then power requirements are greatly reduced.

The simplest attenuator is a series resistor, but this changes the load presented to the crossover network. Unless the network is designed for the impedance presented by the combination of driver and attenuator resistor, this is unacceptable. As a result, the most common attenuator is an "L" pad. This is shown in Figure 6.1, and maintains an impedance of 6 ohms to the crossover, but reduces the tweeter level by 2dB.

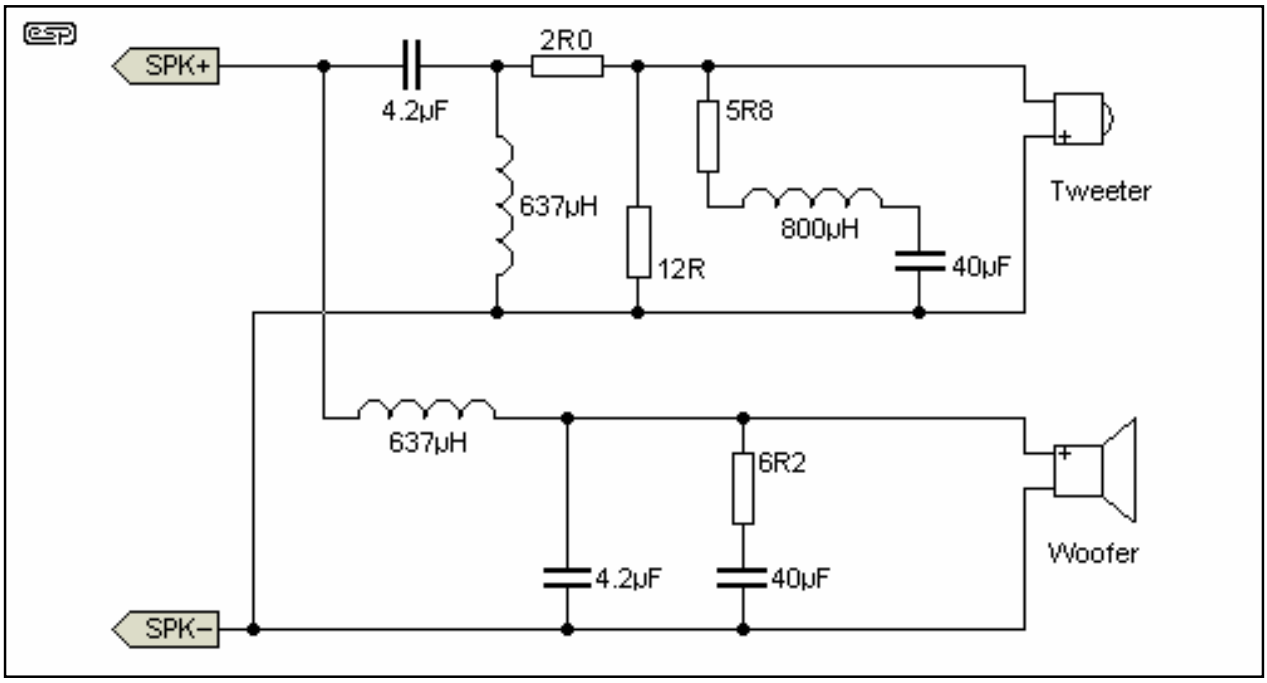


Figure 6.1 - 2dB L-Pad Attenuator

The calculations are quite irksome, and (as always) can be tiresome. The problem is that the resistances are interdependent - if one changes, then so does the other. The idea is to maintain the same impedance we had before, or re-calculate the crossover network. The latter is very much easier, but means that the high and low pass sections will no longer have the same component values. For the purposes of the exercise, I will maintain the crossover, and make the resistors provide the same 6 ohm load as before.

Essentially, there are two ways to calculate an L-Pad - the simple way and the hard way. I am going to use the hard way, because it is simpler! How can this be? If I give you a formula that just spits out the value, that is easy, but you will *never* remember the formula. On the other hand, if I show you how to derive the formula using simple ohm's law, then you will be able to use the basic method, and you will have a much better chance of remembering it.

As suggested, the tweeter in this example is 3.5dB more efficient than the woofer + inductor combination, so must be attenuated by this amount. Please remember that these are examples only, and your situation will probably be completely different. First, we need to convert 3.5dB into a voltage ratio (Vr) ...

> **dB = 20 log (V1/V2) = 20 log (Vr)**, so reversing the formula we get ...
> **Vr = 1 / (antilog (dB / 20))**

For this example, we have 3.5dB, so substituting ...

> **Vr = 1 / (antilog (3.5 / 20)) = 1 / (antilog 0.175) = 1 / 1.496 = 0.668**

The voltage presented to the tweeter will therefore be 0.668 of the input voltage. Since we

want to preserve the impedance presented to the crossover network (rather than redesign the rotten thing :-) this makes the calculations a little harder.

As I said above, it is not impossible to derive a formula for the pad, but it is more convenient to work it out the long way - largely because this provides a better understanding of the process. We will assume an input of one volt - simply because it is convenient to do so.

Firstly, we need to determine the current that will flow into the load (**Z**) ...

I = V / Z = 1 / Z

Now, find the voltage drop across the series resistor **Rs**, then the value of **Rs** ...

Vs = 1 - Vr
Rs = Vs / I

The process is quite simple so far. Now we need to determine the value of the parallel resistor, **Rp**. We already know that the voltage across the parallel combination of **Z** and **Rp** - it is equal to **Vr** (I told you that an input voltage of 1V was convenient, didn't I? . The value if **I** (current) does not change, so we can determine the current through **Z** and **Rp** easily, and then **Rp** itself ...

Iz = Vr / Z
Ip = I - Iz
Rp = Vr / Ip

Now, let's substitute all the values for the example into the formulae. As I said, this is a little tedious, but easily remembered. Recall does not come from rote learning, it comes from *understanding*, and this is just simple arithmetic and Ohm's law.

I = 1 / Z = 1 / 6 = 0.1667 Amps
Vs = 1 - Vr = 1 - 0.668 = 0.332 Volts
Rs = Vs / I = 0.332 / 0.1667 = 1.99 (2.0) Ohms

That was easy enough, so now for Rp ...

Iz = Vr / Z = 0.668 / 6 = 0.111 Amps
Ip = I - Iz = 0.1667 - 0.111 = 0.0557 Amps
Rp = Vr / Ip = 0.668 / 0.0557 = 11.99 (12.0) Ohms

Now we might want to check that the values really will give us what we wanted - I recommend this final check, because there are resistor values that are easily created (or are standard), and we want to use these if possible. As a result, we will substitute 2R (2 x 1R in series) for

Rs, and 12R for Rp, as these are standard values. The first thing we need, is to determine Rt - the total parallel combination of Z and Rp (Z || Rp). We could do that from the current calculated earlier, but that may re-introduce any error made earlier.

Rt = 1 / (1 / Rp + 1 / Z) = 1 / (1 / 12 + 1 / 6) = 1 / (0.0833 + 0.1667) = 4 Ohms
Vd = (Rs / Rp) + 1 = (2 / 4) + 1 = 1.5
dB = 20 log(Vd) = 20 log (1.5) = 3.52 dB

Damn! That was close

An error of less than 0.1dB is completely insignificant, and can be ignored completely, but in this case, we got almost *exactly* the attenuation we determined right at the beginning. In case you were wondering, this was not deliberate - it just turned out that way. If you are careful with your calculations, it will always turn out this way.

The resistors should be wirewound power types, and the actual power is determined by the input power from the amplifier.

7.0 Determining Power Losses

There will always be power losses in a passive system - this is generally referred to as "insertion loss", and all resistors and inductors will create power loss and thus, heat. Capacitors will generally contribute very little loss, and will not get hot - a potentially notable exception being bipolar electrolytics. This is another very good reason not to use them, but the main reason is that their value will change over time, and will upset the crossover frequency.

The power loss is naturally proportional to the input power, and for our example, I shall assume a maximum amplifier power of 100W. Use the chart below to determine how much power will go to the tweeter, using a crossover frequency of 3.0kHz.

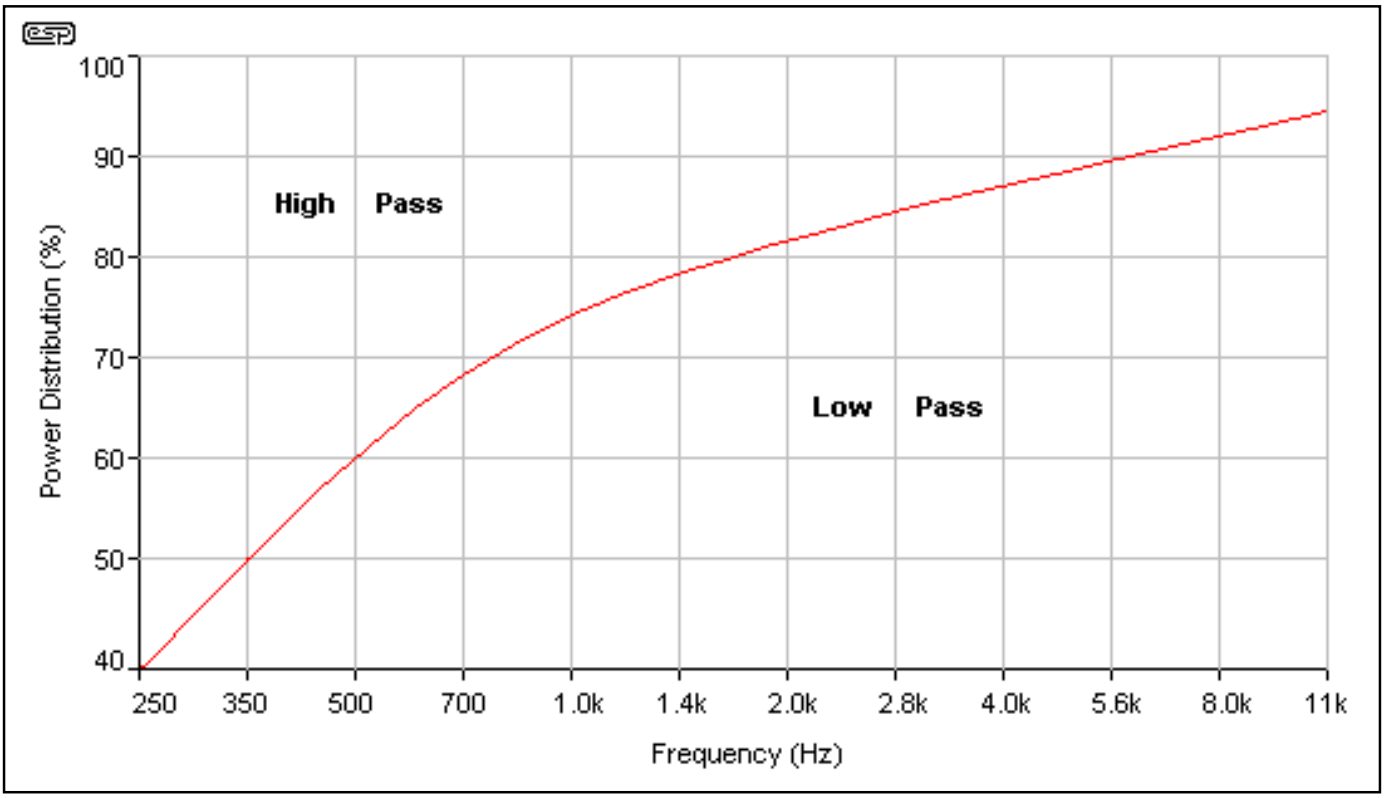


Figure 7.1 - Power Distribution Chart

Working along the frequency axis, we see that at 3kHz, the power in the low pass section will be about 85% of the maximum (85W), so the high pass power level is about 15%, or 15W for our 100W system. We are now in a position to work out some power ratings for the resistors, and can also calculate the inductor losses. Not that we can do anything about losses in inductors, but we can at least decide whether they need to be mounted on fire proof material (just kidding).

We already know that the woofer's Zobel network only needs to operate at over 650Hz, so using the chart again, we see that the low pass section will get about 65% (65W) at that frequency. The Zobel network is a high pass section (yes it is!), so the maximum power will be at above 650Hz, where high frequency energy is less than 35W. Since the power progressively reduces, the resistor will never have to deal with more than about 20W peak. Since no-one (well, no-one who is going to go to all this trouble to make an almost perfect crossover) will listen at 100W continuous, we can safely assume an average power of about 10W - this corresponds to the "typical" peak to average ratio for music of 10dB.

Woofer Zobel resistor ...
I would suggest a power rating of 10W for the Zobel resistor - this provides a very large safety margin.

Tweeter resonance compensation ...
The tweeter power at resonance is more than 20dB down from the maximum level using a 12dB crossover. Since this is the case, a 5W resistor is more than adequate.

Tweeter L-Pad ...
The L-Pad will be subjected to a maximum of 15% of the power, but will dissipate very little of this. 5W resistors are again more than enough to handle the power.

Woofer inductor ...
Since we determined that the resistance will be about 0.53 ohms, so at full power it will dissipate less than 10% of the 100W input, which is about 10W. (0.53 ohms is 8.8% of 6 ohms) The average will be much less than this, and heating will not be a major problem.

Note that all the losses above are wasted power, so in all, only about 90W of the amplifier power will ever get to the speakers themselves, and about 10% will simply be wasted as heat - inside the cabinet! This will eventually cause the air inside the box to heat up enough to change the characteristics of the enclosure - especially sealed boxes. This is added to the heat generated in the loudspeaker voice coil, a good proportion of which will remain in the cabinet, having no means of escape.

Perhaps you would be better off either mounting the crossover network on a heatsink on the outside of the box, or install it in a false (ventilated) base. And yes, I am serious. Apart from the heat (which is actually relatively low compared to that generated in the voice coil), the high sound pressure will cause vibration of the components. While I think that this is unlikely to be audible if everything is well fastened, there are claims from some quarters that microphony at this level is definitely audible. For the small additional effort, external mounting is recommended. It also make minor tweaks and adjustments much easier, since you don't have to remove speakers to get at the crossover.

8.0 Winding The Coils

Using the drivers simulated in this article, and using all the networks that were devised along the way, we have examined the complete 12dB/octave crossover network implementation. The tweeter is 2.8dB more efficient than the woofer (not allowing for the additional inductor loss), and is padded back with an L-Pad that reduces the signal while maintaining the impedance.

Use the examples above to work through your own crossover design - the final result is quite complex, and will not be inexpensive to build. This is the price to pay when the best possible performance is to be obtained from a passive crossover (now, wouldn't it have been easier to tri-amp the system instead?).

All that remains is to explain how to wind the coils you will need. It may be possible to obtain them commercially, but I doubt it, since the values are unlikely to match those you will find at most electronics outlets. You may be able to get coils that are slightly above the values you

need, and remove a few turns until you get it right.

An inductance meter is essential for any of this - many are available in (even relatively cheap) digital multimeters, and these should be sufficiently accurate. It will be virtually impossible to get the coils right without a meter - you can measure their resonant frequency with a known capacitance and calculate inductance from that, but it is very tedious!

If you want to wind your own coils, I suggest you use either of the following the online calculators, available at ...

Shavano Online Music - Inductance Calculator
Barry's Inductor Simulator

Both of these are very much easier than using the formula below. You will still need to measure the final result to make sure that it is within a reasonable tolerance. Much as it might appeal, don't use iron or ferrite cored inductors for crossovers. There are advantages in that they are smaller and have less power loss, but the distortion and risk of saturation (at which point inductance drops dramatically) are not worth it (IMHO).

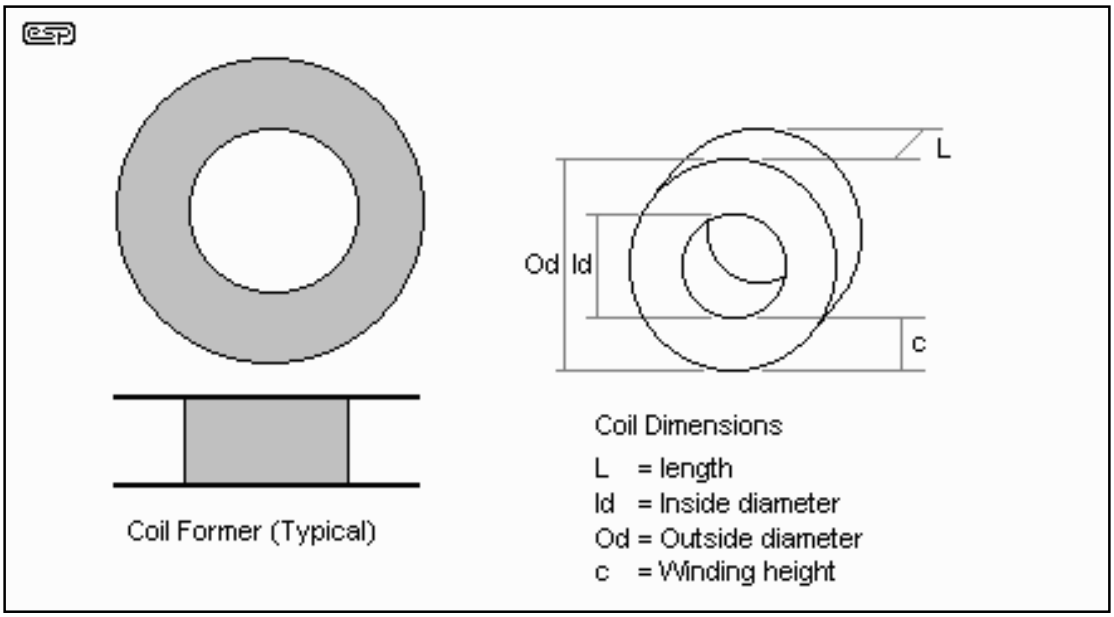


Figure 8.1 - Typical Coil Former and Dimensions

From the above, the necessary dimensions used in the formulae can be easily determined. These apply to the winding itself, not the former (which may be a temporary affair, as it is only needed while winding the coil). There are actually several different formulae that may be used - all are empirical, and require some experimentation to arrive at the correct value (or the correct number of turns for a given inductance and wire size).

Some I found are not worth the paper they are written on - they either don't work, or work only under certain limited circumstances. The formula below is due to Wheeler ("Simple inductance

formula for radio coils" 1928), and is still probably the most accurate so far (which is scary - a formula over 70 years old has never been surpassed). The original (using inches (mutter, mutter)) is ...

$L = 0.8 * a^2 * N^2 / (6a + 9l + 10c)$ uH

Where

N is the number of turns.
a is the average radius.
c is the height of the windings
l is the length of the coil.

All dimensions are in inches. This is easily converted in a spreadsheet or program, but modifying the formula itself is too tedious. For a real example of a coil wound with 0.83 mm (20 AWG) wire, having a design inductance of 637 uH and a resistance of 0.53 ohms has the following dimensions ...

N = 99 turns
l = 11.2 mm (0.44")
Id = 44.8 mm (1.76")
Od = 58 mm (2.28")
c = 6.64 mm (0.26")

These figures were arrived at from one of the Shavano online simulators, and when the data is plugged into the very basic spreadsheet I have done so far indicates that this coil will have an inductance of 635.17 uH. This is a very small error, and will be of no consequence in practice. According to the simulator, power handling is 180 Watts, and it will require approximately 15.9 metres (52.3 feet) of coil winding wire.

Never use plastic coated wire for winding inductors - enamelled winding wire is essential. In most cases, it is probably simpler to buy ready made coils (of higher inductance than required) and remove turns to get the exact value you need.

When the inductance is correct, soak the coil in varnish for an hour or so (most clear floor varnishes are quite acceptable), drain, and let it dry thoroughly before use. This ensures that the turns cannot move - we don't want to add vibration sensitive coils to the already suspect passive crossover. It will also prevent rattles - of the type that will drive you nuts, because you will have no idea where (or what) is rattling! Vacuum impregnation is nice, but few constructors will have the necessary equipment. If desired, coils may have potted centres to facilitate mounting, but you *must* use nylon or brass screws - steel screws will increase the inductance significantly! Brass screws may decrease the inductance ever so slightly, but the error is unlikely to be significant.

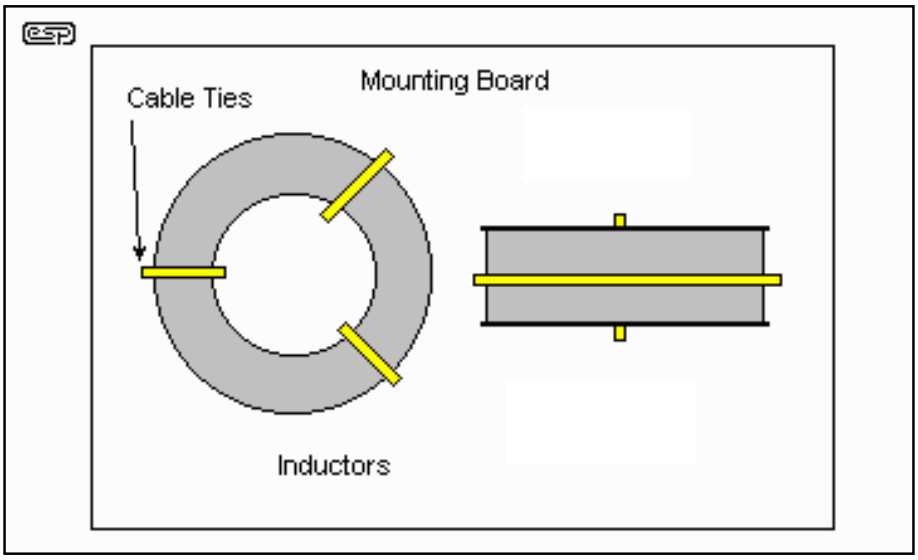


Figure 8.2 - Mounting Inductors to Minimise Coupling

When mounting the coils, keep them well separated, firmly attached to the mounting board, and ensure that there is minimal mutual coupling by placing the axes at right angles to each other as shown in Figure 8.2 - transformers in the crossover you ***don't*** need. Use cable ties and silicone, hot-melt or epoxy glue to make sure the inductors are firmly fastened to the mounting board and cannot move. Any movement will eventually fracture the copper wire if it is rigidly attached, so a loop of wire from the coil (or stranded wire leadouts) will ensure that you don't have electrical failures.

It has been pointed out by a reader that one can economise on the inductor wire if the finished product is going to be installed with a series resistance (to equalise impedances, or provide a level of correction for example). Simply using a thinner wire will make the inductor smaller and cheaper, and if you are very lucky may eliminate the resistor altogether. One must be careful that the coil resistance is the same or lower than any external series resistance - if resistance is too high, there is nothing you can do to lower it again.

9.0 Conclusion

This article has covered the topic in far more detail than you will find in most references, and explains some of the things that most articles don't even touch upon. One thing that should be quite clear by now, is that a full 3-way passive crossover, with everything done properly will be ***very*** expensive to build. It is also time consuming, and the final result will only ever be as good as the effort you are willing to put into getting everything right.

A few generalised recommendations are in order ...

- Sub-Bessel filters are to be preferred for flattest overall response (Similar response to Linkwitz-Riley alignment and easier to design)

- There is a vast amount to be gained by using a biamped system to cover the bass and mid+high crossover. Keep passives for higher frequencies, where their bulk, cost, power loss, and other flaws are minimised.
- A really well designed crossover is of no use if the box is not designed correctly, is inadequately braced, or has drivers mounted equidistant from two or more edges - these cause high frequency refractions that "smear" the stereo image.
- Likewise, voice coil (time) alignment can make a huge difference to the linearity of the system as a whole.
- The crossover will never compensate for a poor selection of drivers, regardless of the work you put into it.
- Some very simple crossovers may appear to give a more "musical" sound reproduction, but are not accurate - in the long term, they cause listener (and driver) fatigue, and are adding things to the music that was never there in the first place.
- Your listening room has more effect on the sound than any of the other points made above! However, a good system has a much better chance of sounding acceptable in a bad room than a bad system (which will sound bad everywhere!).
- A fully active system (using electronic crossovers and separate amps for all drivers) will almost certainly give a better result than the most carefully designed passive system, and may even work out cheaper ... Some passive crossover networks can become very complex and expensive indeed.

A Quick Rant

The "simple" passive crossover is actually vastly more complex than is commonly believed. The "new" Diaural™ "inductor only" crossovers are ***not*** a panacea for the ills of the world (despite *massive* marketing hype to the contrary), but fall into the 'overly simple and grossly coloured' category. There is nothing (repeat - nothing!) about the Diaural system that is new, or will benefit the vast majority of systems. They will probably sound "lively" and perhaps "musical" by initial direct comparison to a conventional crossover, but require very careful driver selection indeed if gross response and phase errors are to be avoided. These are even patented - how in God's name those ratbags got a patent on something that has been done by others for years, we will never know. I have a copy of the patent, and all the variations are shown - so much for the "cone of silence" that was placed on anyone who saw them in the early days. There is absolutely nothing remarkable about the principle, other than the complete and total neglect of almost everything I covered in this article. Especially noteworthy is the fact that they use ... inductors, which as discussed above have more compromises than any other passive component.

So there we have it. I doubt that this is the "last word" on passive crossovers, and I'm equally sure I have left out something that should have been included. I don't profess to be an expert on the design of passives, but my design background and experience (as well as that of a few others) has helped in the analysis of the loudspeaker driver behaviour, in the electrical sense at least.

The mechanical behaviour is something else again, and many excellent papers have been written on the subject. In particular, I suggest Lynn Olson's series of articles (of which I was made aware just as I had almost completed this paper). These are not (IMO) definitive articles, but have sufficient good information to make them recommended reading. Some readers may be mystified that I would include 'subjectivist' material in my reading list, but the fact is that some of the points made are just too important to ignore, and are not often measured by hobbyists and even some manufacturers. That the effects are measurable is not in doubt here ... intermodulation and harmonic distortion are easily measured, but no manufacturer of loudspeaker drivers is going to do so until this is demanded by their customers.

Lynn Olson ...

- Looking Over My Shoulder, Parts I and II
- The Soul of Sound

The material presented is mainly to do with cone break-up effects, but some mention is given to impedance correction as well. There are obviously many others, and an enormous amount of research has been done by a vast number of people, all striving for the same thing - the perfect loudspeaker. We don't have it, and may never get it - but when (or if) we do, there will always be someone who says it stinks. Such is Hi-Fi.

References

[Compendium of facts and recommendations]

Most other reference material is either based on general knowledge, or is gathered from a wide range of sources (including ESP articles) too numerous to mention.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 20 May 2001 Updated 09 Aug 2004 - added material, and corrected some small errors./ 04 Nov 2003

- corrected attenuator calculation./ 17 Jun 2002 - added spreadsheet and impedance info./ 08 Oct 05 - Included component selection info, table 2, etc.

The Audio Pages

Elliott Sound Products **Phase, Time and Distortion in Loudspeakers**

Copyright © 2001 - Rod Elliott (ESP)

Page Created 30 May 2002
Updated 09 Apr 2005

Articles Index

Main Index

Contents

- Introduction
- Time Delay
- Crossover Filters
- Distortion
- Phase Audibility
- Conclusion
- Updates
- References

Introduction

One only needs to look at a few web sites to realise that there is actually very little useful information on phase in audio systems in general, and loudspeakers in particular. There are a great many conflicting claims and counter claims, but little real data. There is naturally a great deal of rubbish, mostly describing why "Brand X" loudspeaker (for example) is demonstrably superior to every other speaker on the planet (which is why no-one has ever heard of them). Expect to see claims that "this speaker is the only design that will accurately reproduce a square wave" or something similar. As we shall see, this is realistically possible, but has (or should have) a huge "who cares" factor that will be

discussed in greater detail a little later.

This article is not for the faint hearted, as it discusses amplitude, phase and delay, and the complex interactions between them. There is only so much that can be accomplished by diagrams and graphs, and many of the concepts do not lend themselves to easy analysis. I have tried to keep the information in a logical form, but unfortunately, all of the things discussed occur simultaneously. This is not always easy to visualise, and is even harder to write.

The many diagrams and graphs were produced using SIMetrix, an excellent simulator available from Catena (formerly Newbury Technology) in the UK (Catena Software). It is available as a free demo system, and is the best simulator I have used so far. All images are reduced for speed - to see the full version, just click on the displayed image.

Since I am going to be using a 6 dB/ octave filter for many of the examples below, Figure 1 shows the response of a conventional 1st order (6 dB) filter. This is normalised to 1k ohm and 1uF, which gives a crossover frequency of 159 Hz. Although most discussions will revolve around higher frequencies, this is of no consequence. The graph is designed to show the rolloff slopes of the high and low pass sections - not the absolute performance at any specific frequency.

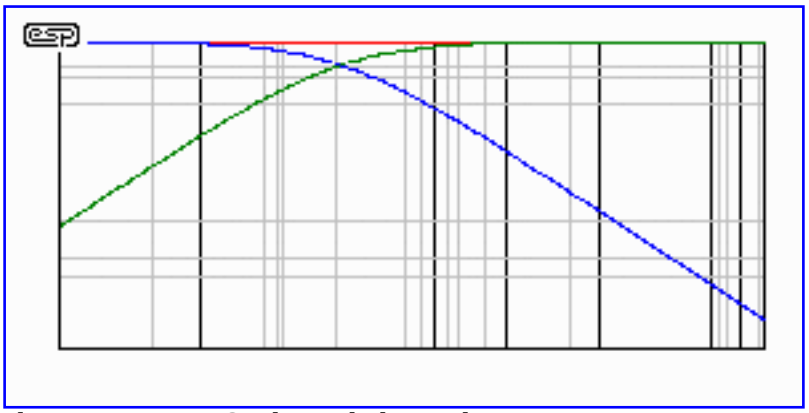


Figure 1 - 1st Order High and Low Pass Response

Note that *all* graphs on this page are shown as reduced size. For the full sized version, click on the image - this uses Java Script to open a new window for the full sized graph.

The line labelled "1-Input" is the input applied to the filter, and is also equal to the electrically summed outputs from the two sections. Both are flat, and are at 1 Volt relative level. Note that the simulator claims slightly different -3 dB frequencies for the two signals - this is not a simulator aberration, but the result of the simulator calculating to the absolute limits of accuracy. The crossover frequency is in fact 159 Hz as calculated, and at that frequency, the level is exactly 0.707 volt. If expressed accurately, -

3 dB is in fact 0.7079, and not 0.707 as is commonly used. This is a small error, and may safely be ignored.

All filters come with some pretty rigid rules - these are determined by the laws of physics, and are not open to discussion, although some of the snake oil vendors will still try. Filters are described in "orders" - 1st, 2nd etc. Each order has an ultimate rolloff (i.e. achieved at some point distant from the cutoff frequency) that increases by 6 dB steps for each successive order, so 6, 12, 18, 24 dB/ octave is a common way to describe the filter's response. They are further divided into "even" and "odd" order (even and odd numbers - it doesn't matter much, but is commonly used anyway).

A brief numerical description of each filter type is shown below, along with its rolloff characteristics and power level above the "cutoff frequency", typically defined as that frequency where the response is reduced by 3 dB. This is not always used as the crossover frequency - Linkwitz-Riley aligned crossovers use the -6 dB point instead, and achieve a flat response as a result (not applicable to 1st order filters).

Order	Slope	Voltage	Power	Theoretical
None	Flat	1 Volt	1 Watt	1 Watt
1st	6 dB/ octave	439 mV	193 mW	250 mW
2nd	12 dB/ octave	371 mV	138 mW	64 mW
3rd	18 dB/ octave	195 mV	38 mW	16 mW
4th	24 dB/ octave	122 mV	15 mW	4 mW

Table 1 - Filter Characteristics

Voltage in the above table is the voltage one octave above the -3dB frequency (assuming an input of 1 Volt and a low pass filter), and power at the same frequency, referred to 1 Watt. For example, 138 mW is about 1/7th Watt. The performance of a high pass filter is exactly the same as shown. The "Theoretical" value quoted is the power that should appear in theory - you may even see it quoted by manufacturers who have neglected to actually perform the maths, and have simply used the filter rolloff to arrive at a convenient looking number.

The above is not exhaustive, but it covers the filters most commonly used in audio. For all filters above 1st, the table is based on a sub-Bessel (minimum settling time) alignment

having a Q of 0.5, which is also typical of Linkwitz-Riley designs.

Time Delay

Firstly, there are many ways that the phase of a wave can be shifted, with the most common being time delay. At its most extreme, there is a delay of days to decades between the material being recorded and you listening to it - and no, this is not meant as a marginally humorous comment - this is a genuine time delay. The important thing is that all of the signal is delayed by the same amount, and it doesn't matter if this delay is measured in milliseconds or millennia, the sound will emerge intact and completely recognisable.

The situation is very different if some of the sound is delayed, while the rest is not. The listening experience would not be enhanced if the high frequencies were to be reproduced half an hour later than the bass or vice versa. This is quite obvious, but let's reduce the time to something more realistic. What if the treble were to be delayed by 20 milliseconds? The effect would be awful - this is a time difference we can easily pick, and we use these cues to determine the original sound from reflected sound for localisation.

We can continue reducing the time delay, and the effect will become less and less discernible as the time is reduced. Finally, we get to a point where the delay represents less than a wavelength (in air), and (perhaps surprisingly), the differences are still audible. Consider a 1 kHz sine wave, reproduced from two sources, but with one delayed by 500 us - just 1/2 millisecond. As one source creates a compression, the other creates a rarefaction - the waves are 180° apart, and will attempt to cancel each other. Early reflections and a multitude of other effects will ensure that we still hear the sound (at least at that frequency), but there will be a noticeable drop in level.

Now, there are some who will claim that reversing the phase of one source will bring everything back to where it was, so there is no harm done, and the net result is the same as if the two sources were not delayed at all. While this will obviously work at 1 kHz, at other frequencies this is not the case.

Now, let's look at some of the physics involved here. How would a 500 us delay be introduced in the first place? In reality, this is not uncommon, but we shall reduce the time delay to something more realistic before continuing. Any two loudspeakers that reproduce the same signal at the same time will exhibit this phenomenon, but for our purposes on a smaller scale.

If we look at a midrange driver and a tweeter, in the common vertical alignment in an enclosure, we have a time delay. The "acoustic centre" of the tweeter will most likely be a small distance closer to the listener than that of the midrange driver, and for the sake of this discussion, let us assume a difference of 50 mm, because it is a realistic and typical offset for common loudspeakers.

Before continuing, it is important that the concept of "wavelength" is properly understood. Sound travels at about 345 m/s in dry air at sea level. This changes with temperature, humidity and altitude, but we shall not concern ourselves with this, and there is little we can do about it most of the time. A sound at 345 Hz has a wavelength of 1 metre, at 34.5 Hz the wavelength is 10 metres, and at 3450 Hz, it is 100 mm. This is quite linear, and works for all frequencies. Another useful thing to know is the period (the actual time required to reproduce one cycle at the selected frequency).

wavelength = velocity / frequency
period = 1 / frequency

From the above, we can calculate the wavelength for any frequency we like. 3000 Hz has a wavelength of 115 mm, for example.

If we return to the midrange and tweeter mentioned above, their acoustic centres are offset by 50 mm - this is exactly 1/2 wavelength if the crossover frequency is 3450 Hz. We can account for the 1/2 wavelength by reversing the wires to the tweeter, so it is 180° out of phase with the midrange. The two drivers are now aligned in phase, so in theory, they are time aligned. Unfortunately, this is not the case. Although the signal is in alignment at the crossover frequency, it will not be aligned any more when the frequency changes.

What is really needed is to delay the signal going to the tweeter by 145 us (1/2 of the period of a 3450 Hz waveform), or align the acoustic centres of the two drivers in the vertical plane. Such "time alignment" is commonly achieved by angling the baffle so that at the listening position, the signals are properly in phase and time. Stepped baffles have also been used, but often create more problems with diffraction than are solved by the time alignment.

In short, time alignment is a good goal, but does not necessarily guarantee that the sound will be any better than a conventional flat baffle, with the phase of the drivers appropriately switched to ensure that the signal is in phase at the crossover frequency. It must be understood that with any flat baffle, an octave each side of the crossover frequency will see the phase out of alignment again, so it is essential that a high order

crossover is used to prevent unwanted cancellations and reinforcements at different frequencies.

With a flat baffle and a time displacement, above or below the crossover frequency the signals are in and out of phase - the exact amount can be calculated, and this can be very important in the greater scheme of things.

Octave	Frequency	Wavelength	Phase Angle
-1	1725 Hz	200 mm	90°
-1/2	2439 Hz	141 mm	45°
0	3450 Hz	100 mm	0°
+1/2	4878 Hz	70 mm	90°
+1	6900 Hz	50 mm	180°

Table 2 - Acoustic centre displacement 50 mm
(145 us time delay) 1 driver reverse phase

Expect a dip at an octave above the crossover frequency, since the two signals (from the midrange and tweeter) are 180° out of phase at this frequency - not because of the crossover, but because of the time delay of 145 us. The only way to ensure that this dip is inaudible is to use a steep filter! If a 6 dB/octave filter were to be used, the signal level is only down to 0.447 of the total (7 dB). On the other side, at 1 octave above crossover frequency, the tweeter will only have 0.894 of the full signal (0.97 dB down). These voltage relationships can be seen in Figure 1, above.

Hang on - this is a 6 dB/ octave filter, and it's 7 dB down an octave from crossover frequency. How can that be?

Remember that we are already 3 dB down at the crossover frequency, but because a 1st order crossover has a very low Q (or in other words is highly damped), the rolloff is not as steep initially as expected. It should be down by 9 dB an octave away, but this will never happen.

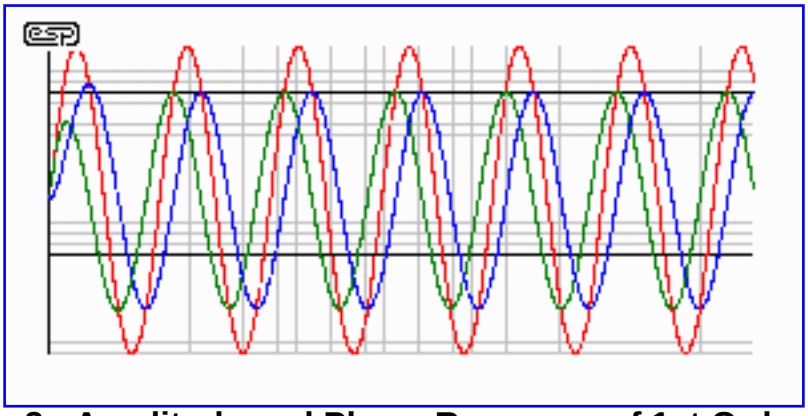


Figure 2 - Amplitude and Phase Response of 1st Order Filter

Ignoring the acoustic centres of midrange and tweeter for a moment, Figure 2 shows the waveform response of the filter at crossover frequency, together with the input signal. The RMS voltages of each are also shown. The waveform at 1 octave above crossover frequency is not shown - the absolute phase will be different, but relative phase (between outputs) remains at 90° for all frequencies. This is the electrical response only - the acoustical response will be different if the drivers are not time aligned!

Now (and this is where it gets tricky), what happens if we sum the electrical signals reproduced by the simple 1st order crossover? Assume an input of 1 volt for convenience. Adding 894 mV and 447 mV electrically (at any frequency) will give an output of 1.34 volts - this is clearly not correct, since the input is only 1 volt to begin with.

Analysis of a 6 dB/ octave crossover shows that the high and low pass signals are in fact 90° out of phase at all frequencies ("Yes but ... isn't the 1st order crossover supposed to be phase coherent?"). Yes and no. It is phase coherent in that all signals at all frequencies are 90° out of phase. I know that you have seen web sites that say that there is no phase shift through a 1st order crossover, but this is simply untrue! At crossover, the high pass section is leading - the signal appears to emerge from the filter 45° before the input, not possible it would seem. This sort of behaviour is standard with all filters with a "steady state" signal - you don't have to really understand it, so I suggest that you just live with it.

The low pass filter has a lagging response, so the signal emerges 45° after the input. This is easier to comprehend, but may still seem a little strange (which I suppose it is for a filter that many claim has no delays).

So, if we make the essential correction, and shift the relative phase of either signal by 90°, we can recalculate the summing of the two signals. Predictably, 894 mV + 447 mV with a 90° phase shift now gives a summed response of 1V - this is as we would expect, and is shown in Figure 3.

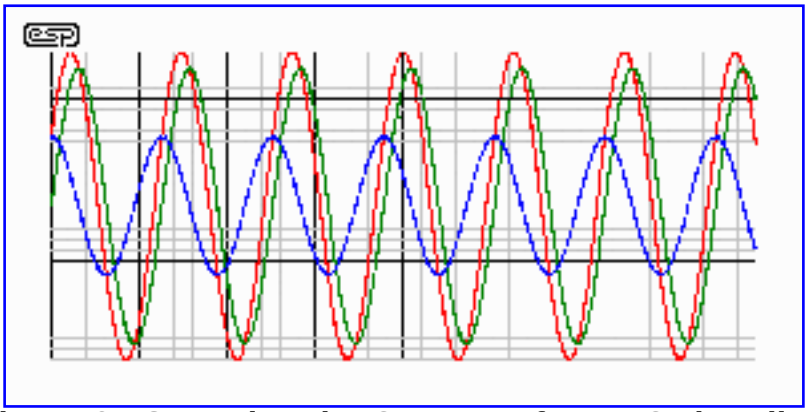


Figure 3 - Summing the Outputs of a 1st Order Filter

You can see the phase relationship between the 3 signals quite clearly. I doubt that this will be terribly meaningful for the most part, but it is essential to the understanding of the relationships - time and phase are inextricably entwined with each other, and cannot be separated.

The electrical and acoustical relationships only coincide if the acoustic centres of the speakers are in exact alignment. As soon as there is a misalignment (introducing a time delay), everything changes. To see the effect, imagine the original setup, with the acoustic centres misaligned by 50 mm. The tweeter's output will now be heard 145 us before that of the midrange. For the purpose of explanation, we shall ignore the 90° phase shift introduced by the crossover, and indeed, this is only present in the 1st order design. In fact, for many of the following explanations I will use signals of equal amplitude, and will ignore the crossover altogether. This provides for a worst case - reality will be somewhat tamer.

If we use two signals of equal amplitude, when summed we get a signal of double that of each signal - after all, the concept of 1 + 1 = 2 is not uncommon (except in corporate financial circles :-) If the level is any different, then there is phase shift (or delay) that causes the error.

Figure 4 shows what happens when the 3450 Hz signal is produced from both speakers simultaneously, but with a 145 us time delay (representing the 50 mm offset). The red line is the combined signal - there is no signal! This is electrical summing, which is much more critical than acoustical summing, so in reality we will still hear something, but nowhere near what we should. This is commonly referred to as a "suckout" by reviewers, and there will be a pronounced dip in frequency response. Now, we know that this is easily fixed by reversing the phase of one driver, and everything will be back where it should be - but (and this is the clincher here) - only at one frequency! At all other frequencies there will be interference effects, and the lower the filter order, the worse it

becomes.

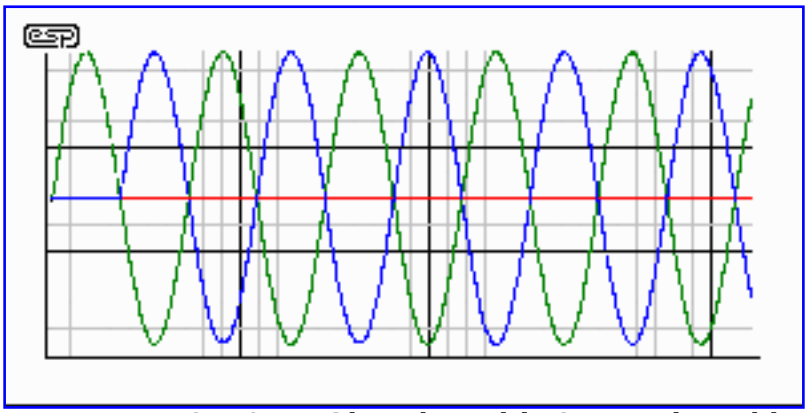


Figure 4 - Two 1V, 3450 Hz Signals, With One Delayed by 145 us

Rather than take vast amounts of bandwidth to display as whole series of similar waveforms, I have tabulated the resultant signal level below, for 2 signals of equal amplitude but with one delayed by 145 us. These are the same frequencies we looked at earlier. In all cases, the result should be 2 volts ...

Octave	Frequency	Amplitude
-1	1725 Hz	1.414 V
-1/2	2439 Hz	0.887 V
0	3450 Hz	0 V
+1/2	4878 Hz	1.195 V
+1	6900 Hz	1.959 V

Table 3 - Summed Signals

Now, bear in mind that the above table is actually meaningless (it looks impressive though). All of the information must be presented in a simultaneous manner for any of it to make real sense. To expand on this a little further, have a look at a frequency scan of two drivers reproducing the same signal, but with one delayed by 145 us. This produces a comb filter effect. Now, in real life, the signals will not be at the same amplitude, so the effect is reduced. The signals are also summed acoustically, reducing the effect even further, but the crucial point here is that the crossover and acoustical summing reduce - not eliminate - the problem. But this is still not real! (It is marginally useful though, just so you can see where all this is going.)

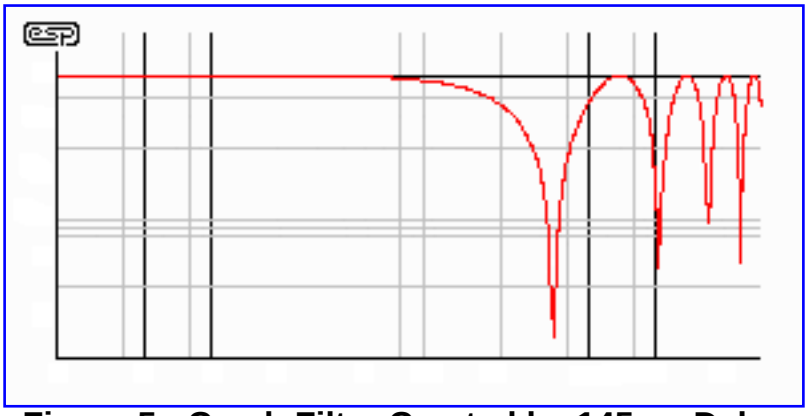


Figure 5 - Comb Filter Created by 145 us Delay

We can see the notch predicted in earlier examples at the crossover frequency of 3450 Hz, but we also see another at 10.26 kHz, and another at 17.4 kHz. The final notch shown is unlikely to be audible for most of us at 24 kHz. If the delay is increased, the effect becomes worse. It is also worth noting that even with the relatively small delay used for this example, the combined signal is down 3 dB at 1737 Hz. Remember that this is worst case, with no crossover network.

The combined effect of the delay and crossover can be expected to be a little less daunting, so the trusty simulator has been stretched a little here, and Figure 6 shows what happens when both the delay and the crossover are used, with the phase of one driver reversed as required to prevent the cancellation at crossover frequency. Oh dear! There might not be a major problem at the crossover frequency, but an 8 dB dip at 2 kHz is less than desirable. The 2.8 dB peak at 5.1 kHz is no bonus either. Less daunting? When all the material is presented, then the whole picture is available.

Note that this was missed in the table above, since I only looked at the 1/2 octave boundaries and with equal amplitudes. Little omissions can leave major gaps in ones actual knowledge! A small (cunningly disguised) trick of calculation or description can leave one thinking that a designer has achieved something special, so always make sure you have all of the information.

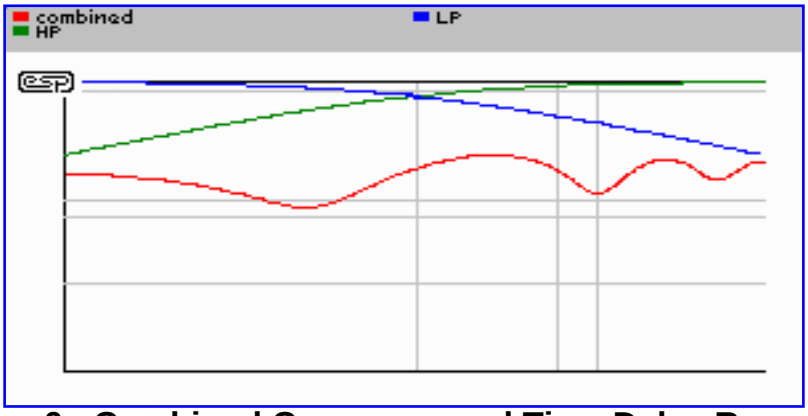
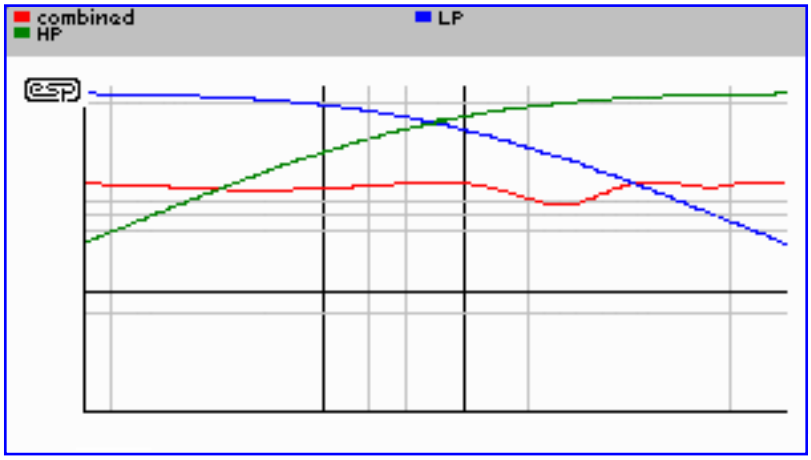


Figure 6 - Combined Crossover and Time Delay Response

The effect is not as severe (note the depth of the notches - read the voltage levels!), but in quite a few respects it is actually worse than the "fake" graph of the previous example!

Just to make sure, I reduced the time delay to 10 us then 1 us, to verify that nothing was awry with my simulations. As expected, the response was almost flat, and with no delay at all, the response was completely flat.

So, the next question has to be ... What difference does it make if the filter order is increased? Figure 7 shows the response with a 2nd order filter, using a Linkwitz Riley alignment. The ripples have been reduced considerably, but are still quite obvious. Figure 8 shows the response with a 24 dB/ octave L-R crossover. In both cases, the signal to the tweeter is inverted to account for the 145 us time delay, which as we know reverses the effective phase of the driver.



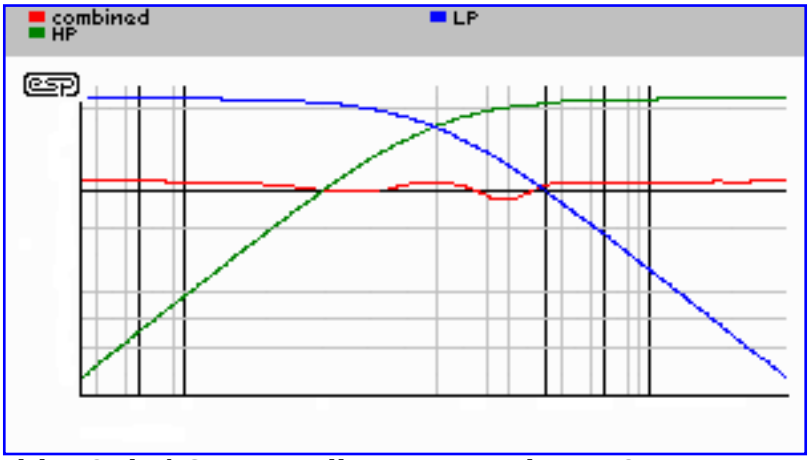


Figure 7 - Response With 12 dB/ Octave Filters **Figure 8 - Response With 24 dB/ Octave Filters**

As can be seen, the ripple is reduced as filter order is increased. Remember that all filters shown will sum electrically and acoustically flat if there is no time delay. All ripple is a direct result of the time misalignment. To put this into perspective, the room and

furnishings (including the speaker box itself) will have a much greater effect on the response than the 12 or 24 dB/ octave filters introduce - however, there is no good reason to muck up the response before the room has a chance.

Using DSPs (Digital Signal Processors), it is possible to delay the signal to speakers to compensate for the physical offset. At present, this is still frightfully expensive, but we can expect digital crossovers with adjustable time alignment delays to become commonplace in a few years. They exist now, but few of us can afford the luxury, and many will be unwilling to insert yet another set of analogue-digital-analogue converters into their system

---

Crossover Filters

I have always liked 1st order filters. Most loudspeaker drivers do not like 1st order filters. The ideal system would use no filters at all. With the partial exception of electrostatic loudspeakers (ESLs), the ideal speaker does not exist. Why "partial" exception? ESLs are bi-directional, and as a result of a relatively small baffle, do not reproduce low frequencies well. ESLs are also hardly a point source - the radiating panel is quite large, and this makes for a small "sweet spot" for listening because of the off-axis response of any large radiating surface.

As always, we must make compromises, and the ideal would be to have a single point source driver that could reproduce all frequencies equally well, and with no distortion. The smaller the driver, the better it will reproduce high frequencies without lobing, most easily described as listening angle dependent response peaks and dips. Low frequencies require that a lot of air be moved, so the small driver will do a very poor job - larger drivers are needed. This is the reason that most high fidelity speakers use at least two, and commonly three different loudspeakers to cover the audible range.

This is where the filters come into play - they are an essential part of the compromise, and separate the signal into ranges that can be accommodated by the individual drivers. The 1st order (6 dB/ octave) filter has the lowest phase shift and the best transient response of all the possibilities. It also has the slowest rolloff, so undesirable effects will be heard from the loudspeakers as they are excited by the signals outside their optimum operating frequency range.

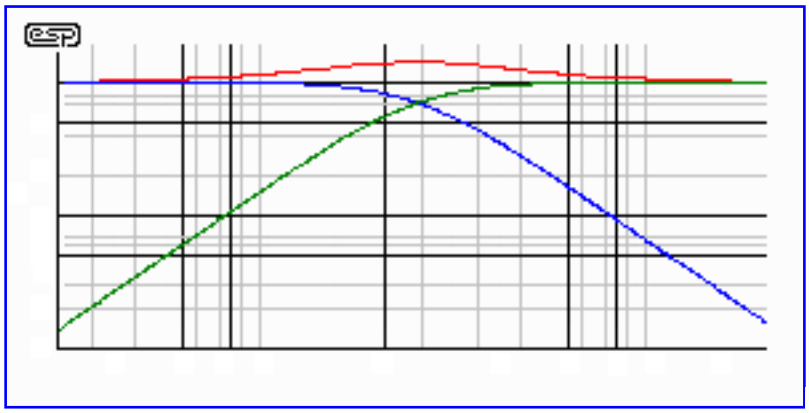
Contrary to what you may read elsewhere, all crossover networks (filters) *bar none* introduce phase shift. This is actually the least important characteristic of a filter, and provided that the low and high frequency waveforms remain in phase with each other,

their absolute phase is not important. Such a filter is described as phase coherent, and this is extremely important to the sound quality obtained.

Since filters introduce a phase shift, they also introduce a time delay. This is not the fixed delay referred to above, but varies with frequency. Perhaps surprisingly, this frequency dependent delay is not overly important to the overall sound, but it requires considerable care to ensure that audible artifacts are not created as a result of the delay.

The conventional crossover of old was the Butterworth. Maximally flat frequency response, a Q of 0.707 (damping factor of 1.414), and 3 dB down at the crossover frequency. It has been shown by many workers in acoustics that this is actually wrong, as a 3 dB peak is experienced at the crossover frequency. It should be noted that this only occurs with 2nd order (12 dB/ octave) filters and higher - a 1st order filter does not have that problem.

The response of this filter is shown in Figure 9, and the peak at the crossover frequency is clearly visible. Figure 10 shows the phase response at one octave below crossover frequency - the signals are perfectly in phase (after inversion of one signal - the 12 dB crossover *always*inverts one signal with respect to the other.



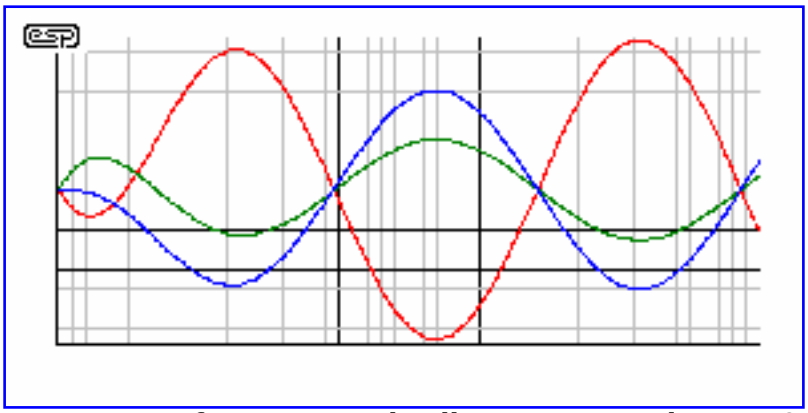


Figure 9 - Amplitude Response of Butterworth Filter ... **Figure 10 - Phase Response, 1 Octave Below Xover**

What about a square wave? This is supposed to be the most telling aspect of a design, which is interesting in itself since a synthesiser is the only instrument that is capable of producing a square wave, and no-one ever uses an unfiltered square wave anyway. Well, the result is shown in Figure 11, and the combined signal looks nothing like a square wave. The fact of the matter is that all frequencies that make the square wave are still present in their exact amplitude relationships, but they are shifted in phase. This is completely inaudible, and that has been proven many, many times. Human hearing is not sensitive to absolute phase, and responds to relative phase only if it causes a peak or dip in the frequency response. I suggest that you treat any claim to the contrary with the utmost suspicion, as the writer has a hidden agenda (to sell you his product being the most common).

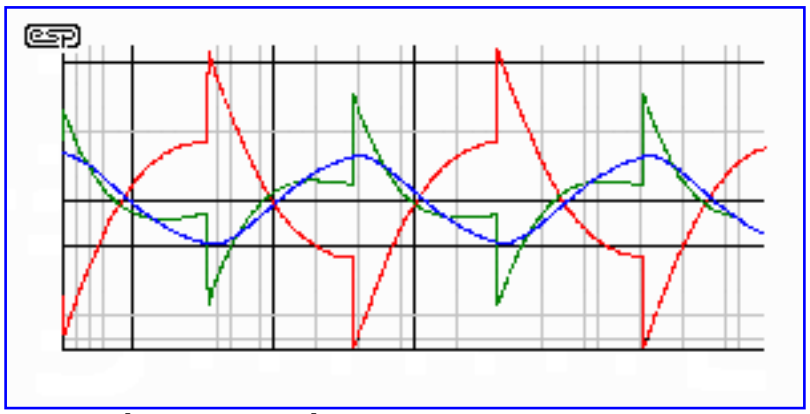


Figure 11 - Phase Response at Xover

Now, for reasons that are unclear (to me anyway), to obtain a license to use the term "Time Aligned", the speaker must be demonstrably capable of reproducing a square wave. Que? License?? Oh yes - the term is trademarked, and one may not advertise speakers as "Time Aligned" unless the appropriate fee is paid (presumably - I have no idea how much this costs), and the requirements are met. The biggest problem faced with getting any crossover to pass a square wave is simply phase shift. 1st order filters do it, but few drivers can cope with the low rolloff. An interesting tradeoff is the so-called "subtractive" crossover. This uses a single filter (of any slope), and subtracts the output of that from the input signal. The result is perfect square wave response, and a flat summed response.

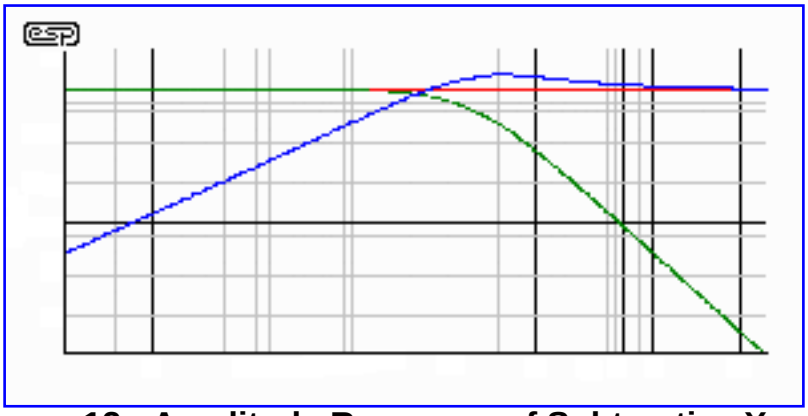


Figure 12 - Amplitude Response of Subtractive Xover

Do you see the anomalies? There is a bump in the high pass response, and although the low pass is 12 dB/ octave, the high pass is only 6 dB/ octave. Even if the "real" filter is 24 dB/ octave, the subtracted one is still 6 dB/ octave. Figure 13 shows the combined waveform and the high and low pass waveforms (input is a square wave).

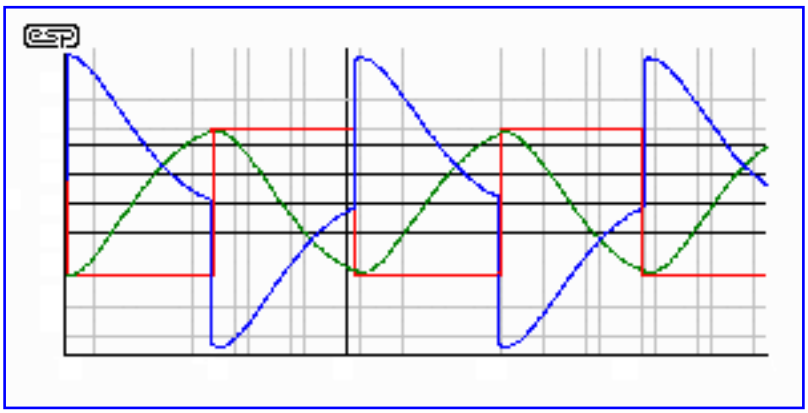


Figure 13 - Square Wave Response of Subtractive Filter

One driver will have an easy enough time, as it will be prevented from entering into the region where it becomes "hostile", with unpleasant lobing effects and possible cone breakup. The tweeter has no such luck! The design frequency is not as expected either (the diagrams shown used the same filter that gave a crossover frequency of 2.8 kHz in Figure 9). When we obtain a crossover frequency of 3450 Hz, the signal to the tweeter is down by only 11 dB at 1 kHz (typical of the resonance frequency of many high end tweeters).

The primary issues that confront the crossover designer are the constraints of the drivers themselves. As soon as the diameter of the radiating surface (the cone) of a driver becomes "significant" with respect to wavelength, you will have problems with lobing. This causes poor off-axis response, and makes the overall sound power output something of a gamble. A safe enough rule of thumb is that no speaker should be asked to reproduce any frequency where the cone diameter is greater than one wavelength. A typical 150 mm (6") mid-bass driver should not be operated above about 2300 Hz, and a

100 mm (4") driver is limited to around 3450 Hz. In addition, all loudspeakers will have cone breakup at some frequency - this can be "soft", causing no gross unpleasant sounds, or "hard", where the sound is quite objectionable. Generally, the more rigid the cone material, the worse it will be when it is finally incapable of true pistonic movement. This is one of the reasons that paper cones are so popular. I do not propose to cover this particular area in detail - further information is available on the Web (right or wrong, subjective or measured - this is up to you to determine).

It is very important that no appreciable power is supplied to a driver at or above the frequency where the cone breaks up or where the cone diameter exceeds one wavelength. The result is almost always a sonic disaster at the high frequency end. A relatively steep crossover is the only way to ensure that this colouration is kept below audibility.

Likewise, no speaker should be operated through its resonant frequency (pity about the bass driver!). For typical tweeters, this is between about 900 to 1500 Hz, and it is imperative that no appreciable power is allowed to get to the tweeter at its resonant frequency - the result is audible, not always insufferably unpleasant, but usually fatiguing and the sound is definitely coloured. With passive crossovers, the resonant peak also changes the characteristics of the crossover network (see High Quality Passive Crossover Design for more details).

This is surely one of the major quandaries facing any loudspeaker designer. To use a steep rolloff crossover, with its attendant transient response problems (and yes, these are real), or a simple 1st order design, that will allow the signal through that will excite the speaker at frequencies it will handle poorly. Despite some of the claims that you may see, there is no evidence that anyone has actually made a speaker that can handle more than about 6 octaves, and most will not come close to managing that. I would normally expect that a driver (other than most tweeters) will handle about 4 octaves reasonably well. The table below shows a few possibilities. A four way system is required to make it across the full audio band if you limit the drivers to 3 octaves, and with 4 octaves per driver, a 3-way system can exceed requirements in theory - the crossover frequencies may not be suitable for a great many drivers. Two way systems will almost invariably miss out on the lowest octave or three.

Oct/Driver	Low	Mid	High
3	39 - 312	312 - 2,500	2,500 - 20,000
4	< 10 - 78	78 - 1,250	1,250 - 20,000

Table 4 - Driver Frequency Ranges

Distortions

Naturally, if the number of drivers is reduced, the bandwidth they must cover is much greater - ever wondered why some (many?) large 2 way systems just don't seem to cut it? One of the biggest problems (and rarely spoken of) is intermodulation. If a cone is moving back and forth reproducing a low frequency, as well as 'jiggling' back and forth simultaneously reproducing a higher frequency, what will happen to the high frequency?

This is not an electrical system, this is pure mechanics and high school physics. Remember the Doppler effect? As a car (for example) comes towards you, the sound is higher in pitch as the sound waves are 'squashed' together by the forward motion of the vehicle. As it passes directly past you, the pitch falls to normal, and becomes lower as the car retreats from your observation point. Everyone has heard this effect, and many people have equated it with loudspeakers. This is actually not quite correct (IMHO), for reasons that are fully examined (and explained) in an article, but for now, suffice to say that the effect that has been claimed as 'Doppler' distortion is a combination of phase distortion and intermodulation distortion.

The Doppler effect is caused by compression or rarefaction of the wavefront, depending upon whether the object is approaching or retreating from your position. A loudspeaker cone does exactly the same thing! The high frequency tones are phase modulated by the cone movement caused by the low frequency tones.

The real problem is intermodulation, and this is one of the major arguments for using ported enclosures, since it reduces cone excursions at the lowest frequencies, and therefore reduces the tendency of the voice coil to partially leave the magnetic field, and introduce amplitude modulation distortion of the higher frequencies (i.e. intermodulation). The difficult load this presents to the power amp, and the phase irregularities of ported enclosures are well known, and I will not dwell on them here. Other alternatives exist ...

- transmission lines - usually very good performance, but rather bulky and hard to build
- horn loaded enclosures - not to everyone's taste, in size or sound
- multiple speakers to share the load - Ok at low frequencies, but causes problems when the distance between drivers exceeds 1 wavelength
- remove all the low bass from the main speakers and use a subwoofer - often an excellent choice, but not to everyone's taste (and subwoofer positioning can be

almost impossible in some listening spaces)

Other distortion generators have been discussed - cone breakup, tweeters receiving significant power at their resonant frequency, and drivers expected to extend their response way past the point where they become highly directional.

The major effect we hear is simple loudspeaker intermodulation distortion. A loudspeaker driver uses a motor, consisting of a voice coil, which is immersed in an intense magnetic field. The radiating element (usually a cone or dome) is coupled to the motor, and supported by a surround of corrugated material, rubber (usually synthetic) or foam. Additional support is provided by the spider, which is attached to (or near) the voice coil former - this is essential to prevent the cone from shifting, and causing the voice coil to rub on the magnetic pole pieces (poling).

The surround, spider and the motor itself are linear over a limited range. The maximum excursion of a driver (Xmax) describes the maximum physical movement allowed, but usually does not guarantee that this full range of movement will be linear. If it is not linear, the speaker will distort - subtle with some, gross with others. How do you know what a driver will be like at its limits? You can ...

- believe the manufacturer - not always a good idea
- test it yourself - be prepared for the odd damaged driver if you go too far
- ask others for their experiences - expect many conflicting responses
- make sure that your design will stay well within limits, perhaps 25% of the maximum - pay more for the driver
- use a servo system - feedback from the cone to the amp will linearise the movement, but it only works at low frequencies

Now we know that there will be intermodulation products generated when the speaker driver is outside its (often limited) linear range, causing the higher frequencies to be distorted as the bass forces the cone towards its limits. This is similar to amplifier clipping, except that it is progressive, and much more subtle - and therefore more insidious, because it is so difficult to detect reliably. Some musical passages will just not sound right at high volumes, but are fine at lower (often unrealistically low) levels.

The ideal is naturally to limit the excursion to the absolute minimum, but this is not always possible, especially with bass drivers. In this case, it is far better to relegate the bass to its own speaker altogether - a subwoofer is not just for home theatre - it can work absolute magic on normal musical programme material as well, including music that does not appear to have a great deal of low bass.

Phase Audibility

The audibility of absolute phase is nil.

I must explain this further, as this is a somewhat contentious issue. It can be proven in ABX tests that there are some signals where the difference between a non-inverted and inverted signal is audible. Certain waveforms and instruments are highly asymmetrical, and if listened to in isolation *will* sound different if the phase is reversed. The difference is not subtle, either - it can be very pronounced. This is much more likely to be a result of loudspeaker driver behaviour than anything else, and the "correct" phase is anyone's guess - should it be inverted or not? We don't know the answer, since we will be unsure of what the instrument sounded like "live" - it is possible that neither the inverted or non-inverted recorded signal will sound like the original, so the point is moot.

The key issue here is that if we listen to a saxophone (a good example of an asymmetrical waveform) with the phase normal then reversed, all we hear is a difference - there is not necessarily a "right" or "wrong" phase, since it depends on the way the instrument was miked in the first place. If the period between listenings is extended to a few minutes, the chance of us hearing the difference will be minimal, and we still won't know which is "right" and which is "wrong" - all that this proves is that there is a difference, and it only becomes audible with some instruments.

This is probably the only case where an ABX test proves something that is not relevant in the general sense - so yes, absolute phase can be audible, but it is (generally) irrelevant. [Update added 30 Jul 2002]

The net result is that our ears do not care if there is a slight misalignment between the fundamental and harmonics of any instrument known. This is likely to cause howls of protest from people who won't actually bother to read this article in its entirety (if at all), but it has been demonstrated a great many times, and by various techniques.

> **Example:** Let's examine this from another perspective, using live sound as the source (for example, a string quartet). Regardless where you stand or sit, you will most likely be a different distance from each instrument. One could insinuate oneself into the very centre of the performers' space, but this is more likely to lead to your eviction from the venue than to improve your listening enjoyment. If one is a different distance from different sound sources, then the absolute phase of those sources will all be different too. Not only that, but the phase variation changes with frequency.
>
> At 40Hz, a two metre difference in the relative path lengths amounts to only about 42° phase

shift, but the second harmonic is shifted by almost 84° while the 5th harmonic (200Hz) suffers 208° phase shift. Move a little, and it all changes. Do we hear a huge variation in the sound (assuming a reasonable listening environment)? No, of course we don't. No-one will ever convince anyone who has been to a live performance that there is one (and only one) specific location from the musicians where the sound is somehow "right" (ignoring major auditorium problems, of course). The fact is that phase varies with distance and frequency, and this will not change. The complex nature of music using real (as opposed to synthesised) instruments guarantees that things will drift in and out of phase as a matter of course. Yes, the loudspeaker should be able to reproduce this as accurately as possible, but there is only science in the design of a loudspeaker system - no magic.

A simple all pass filter will shift the phase of an audio signal by 180° over a frequency range determined by the component selection, and it is completely inaudible - provided the source is music, and provided the phase sweep is performed slowly enough for our ears and brain to make the necessary adjustments. In fact, I have demonstrated this as the filter is adjusted (very slowly), and the sound quality remains the same. Nearly every (Ok, not nearly - every) recording ever made has been recorded using a microphone, had some equalisation applied, and/ or has had some additional treatment in the recording process. All of these introduce some degree of phase shift, but does it ruin a good recording? No. As the signal emerges from the vast majority of crossover networks, there are huge shifts of phase, as has been described above. A square wave subjected to phase shift still has all of its harmonics present, they are just slightly misplaced in time.

The sort of delay we will experience is dependent on the frequency, but it doesn't matter. Vented speaker boxes do 'awful' things to phase, as do many highly regarded 'feedback free' single ended triode (SET) amps. Any equaliser, be it a constant Q graphic, parametric, or just a simple tone control, will introduce phase shift as well as equalisation. The phase of a waveform changes as you move about - but your best friend sounds like your best friend regardless of your relative positions in a room, even though there are massive changes in phase as we walk around.

If we believe the 'absolute phase' lunatics, this would not be the case, so your wife may sound like your wife in one part of the room, but sound like the milkman in another. We all know that this doesn't happen - the tonal structure of a sound does not rely on the phase integrity of the received sound, only the relative amplitudes of the fundamental and harmonics. So a speaker that has perfectly flat frequency response but is not 100% phase coherent will sound the same as one that is also flat, but totally phase coherent. This does not include colouration caused by the cabinet or drivers - of course these are important. Assume the same enclosure, same drivers, but a phase shift applied to one, and not the other.

In isolation, they will sound the same. Put them together, and you will hear strange

reinforcements and cancellations as you move about. This is *relative* phase, and is very audible indeed. What we need to concern ourselves with is relative phase only. Two amplifiers with different phase responses used as a stereo pair will sound terrible if the shift is sufficient. Use two of the same amplifier, and there is no problem.

Absolute phase is inaudible within reason - a 3,600° phase shift represents a time delay that is significant, but a 360° phase shift will not be heard. Inverting a signal (e.g. reversing the connections to a loudspeaker driver) creates a 180° phase inversion, but this is not the same thing as a 180° phase shift! This is a point missed by many.

Relative phase is audible, depending on the amount, the frequency and the context. Two speakers side by side with 90° phase shift between them will sound dreadful - and the sound will change as you move about. The relative phase of two musical instruments playing in harmony *makes* the sound you hear - take away the phase shifts, and it will sound flat and lifeless.

There have been many tests and experiments to look at phase shifts within the audio band, and whether they are audible. Under controlled laboratory conditions (or using headphones), there is strong evidence that with single (complex) tones, there is an audible change. However, in a listening room with speakers reproducing music, there is little evidence that phase shifts are audible with the vast majority of recorded material.

An example of a pair of very typical all pass filters is shown in Figure 14. These are connected differently so I could show the different behaviour (not actually different, the phase of one is simply reversed from the other).

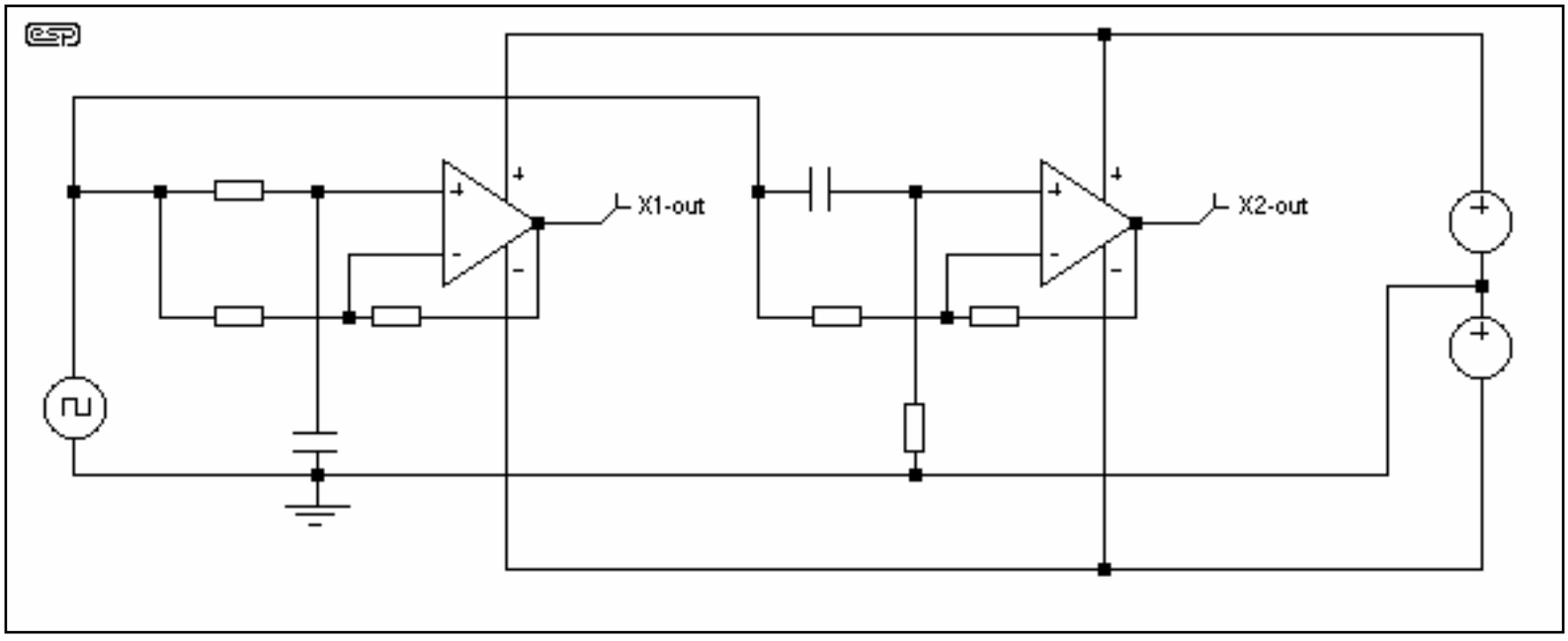
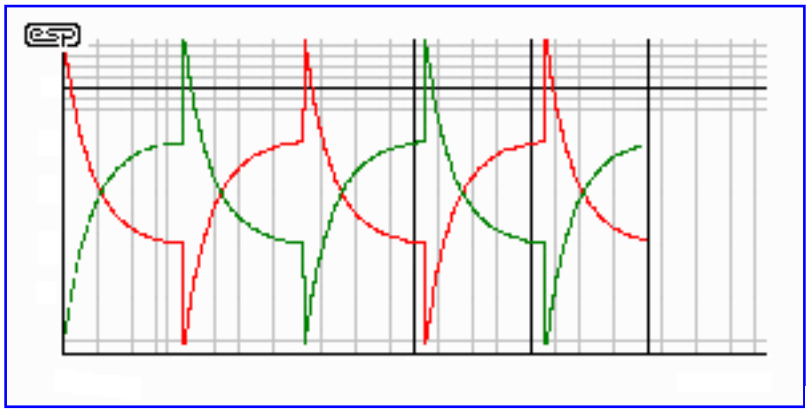


Figure 14 - All Pass Filter Networks

The resulting output and phase response of the filters is shown in Figures 15 and 16 respectively. I only included the phase response graph for one version - the other is simply the reverse of that shown.



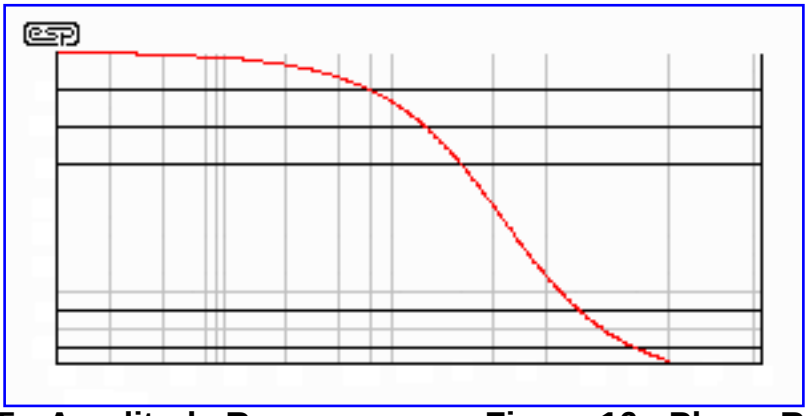


Figure 15 - Amplitude Response **Figure 16 - Phase Response**

Note that this particular filter is called 'all pass' - it passes all frequencies equally (i.e. the magnitude is unaffected). Not much of a filter by normal standards, but a useful tool nonetheless. Interestingly, if the input and output of an all pass filter are summed, the result is an ordinary filter. High and low pass responses are available. Not that there is a great deal of point, since this is vastly more complex than a 6 dB/ octave filter built conventionally. I just thought I'd mention it - someone might be interested :-)

Conclusion

For what it's worth, I originally started this article not to praise, but to debunk the theory that time alignment is the only way a speaker should ever be designed. Having done the research, run tests, and written the article, I confess that I must agree with many (perhaps even most) of the points made by the time alignment proponents. Mind you, there is still a lot that you will hear and read that is either gross exaggeration or a downright lie, and it can be very difficult to tell the difference unless you know exactly

what the real story is.

My overall opinion, based on the research for this article (primarily tests and simulations), is that time alignment is a very good thing, and perhaps all speakers *should* be designed this way. On the negative side, the offset required to achieve time alignment can lead to diffraction effects that may damage the sound quality far more than the misalignment. A sloped baffle means that you are *always* listening off axis from the drivers - not by a great deal perhaps, but off axis nonetheless. This conundrum can be resolved, and it has been by several manufacturers, each in their own way.

Use of 1st order crossovers means that the vertical axis of the speaker is very narrow - the speaker will sound entirely different when you stand up! This means that the signal propagated into the room is uneven, so the natural reverberation of the listening area is not excited evenly at all frequencies. Higher order crossovers are better in this respect, but cause their own problems. Relatively poor transient response is always claimed, but in reality, a great many high end manufacturers are using 24 dB /octave filters, especially with electronic crossovers, and achieve extraordinary results. My own system loudspeakers are triamped using my version of the Linkwitz-Riley 24 dB crossover, and they sound very good indeed. They are not time aligned, but based on the results of my work on this article, I would expect that when (not if) I rebuild the boxes (or just make a new system altogether) they can sound even better.

Reproduction of a square wave is something of a myth. I have received a very passable square wave response from a pair of small hi-fi boxes I use in my workshop. All I have to do is select a good position for the measuring microphone. How many sites have you visited in your quest for "the ultimate loudspeaker", where they claim (or show) the square wave response? How many admitted that the positioning of the measurement mic has a very great bearing on whether a square wave is reproduced or not? From what I have seen, no-one has ever claimed that a square wave is received perfectly, regardless of mic position, nor have they disclosed the actual measurement setup that was used - is this at the listening position in a "typical" room, or 300 mm in front of the speaker in an anechoic chamber? We shall never know.

Indeed, the room itself is still the greatest offender - even a coffee table that is in the acoustic path of the loudspeaker will have a profound effect on the overall response. Very few rooms are acoustically dead enough (IMO), and I have seen a great many photos of people's systems set up on polished marble (or whatever) floors in relatively bare rooms, with almost no acoustic deadening materials to be seen.

Human hearing is very adept at picking the original sound from the reverberant field,

provided the early reflections are not so early that they influence the direct sound. Given the highly reverberant listening rooms of some people, I have difficulty understanding how they can even tell what anything really sounds like - yet they will happily espouse their theories on what makes the sound better, ignoring the fact that their room will destroy the sound of any loudspeaker.

Finally, the quality of much of the recorded material available is absolutely woeful. Equalised to within an inch of its life (so it will sound "good" on crappy systems), compressed, "aurally excited" (ptooey!), and generally mangled beyond all recognition. To be sure, quality recordings are available, but are they available from your favourite artist(s)? Usually not, so you either have to change your musical tastes to experience a decent recording, or put up with the rubbish that is often the only version of the artist/ song available. I have so many CDs and vinyl recordings that I find unlistenable on a decent system that it's not funny - for one CD, I have to switch off my subwoofer or all my windows will fall out!

This article started as a short explanation, intended to dispel some more snake oil, and has become the missive you see due to the vast amount of information I collected as I ran the tests and simulations. I Hope that it has been of value to you - having read this far, I suppose it must have been. Expect an update shortly, after I have had a chance to figure out a way to determine the acoustic centre of typical drivers - perhaps manufacturers could supply this information as a part of their specifications (hint, hint).

Update

1. It was pointed out to me by Geoff Moss (who is still checking my pages regularly - thanks Geoff!) that the all pass filter has been used as a time delay, and this usage is described by Siefried Linkwitz in one of his articles published many years ago in Wireless World (now Electronics World). I have run some simulations of a 4 stage all pass, and it is indeed possible to get a time delay that is reasonably constant for at least a few octaves. Now, while there is no doubt that the principle works, there are not too many people who would actually want to have anything from 4 to 8 stages (all based on opamps) as a series string in line with the tweeter signal. The high frequencies are the most easily damaged, and such an arrangement would generally be considered unacceptable.

Nevertheless, it is a valid usage of the principle, and shows that this is not a new topic - indeed far from it. The original was published in 1978, and was republished in 1980 in Speaker Builder magazine - a copy of the article as published in Speaker Builder can be obtained from ...

www.linkwitzlab.com/sb80-3wy.zip

I do not propose to go into great detail on this topic, but since it has been done (and is described in excellent detail in the article), this will provide you with more information on the topic.

2. I recently saw a posting on a newsgroup referring to a (single driver) speaker reviewer's claims that a single driver is the only way to undo Doppler distortion introduced during the recording process. For various reasons, this is complete rubbish, but mainly because single driver speaker systems introduce vast amounts of intermodulation type distortion, especially if they are expected to cover anything more than a moderate range and at relatively low levels. Even horn loading does not reduce cone movement sufficiently to prevent distortion, but it may reduce it to within acceptable levels (depending on the design of the horn and driver).

By comparison, a microphone diaphragm may move a very small fraction of a millimetre at most, and the distortion introduced is minimal - indeed, with capacitor (aka 'condensor' or 'condenser') mics, the movement is infinitesimal, and distortion can be all but ruled out. In the case of 'heavy' musical styles, there will be separate mics for each instrument, so the most troublesome signals are removed from the equation.

In any case, the claim is fallacious, and highly misleading for anyone without the knowledge to be able to examine the facts properly. This is classic "snake oil" marketing at its very best. I wonder how such a speaker handles electronic music, that has never even "seen" a microphone during the entire recording process? Maybe it is clever enough to know the difference ... no, I didn't think so either :-)

For a speaker reviewer to be spouting this sort of garbage gives some idea as to their overall credibility - would you take any notice of someone who made such absurd claims in any field other than audio? Would you believe it with audio? I certainly don't.

References

Although the majority of this work is the result of tests and simulations I have carried out, there are a few other sources as well. Many are part of the ESP site, and I shall not bother referencing my own work. The only other real references used are shown below.

ASA 130th Meeting - St. Louis, MO - 1995 Nov 27 .. Dec 01 1pEA4. Time Align (registered) loudspeaker crossovers.
Edward M. Long, E. M. Long Assoc., 4107 Oakmore Rd., Oakland, CA 94602

http://www.auditory.org/asamtgs/asa95stl/1pEA/1pEA4.html

For an excellent expose of the Doppler "distortion" myth (be warned, it is highly maths intensive, but very well done), have a look at ...

http://www.geocities.com/kreskovs/Doppler1.html

You may also consider my explanation worth reading - this is what I have to say on the topic ...

Doppler Distortion - Real or Imaginary?

An interesting article covering amplitude audibility, phase (including absolute phase) - well worth reading (make sure you read part 1 as well)

Audioholics - Human Hearing - Phase Distortion Audibility

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 30 May 2002./ Updated 09 Apr 2005 - added javascript to show charts, added Audioholics reference.

The Audio Pages

Elliott Sound Products **Series vs. Parallel Crossover Networks**

Copyright © 2003 - Rod Elliott (ESP)
(With additional material by Gene DellaSala - Audioholics)
Page Published 14 Aug 2003 (Updated 24 Aug 2003)

Articles Index

Main Index

Contents

- Introduction
- 1.0 - First Order Comparison
  - 1.1 - Response
  - 1.2 - Impedance Variations
  - 1.3 - Summary
- 2.0 - Second Order Comparison
  - 2.1 - Response
  - 2.2 - Impedance Variations
  - 2.3 - Summary
- 3.0 - Conclusion
- 3.1 - Series & Parallel Networks are (Virtually) Identical

Introduction

Despite many of the myths that surround series networks and their acclaimed superiority over conventional parallel networks for loudspeaker design, both networks can be designed with identical transfer functions if the load impedance remains constant. Most of the claims regarding series networks are either grossly overstated or blatantly wrong and may cause deleterious effects on system performance. As with all aspects of design, there are compromises that must be made, and it is impossible to make an informed decision if you are unaware of the facts.

This article is intended to show that there are no greatly enhanced features in a series or parallel network - if properly designed their performance is essentially identical in terms of response, phase and (by extension) transient response. It is unwise to claim that one type of network is superior to

the other, when simple logic dictates that if amplitude and phase response are the same, then all of the filter's other characteristics are also the same.

There are other factors than just the response, and this is where the differences between the network topologies exist. Each has good and bad points that must be considered.

1.0 - First Order Comparison

First order (6dB/octave) networks have a strong following amongst many audiophiles, and indeed, they have a number of very desirable features. They have the best possible transient response, and are predictable and easy to design, but as with all things there is a down side. The demands on the drivers are extreme, with significant power delivered to the tweeter even at its resonant frequency, and the risk of cone breakup and off-axis lobing for the mid-woofer.

Nevertheless, at low power, intermodulation products can be kept within reasonable limits with careful driver selection, and they can sound very good indeed.

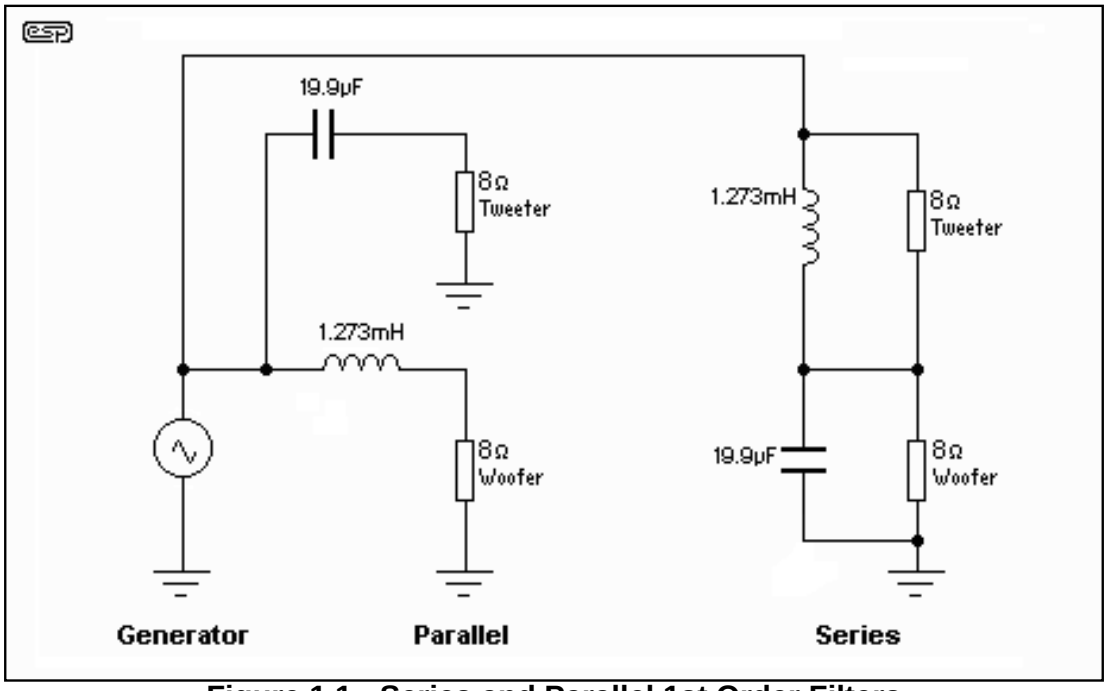


Figure 1.1 - Series and Parallel 1st Order Filters

Illustrated above are equivalent series and parallel first order crossovers with 1kHz crossover points for a fixed load. Note that resistive loads were used in order to minimise analysis variables. There is some material in the conclusion of this article describing further simulations and transient response that encompasses complex load impedances typical of a loudspeaker. 1kHz was chosen for one reason - the crossover frequency is nicely centred in the graphs for best display. The effects shown in this article are identical at any frequency.

Input impedance is exactly the same for each type, and is essentially perfectly flat, with both circuits dropping by 2 millohms at the crossover frequency. This is of no consequence, and may be ignored.

1.1 - Response

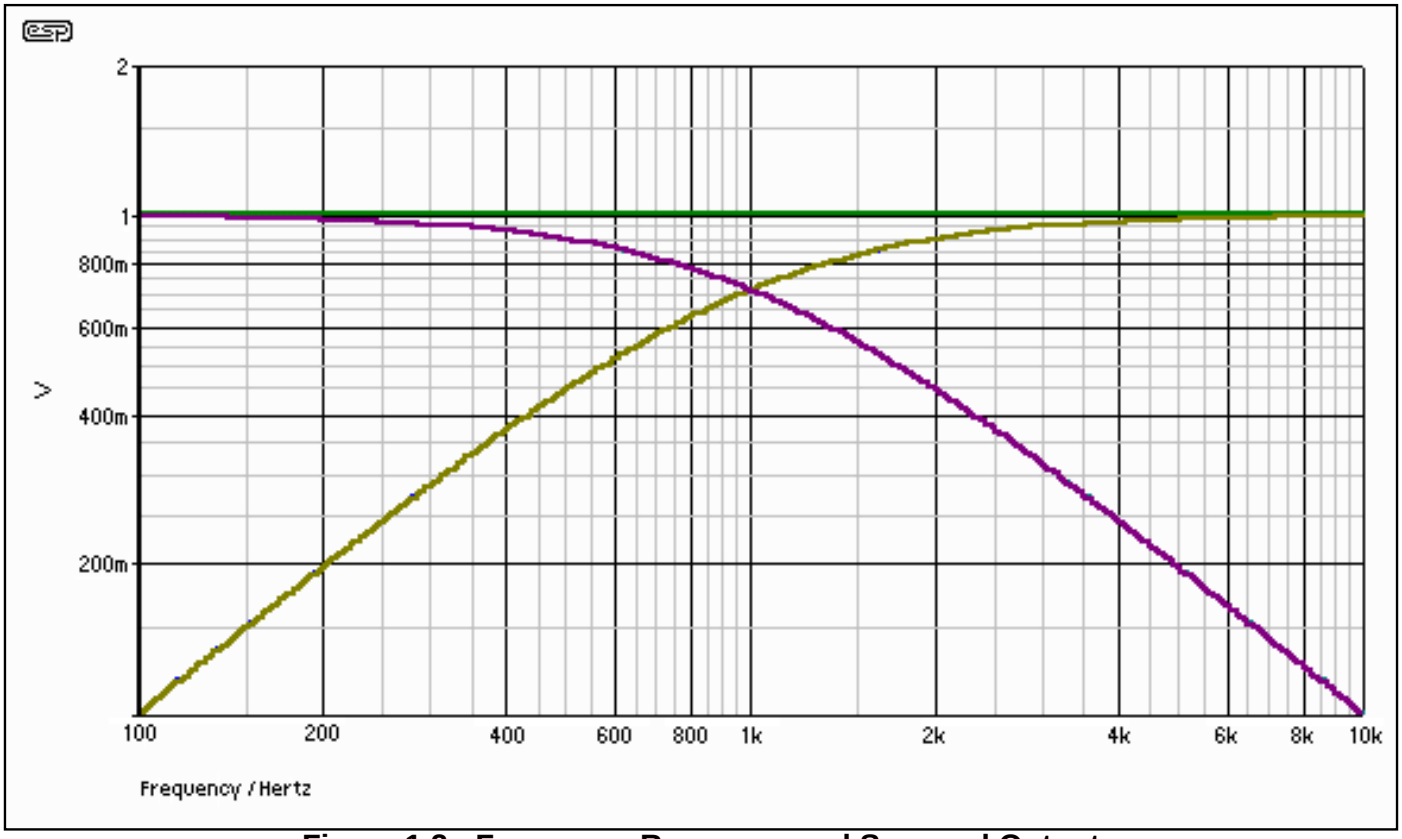


Figure 1.2 - Frequency Response and Summed Output

The frequency response and electrically summed outputs are shown in Figure 1.2 and it is quite obvious that they are identical, since the graphs are perfectly overlayed (there are 6 graphs on the chart, not 3).

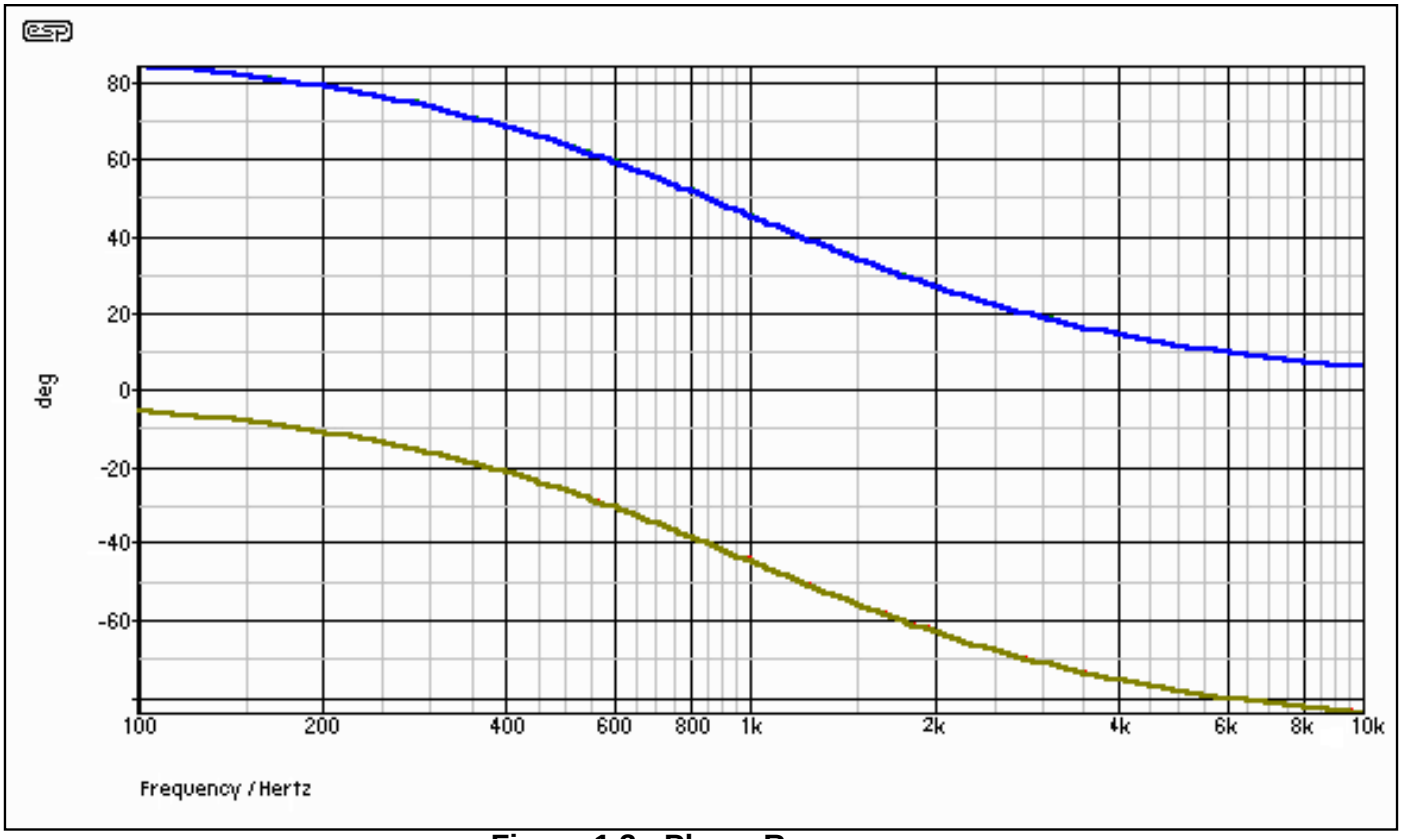


Figure 1.3 - Phase Response

The phase response of the series and parallel crossovers are also identical as can be seen above. There are 4 graphs (2 serial and 2 parallel) and again, they are perfectly aligned.

1.2 - Impedance Variations

Figure 1.4 shows the variation of high and low pass filters and summed response when the woofer impedance is varied by +/-2 ohms. Red shows the electrical sum of the variation with 6Ω impedance, and the green graph is for 10Ω. Note that only the low pass filter response is affected.

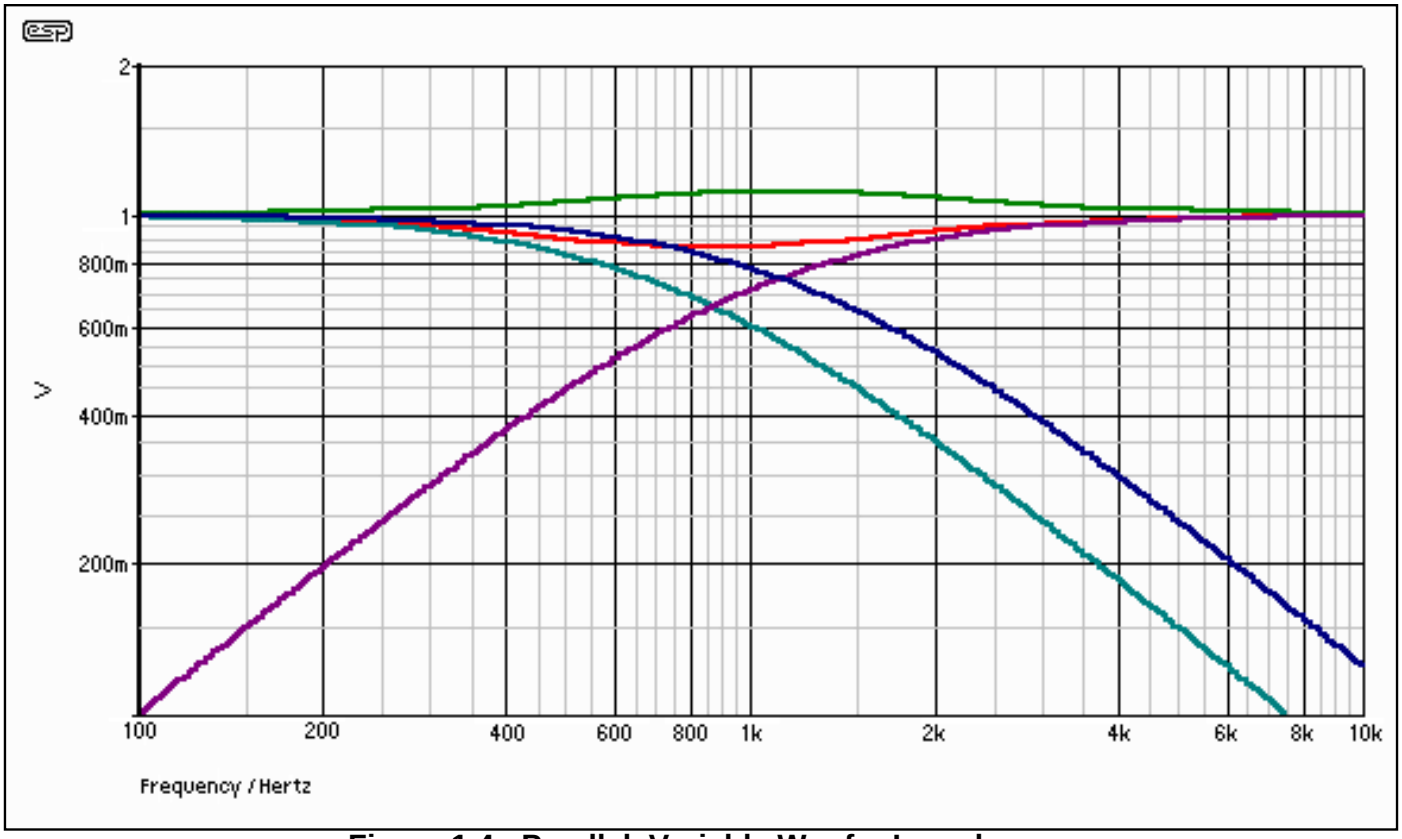


Figure 1.4 - Parallel, Variable Woofer Impedance

The results for tweeter impedance variations are similar (and affect only the tweeter section of the filter), but have not been shown, since the tweeter is far less likely to undergo any noticable change than the woofer.

The graphs below are *very* interesting. The woofer impedance was again changed from 6Ω to 10Ω as was done with the parallel network. Note that although the crossover frequency moves (it becomes higher at higher woofer impedances and vice versa), the summed response remains completely flat.

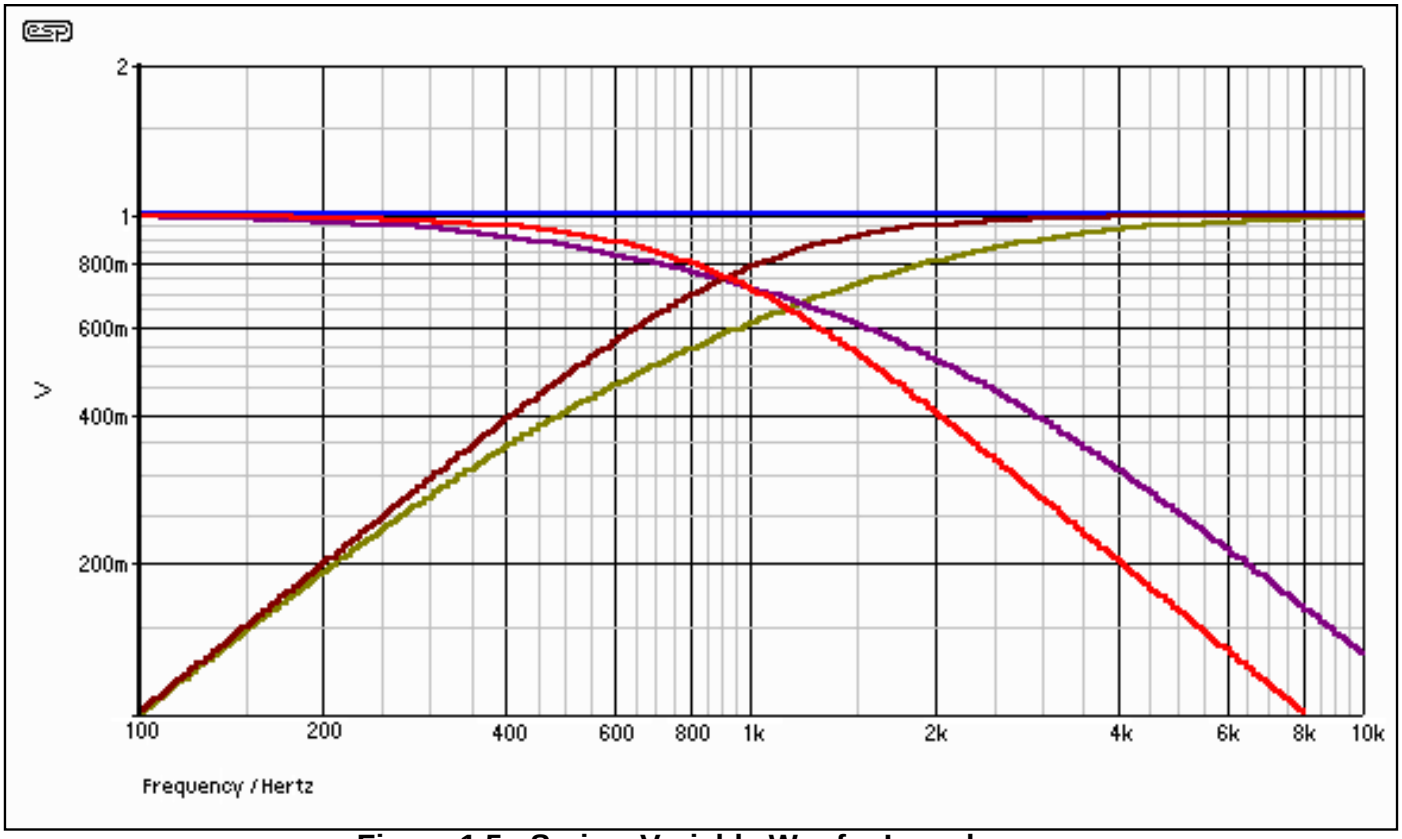


Figure 1.5 - Series, Variable Woofer Impedance

The two sections have a complementary shift - when woofer impedance changes, it effects both low *and* high pass sections, and changes the Q of the filter sections. The result is quite obvious - unlike a parallel crossover, the response remains flat regardless of a shift in the woofer (or tweeter) impedance. If both change in any direction, the same thing happens. In theory, this means that the series network is almost immune from impedance variations in the drivers.

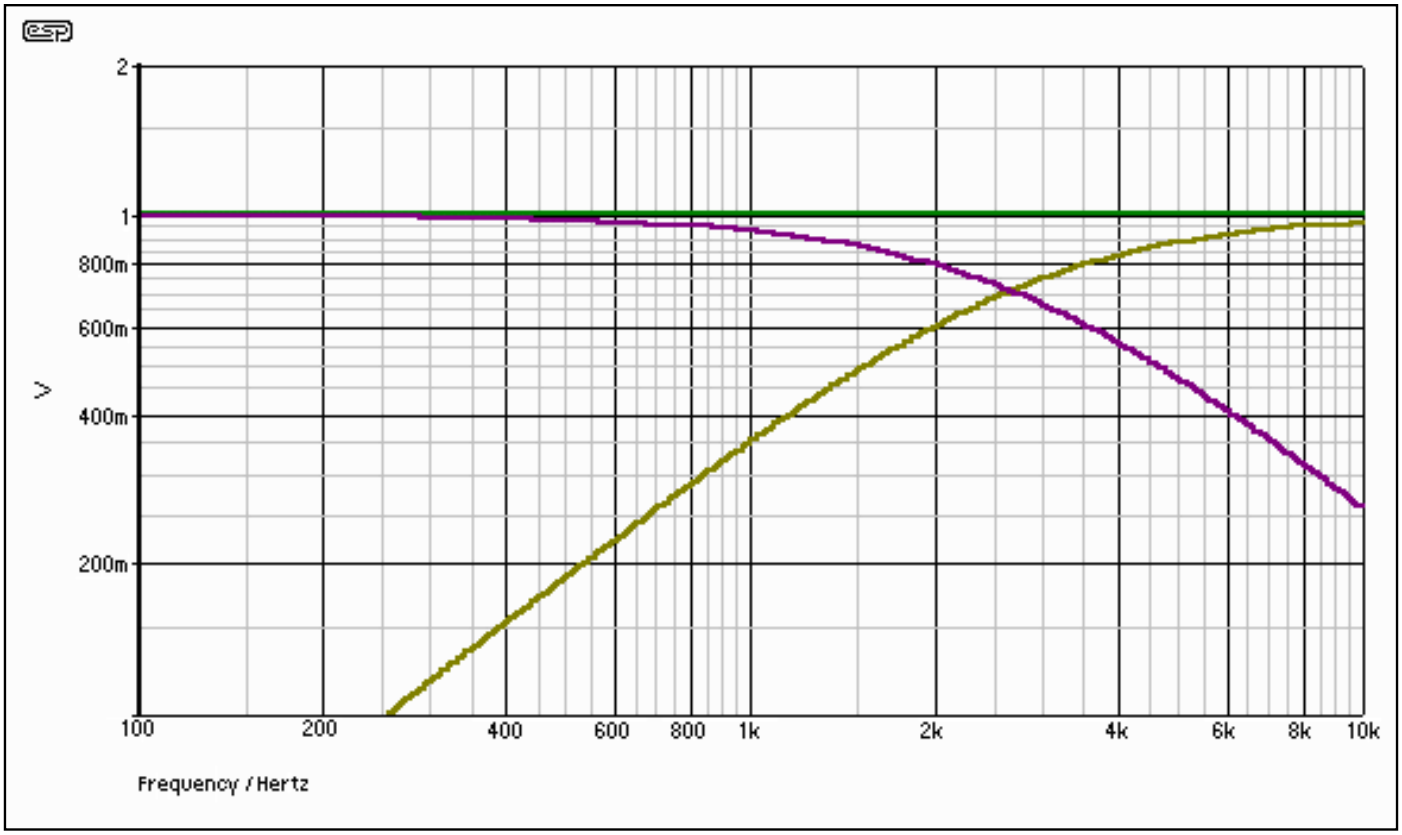


Figure 1.6 - Series, 20ΩWoofer, 3Ω Tweeter

By changing the driver impedances, two things happen. The filter Q changes, and the reflected change affects the behaviour of the other filter section. Although the individual response, Q and phase varies, the net result is that the effective crossover frequency is changed, but nothing more. This is a remarkable property, and the series first order is the only crossover filter circuit that has this ability.

Remarkable though it may be, it is still advisable to design the series network correctly, and maintain everything as close as possible to the design values. Should the woofer impedance increase (with voice coil temperature, for example), the crossover frequency will move upwards, thus providing a small measure of added protection for the tweeter at sustained high power levels.

However, all is not completely rosy. Everything in electronics is a compromise, and the selection of a crossover is no different. There is one final test that needs to be applied, and that is to examine the amount of woofer back EMF that reaches the tweeter. This is an area where the series network is inferior to the parallel.

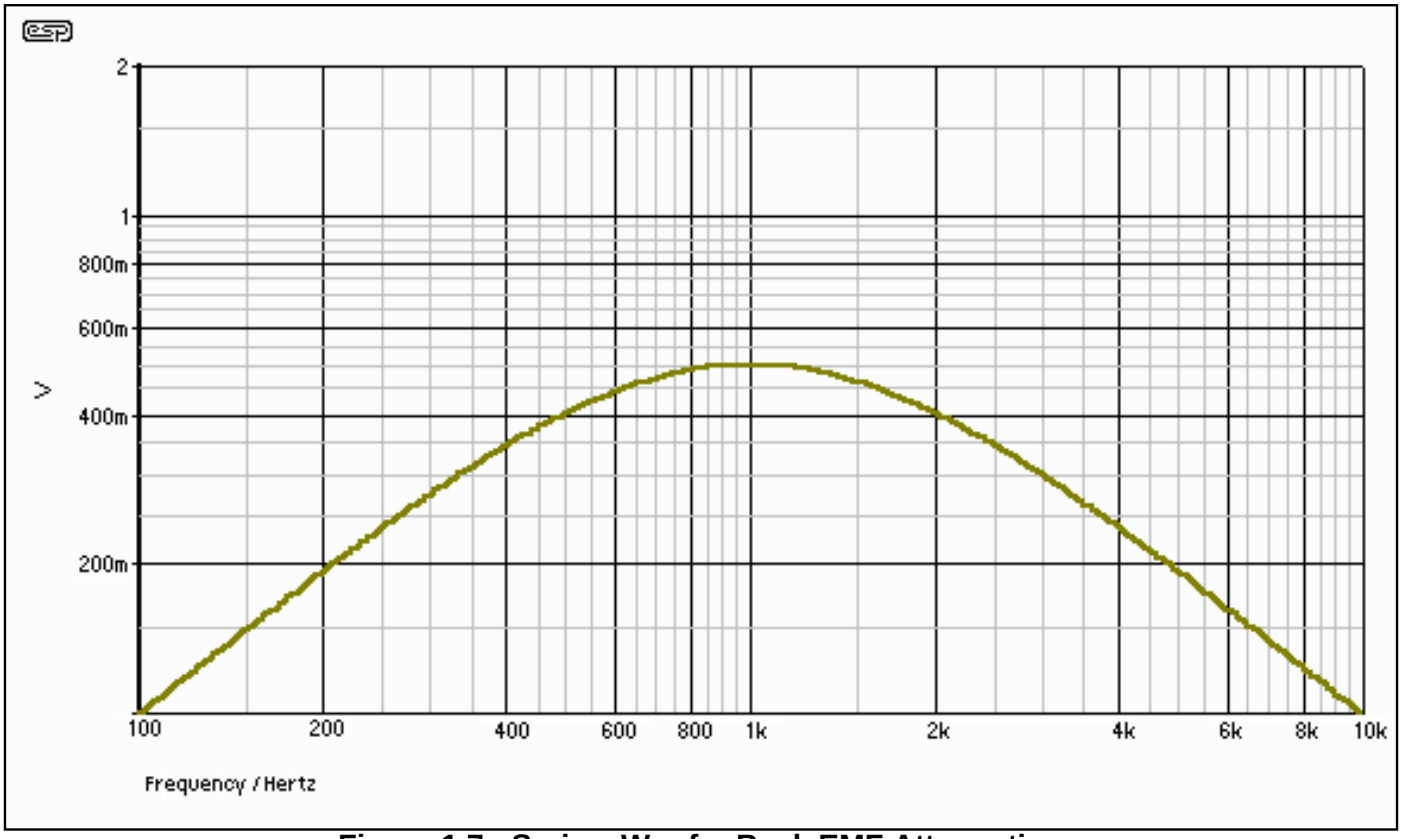


Figure 1.7 - Series, Woofer Back EMF Attenuation

With a parallel network, only the amplifier's output impedance plus the impedance of the cable allows any cross coupling between high and low pass sections. With a zero ohm source, attenuation is infinite, and is not shown above.

A series network relies solely on the isolation of the crossover filters, and as a result, the back EMF from the woofer is not attenuated as well. This may not be a major problem, since the attenuation of back EMF is the same as for amplifier power (actually, it is 3dB better), and the latter is at a far greater amplitude. It is a consideration nevertheless, so be aware that it may increase tweeter intermodulation.

1.2 - Summary

The series network is probably a better choice than parallel for a number of reasons. It retains a flat response even when the driver characteristics change, and is to an extent "self correcting". Implementation is no more difficult than for an equivalent parallel network, and the same component values are used.

On the negative side, woofer back EMF supression is significantly worse than with a parallel network - it is up to the designer to determine if this is likely to cause a problem.

Finally, it must be remembered that any first order network dictates that the drivers will have significant power applied at frequencies where their performance will be rapidly deteriorating, however for a system that will never be operated at high power, the performance can be very

satisfying.

2.0 - Second Order Comparison

The design process for a 12dB/octave filter is completely different for series and parallel implementations of the same design. For a parallel network (assuming a Butterworth alignment for the sake of simplicity), the capacitance and inductance are calculated by …

$$C = 1 / (2 * \pi * f * (Z * \sqrt{2}))$$
$$L = (Z * \sqrt{2}) / (2 * \pi * f)$$

(where Z is impedance, f is frequency, $\sqrt{2}$ is ≅1.414, and $\pi$is ≅3.14159)

A series crossover design is different in terms of the component values …

$$C = 1 / (2 * \pi * f * (Z / \sqrt{2}))$$
$$L = (Z / \sqrt{2}) / (2 * \pi * f)$$

For this exercise, the crossover frequency was arbitrarily selected to be 1kHz, and 8 ohm resistive loads were used. The series network has the advantage of using smaller inductance values, but capacitor values are higher. The difference is unimportant, but capacitors for crossovers are more expensive than inductors. This is a minor point if there is an improvement in performance.

The values used for the simulations were as follows …

Table 1 - Second Order Crossover Values

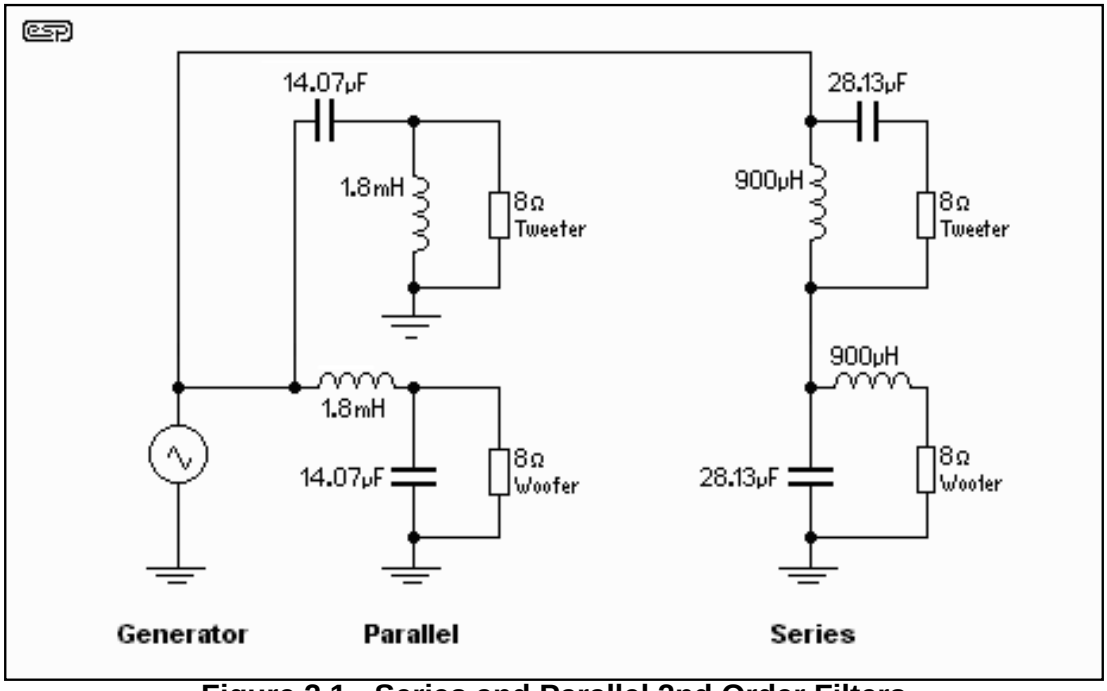


Figure 2.1 - Series and Parallel 2nd Order Filters

2.1 - Response

As with the previous example using a first order filter, when properly aligned, the response is identical. Because the plots look exactly the same as the previous example (other than the rolloff slope), there is little point displaying graphs that show two sets of curves that are perfectly matched.

It can be stated that if two filters, regardless of topology (series, parallel, active or passive) have an identical frequency response, then they must also have identical phase and impulse responses, since these cannot be separated.

Of course, this only holds true as long as the source and load impedances are also identical. Input impedance of both filters is essentially completely flat, having a variation of only 4.6 mdB (i.e. 0.0046 dB). Due to rounding errors in the component values, there is a tiny variance between the two filters, however it is completely insignificant (about 0.17 Hz difference).

One thing that should not be overlooked is the inductor's resistance. While this causes a small loss of level with a parallel crossover network **, it will cause the series network to "shelve" the tweeter rolloff. As a result, a DC resistance of (say) 800mΩ will cause the signal applied to an 8Ω tweeter to drop to a minimum of just over 20dB below the applied signal *regardless of frequency!* This includes DC under amplifier fault conditions.

There is virtually no difference between series and parallel at about 1 decade below crossover (i.e. 1/10th the frequency), but below that the difference becomes apparent. There may be as much as 20dB more level applied to the tweeter at 20Hz with a series crossover vs. an otherwise identical parallel version (with an inductor DCR of 0.8Ω)

** Although there is a small loss of level, the parallel crossover's theoretical response is greatly disturbed by even a 0.8 DCR in the inductor. This will cause a response anomoly of about 1dB, with the woofer output being 0.8dB down at one decade below crossover frequency. Naturally, higher resistance will create more deviation in response. The series network's overall response remains flat. Normally, the inductor's DCR must be factored into the design, regardless of crossover type.

2.2 - Impedance Variations

As was shown to be the case with the first order implementation, by its very nature, the two segments of a parallel crossover are separate, and share only the amplifier’s output impedance, plus the impedance (R, L and C) of the speaker lead. Speaker lead capacitance may safely be ignored as it is insignificant compared to the capacitances within the crossover network.

A series network on the other hand, relies on the integrity of the series elements - all of them. A change in woofer parameters (for example) therefore affects the tweeter, and vice versa. The tweeter is likely to have smaller and fewer changes than the woofer in a practical system.

It is interesting to see the behaviour of the two network types when the outputs are summed electrically. This is a severe test, and in 12dB types, neither crossover is significantly worse than the

other in this respect.

Any change in the parameters of the woofer (the most likely to change) causes a change in the tweeter parameters, and the summed electrical response varies with both types. Since it has been established that the two filter types are identical when all values are at their design figures, there is no point showing this. The following two charts show the extremes - with the woofer impedance at 4 ohms and 12 ohms (the latter value being much more likely).

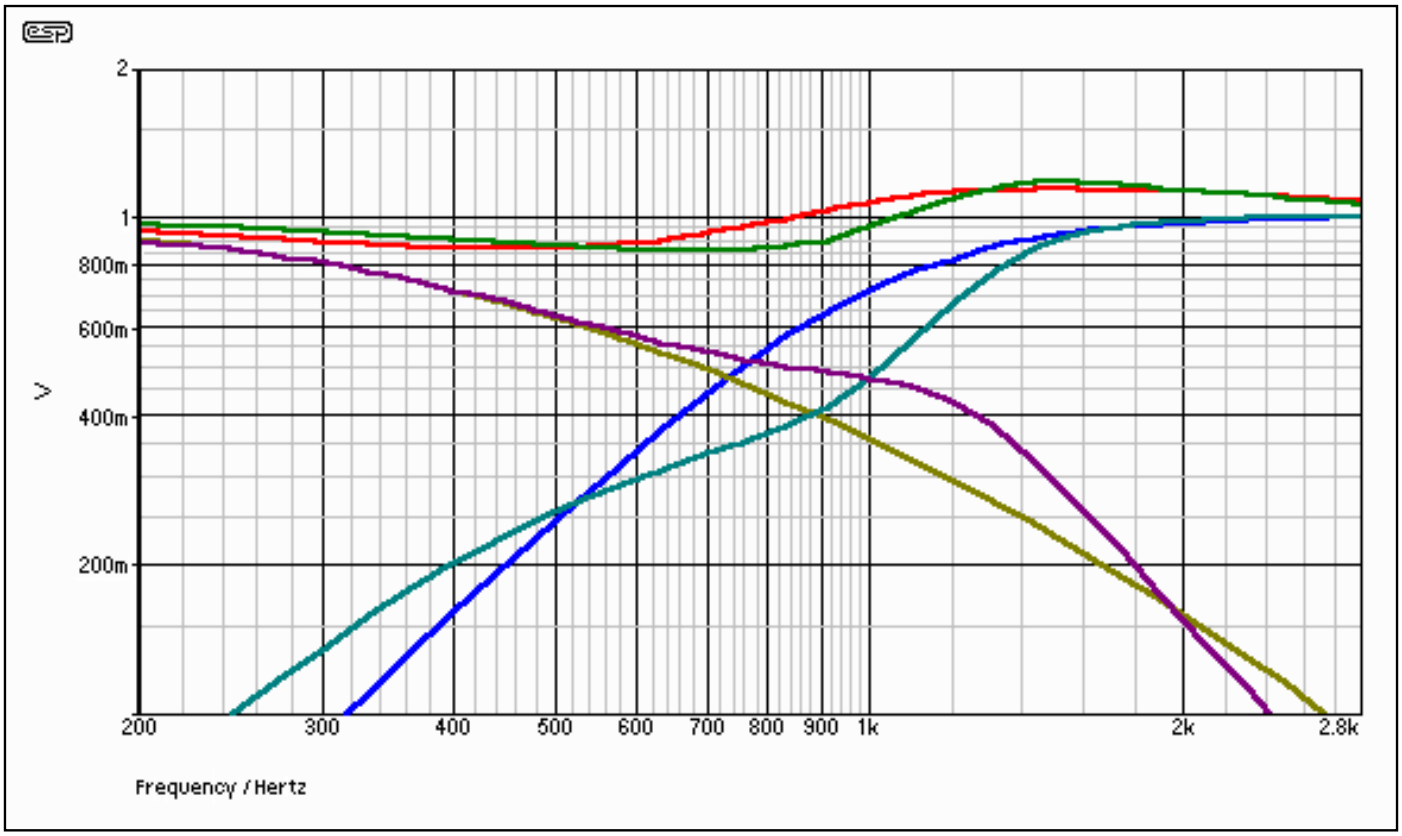


Figure 2.2 - Series and Parallel - Woofer at 4Ω

The red trace is the summed electrical response of the parallel network, and green for series. The dark green and violet traces (with the kinks and bends) are the individual responses for the series network.

Note that although both series and parallel networks have deviated from the ideal, the parallel network has a flatter and less rapid change. Overall, the difference is marginal.

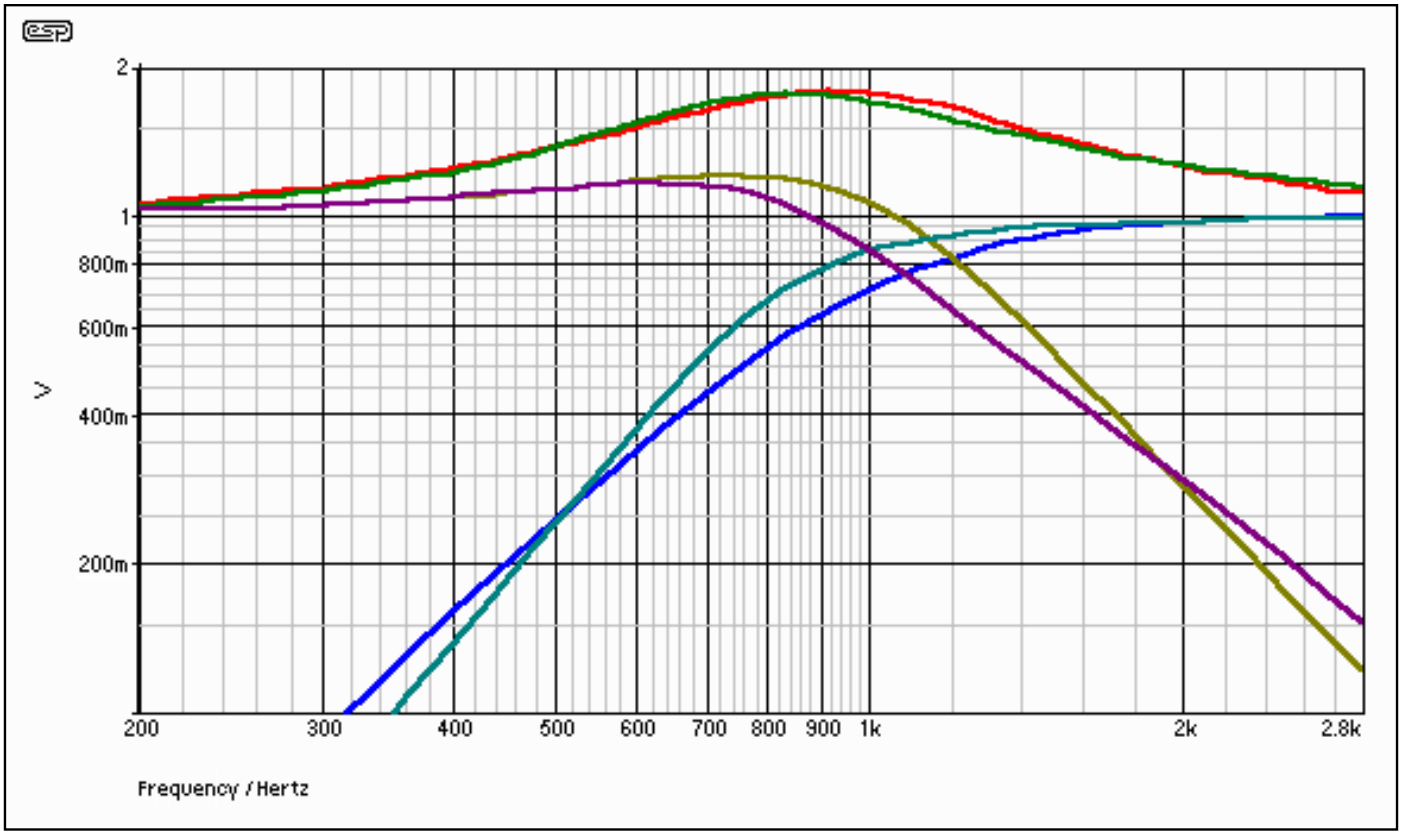


Figure 2.3 - Series and Parallel - Woofer at 12Ω

Here, we see the change when the woofer impedance is increased to 12 ohms. The series network is slightly better, but there is very little between the two. The rise at crossover frequency has changed from 3dB (normal) to 4.9dB - this will be audible in both cases.

The impedance "seen" by the drivers is also important. This may be referred to as "look-back impedance". The woofer is expected to be effectively short-circuited by the amplifier at low frequencies, and both networks achieve this quite well. Interestingly, the parallel network loses control at the crossover frequency. This is shown in the following diagram. The loss of control at this frequency is relatively unimportant if the cabinet is well damped, but may cause colouration with some systems.

In the following graph, each trace indicates the current generated when a 1V source is connected in series with the woofer. This represents the back EMF generated by the cone's momentum when the signal changes. The red trace shows the current in the parallel network, and as can be seen, it drops to a low value (high impedance) at the crossover frequency. A series network maintains relatively good control over this region, tapering off (impedance increasing) gradually.

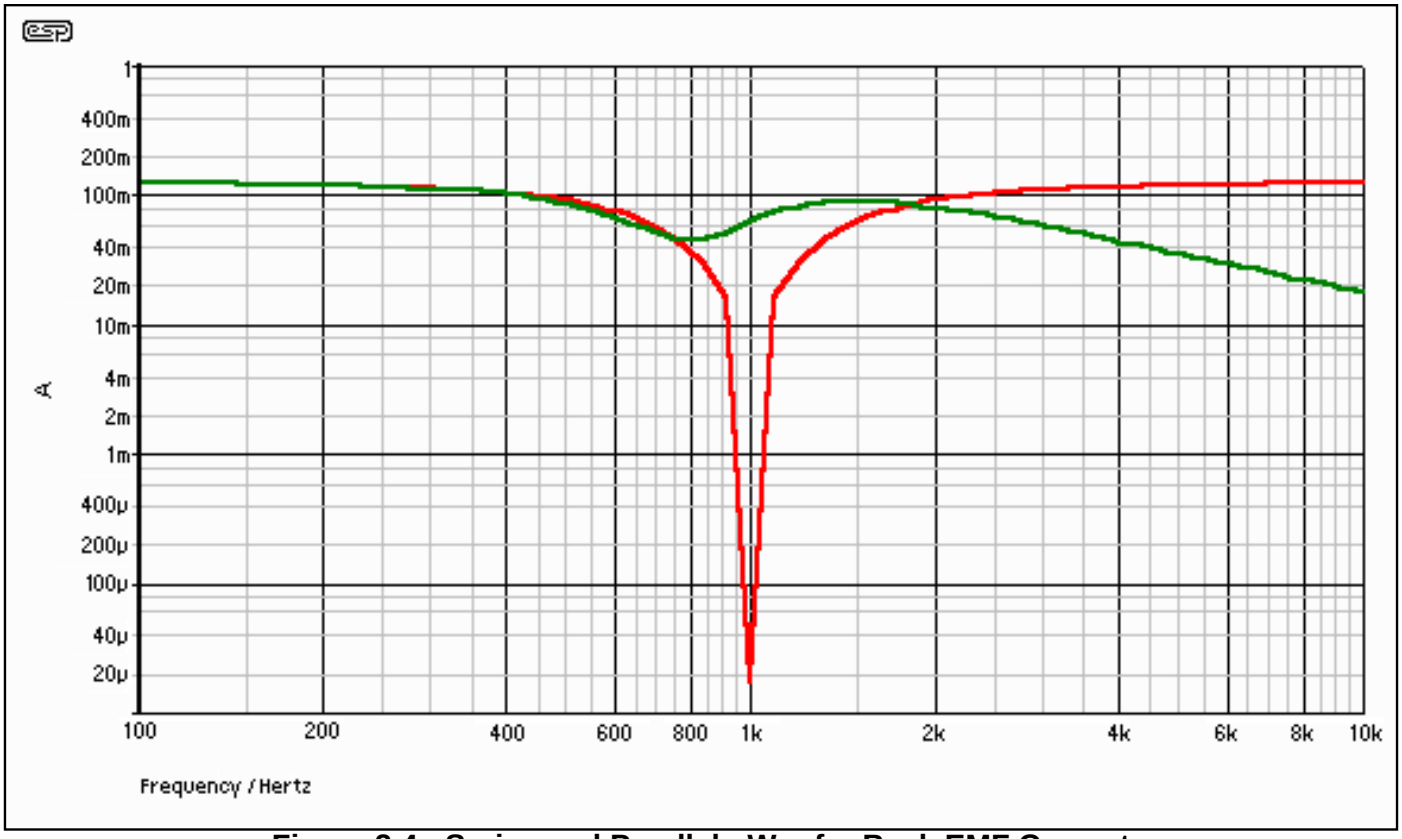


Figure 2.4 - Series and Parallel - Woofer Back EMF Current

The next test is to see how well each network maintains separation of the signal generated by the woofer. It is important that woofer back EMF (in particular) is not seen by the tweeter, as this may create intermodulation. The 2nd order network is the same as a 1st order network in this respect, except that the slope is 12dB/octave as is expected of a second order network.

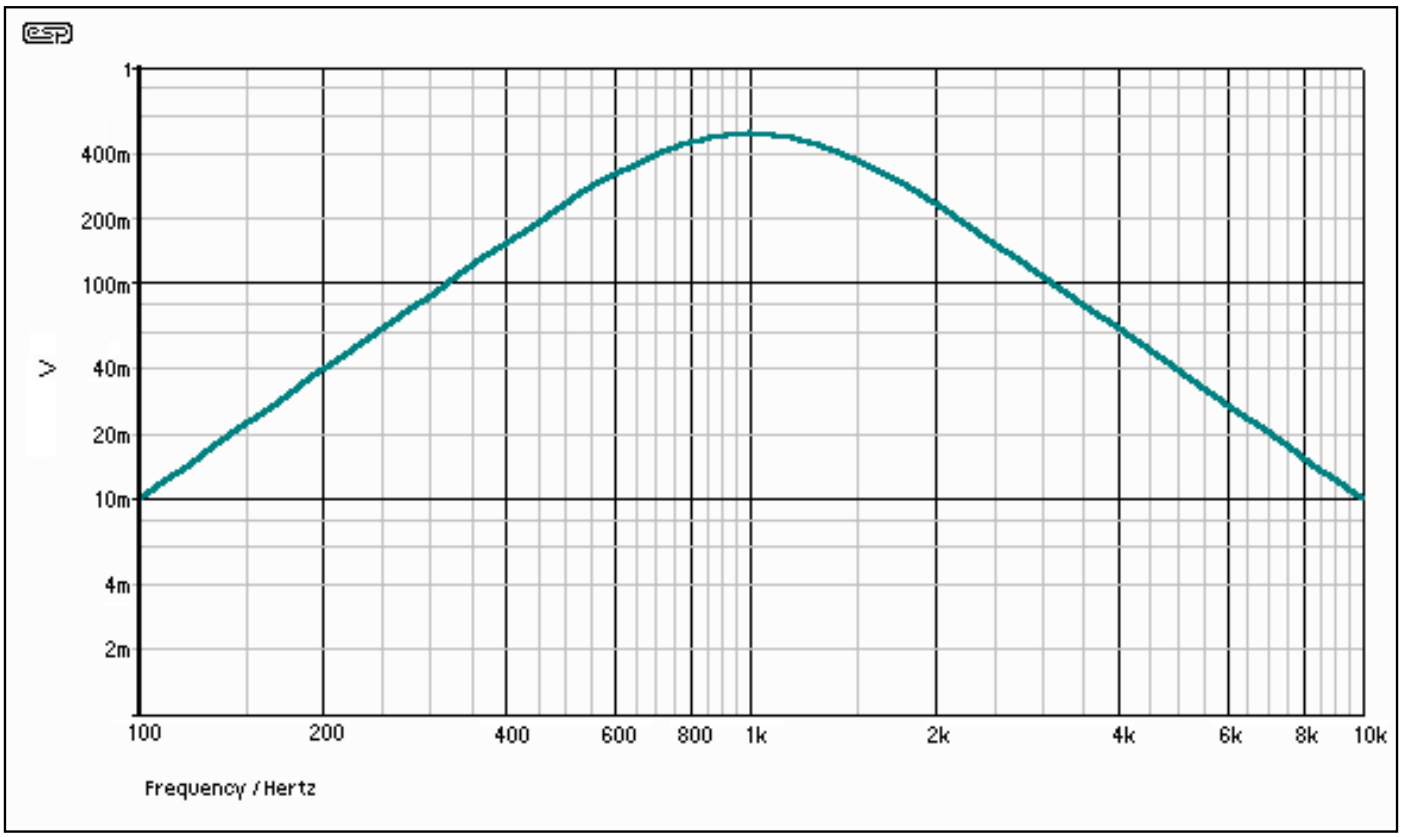


Figure 2.5 - Series, Woofer Back EMF Rejection

The amount of this signal reaching the tweeter should be zero (or close to it). The parallel network is not shown, since it is at zero. Not so good for the series network however, with more than half the generator voltage appearing at the tweeter terminals at the crossover frequency. Even at 300 Hz, the voltage is significant at 100 mV (20dB down from the full 1V applied). As with the series 1st order network, the back EMF rejection is 3dB better than the attenuation of the amplifier signal below crossover frequency.

The levels shown are not a real concern, since woofer back EMF will always be much lower than the amplifier signal. While it would seem ideal to limit such cross-coupling to the minimum possible, the effects are something of an unknown, and back EMF can be expected to be quite low with typical drivers - especially where the box is well damped internally.

Given that valve amplifiers typically have an output impedance of 6 ohms (when operated without global feedback), the differences between the series and parallel configurations become very similar, with the parallel network being only 2.7dB better than its series counterpart.

2.3 - Summary

The differences between second order series and parallel filters are more difficult to rationalise. Each has strengths and weaknesses, but from the above, the parallel version probably has a slight advantage. Both exhibit variations in response when the woofer (or tweeter) characteristics change, and they are quite similar. The parallel filter has better woofer back EMF rejection in the tweeter circuit, while the series crossover has a better woofer "look back" impedance characteristic.

Components for a series crossover will be more costly because of higher capacitor values, but it will have lower losses due to inductor resistance, since they are lower values. For those who feel that capacitors change the sound, the higher values may be thought to have a greater effect

---

3.0 - Conclusion

It is very difficult to make any judgement of series or parallel crossovers as a generalisation. The series first order network is probably a better choice in general, due to its flat response regardless of driver impedance - this can simplify the design, but at the expense of having the crossover frequency shift from the design value.

The choice is more difficult for the second order crossover, since both series and parallel have vices and virtues, with neither standing out as generally superior. Overall, the parallel version is probably a better choice, if only because it is slightly more tolerant of variations, and will probably have marginally lower losses because there is no series connection of the drivers (this adds the resistive losses in the inductors, whereas they are in parallel in the parallel filter - of course).

As for any claims for better transient response or sound quality, this is very doubtful - there is

nothing to suggest that either version if properly designed will outperform the other to any degree. Parallel crossovers are easier to design, and are simple to convert to a (sub) Bessel response with a Q of 0.5 (approximating a Linkwitz-Riley response).

Most constructors who have attempted second order series crossovers have had to spend considerable time tweaking to get it right - they are harder to design than their parallel counterpart, and interactions will always cause problems.

As a final examination, Figure 3.1 shows a series and parallel network, using simulated drivers. There is no compensation applied for woofer inductance or tweeter resonance, yet both effects are present.

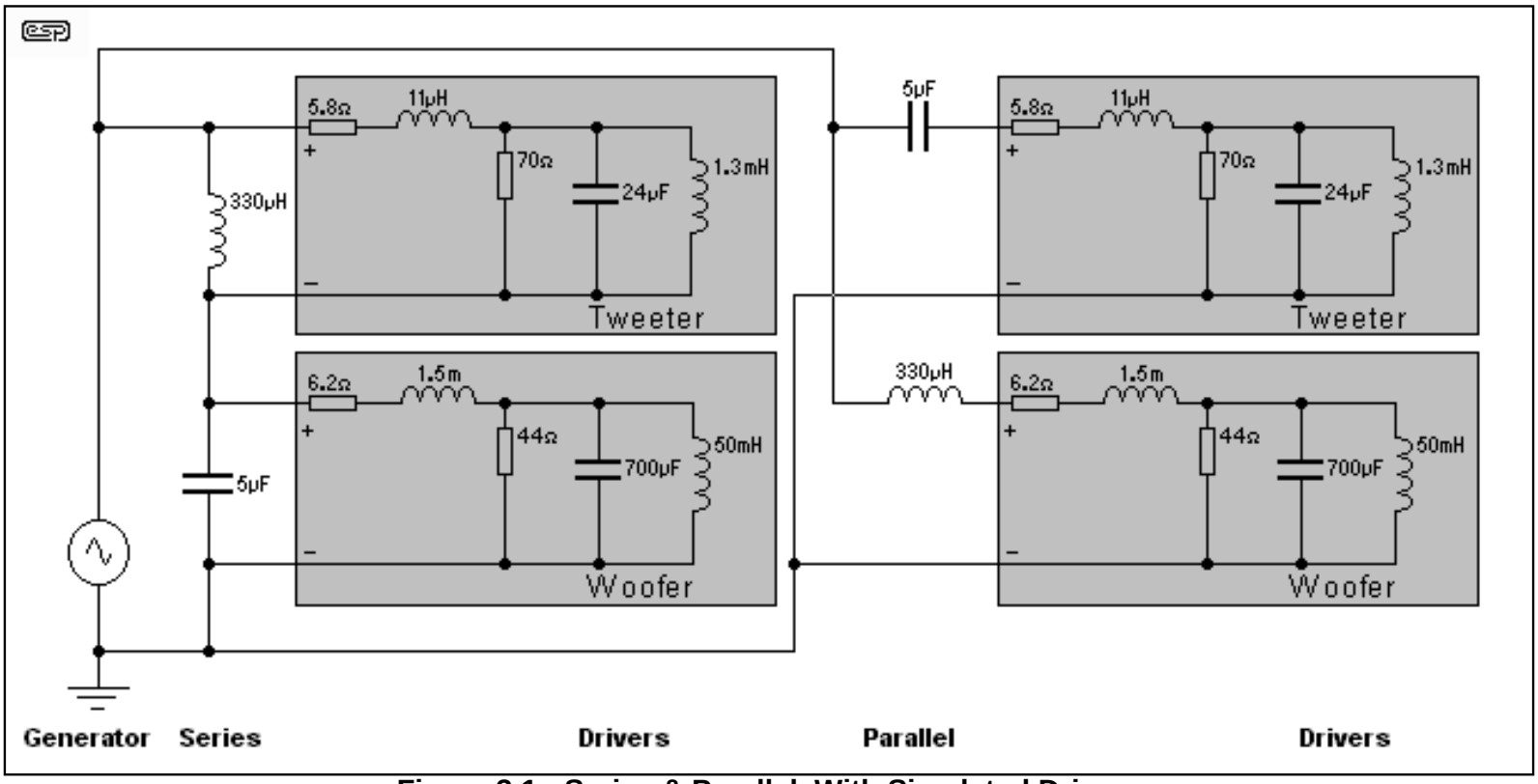


Figure 3.1 - Series & Parallel, With Simulated Drivers

The grey boxes are the drivers (identical in each version), and the area outside the boxes contains the generator and filter networks. As you can see, these are the same in each case, with the values deviating from the previous simulations only in that this design is for a real crossover network (a very similar design is used in my PC speakers, as described in the ESP projects section - see Project 73). The values are slightly different from those shown, but the principle is identical!

A transient analysis shows the following outputs, using a nominal 4kHz crossover frequency (as per the circuits above) and an input signal of 1kHz ...

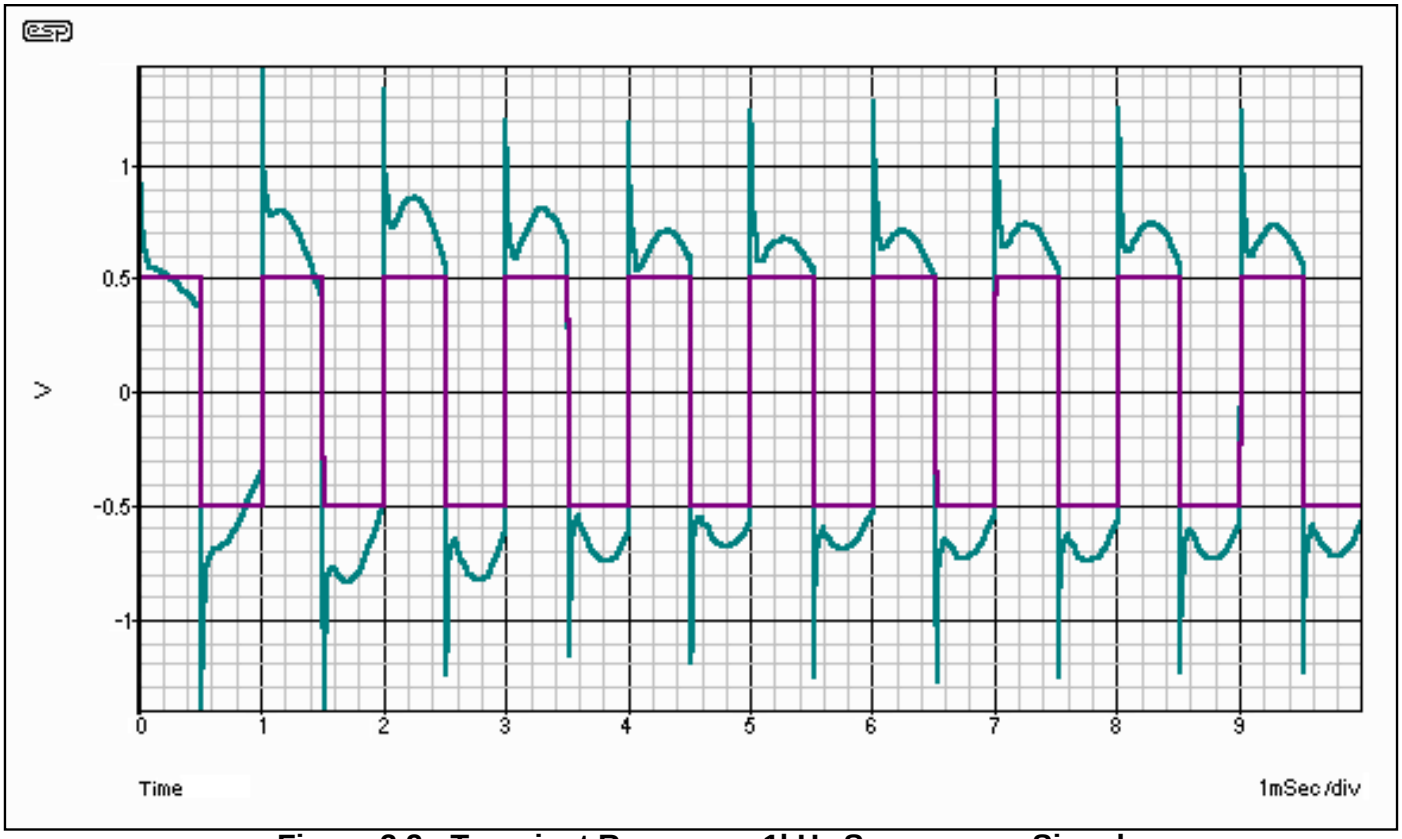


Figure 3.2 - Transient Response, 1kHz Squarewave Signal

The parallel crossover output is shown in Aqua, and the Violet trace is the series network's output. This is an electrical summing, but it shows clearly that the driver characteristics are fully compensated by the series network, and the output is exactly the same as the input. The parallel network by comparison indicates severe waveform distortion, and this implies phase and levels are incorrect - remember that no attempt was made to optimise the driver impedance with zobel or notch filters in either case.

This is fine in theory, so to prove the point one way or another, the following are real impedance and response plots from two identical (inasmuch is possible) boxes, measured under identical conditions, and within a few minutes of each other. The boxes are my PC speakers, as described above, using shielded Peerless tweeters, and small (unbranded) polypropylene woofer drivers.

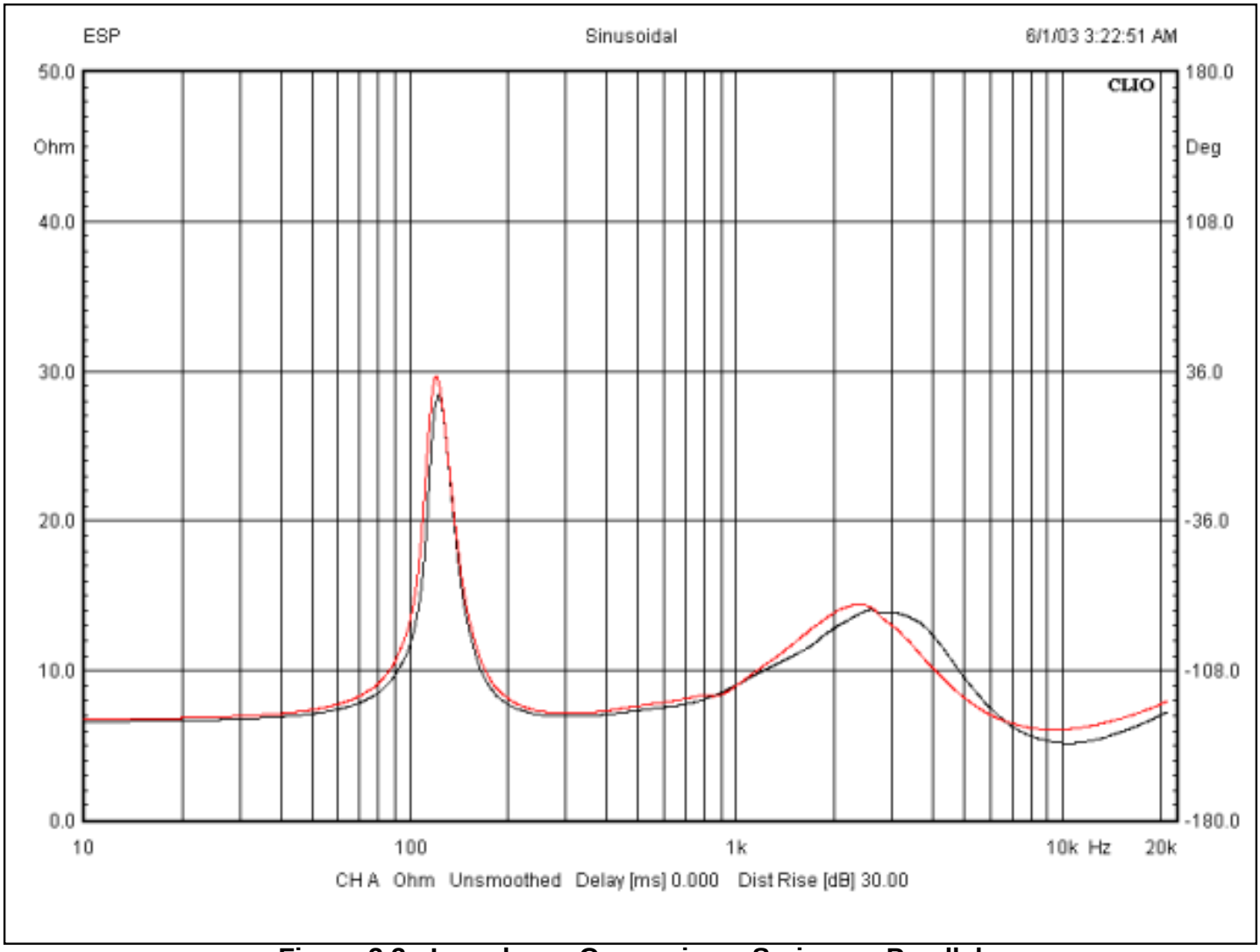


Figure 3.3 - Impedance Comparison, Series vs. Parallel

The measured impedance differences are as likely to be the result of silghtly mismatched drivers as anything else. There is not a great difference at all. The red trace is the series connection, and black is parallel.

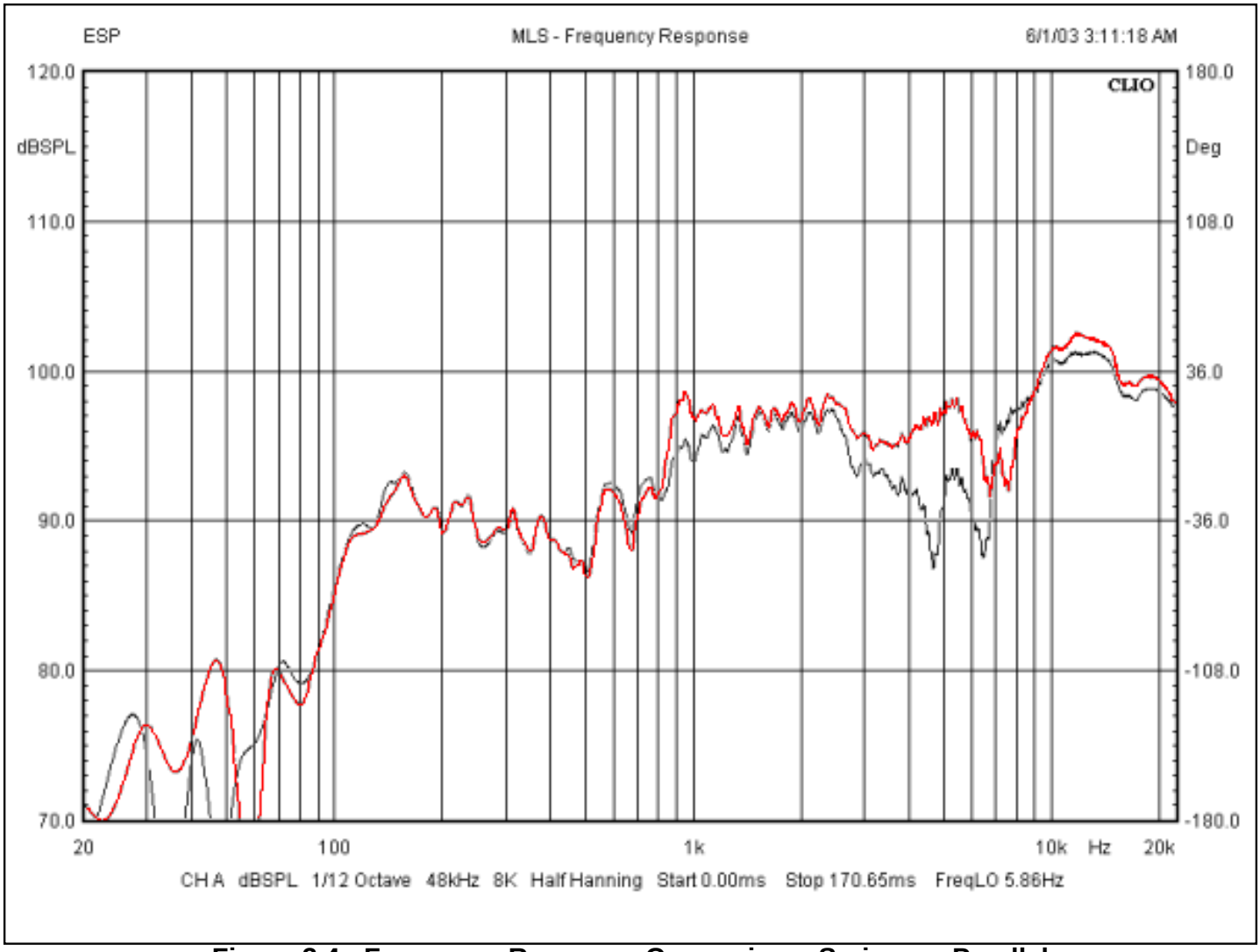


Figure 3.4 - Frequency Response Comparison, Series vs. Parallel

Response differences are a bit more pronounced (again, red is serial and black is parallel), but are not as we should expect based on the simulations. Simulation showed perfectly flat response, but remember that was an electrical signal only, and fails to account for driver behaviour. Note that there *is* a noticable improvement at the crossover frequency of 4kHz - the series network is flatter, indicating that the theory does work (the drivers have no impedance compensation - these are PC speakers, and make no claim to be being hi-fi).

Finally, after converting the second enclosure's crossover to series, I did another response comparison. As you can see, there are still differences between boxes, with one tweeter being more efficient than the other. This alone would account for some of the differences seen in the series-parallel comparison.

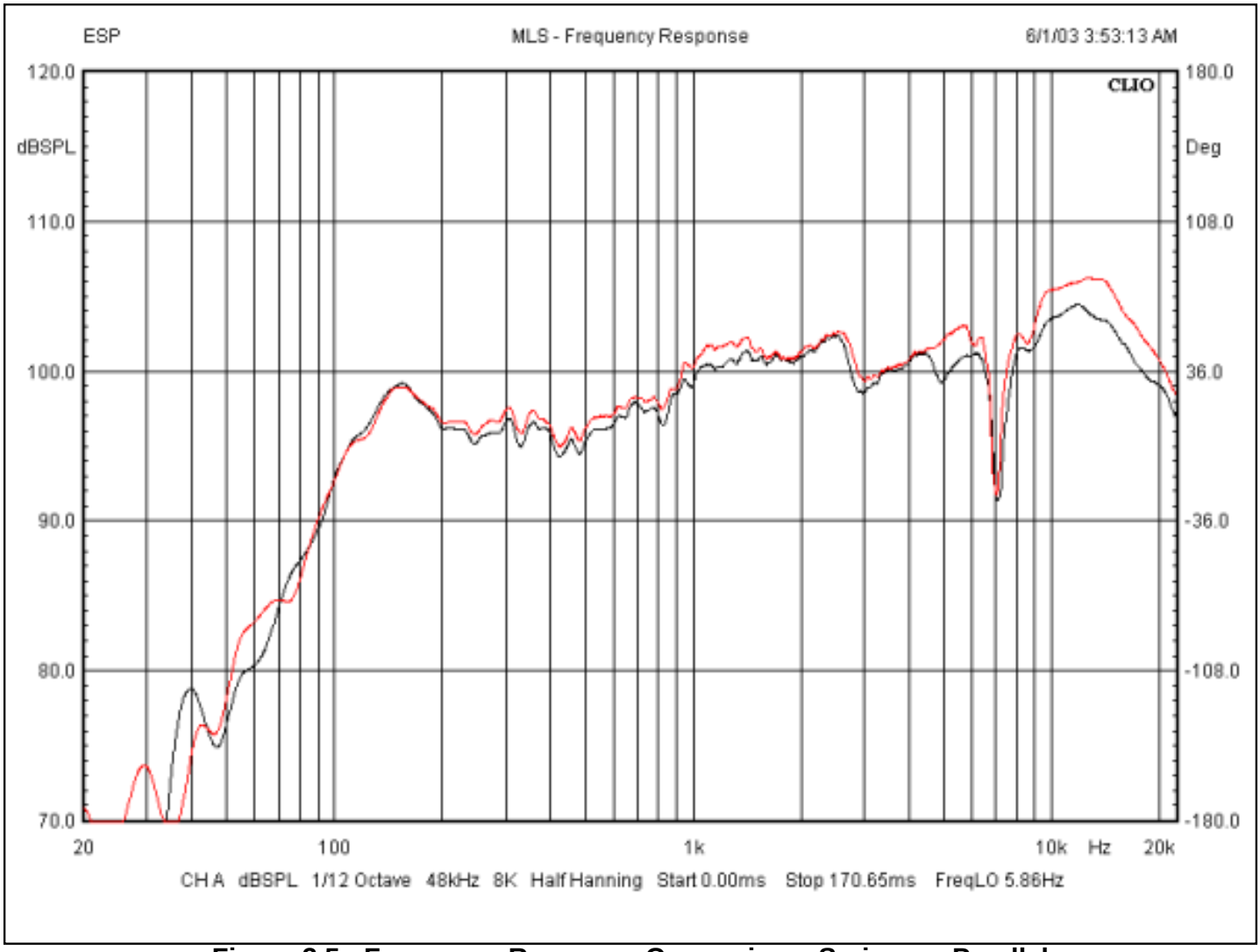


Figure 3.5 - Frequency Response Comparison, Series vs. Parallel

The glitch at 7kHz appears to be caused by a diffraction, probably from the woofer's surround (which projects slightly from the frame, and is at the correct distance for that frequency). As for sound differences between the series and parallel connections, there was very little that I could hear. The microphone is much more sensitive to small variations than the ear, and there are quite dramatic variations in response as one moves around - far greater than the differences measured between the series and parallel connections. This shows up readily if one moves the measurement mic even a small distance, and the fact that the two sets of response graphs look quite different is evidence of this. The mic was moved about 50mm further away from the speakers for the second chart.

Spectral decay plots were also done, but are not shown - there are marginal differences as one would expect from the frequency response variations, but little else.

So, although a simulation shows that a first order series crossover is superior to its parallel equivalent, the fact is that the differences are slight. The evidence was sufficiently compelling for me to change the crossovers in my PC speakers, but the huge difference in sound quality one might expect was not forthcoming. More revealing drivers may well sound better to a critical listener, but the differences are hardly "chalk and cheese" as some may imply.

3.1 - Series & Parallel Networks are (Virtually) Identical ...

Despite the differences that have been shown, the loudspeaker drivers should always be carefully equalised with zobel networks to achieve a flat impedance. Once the impedance is flat, it is resistive, and as has been shown above, the two networks are virtually identical with resistive loads. Therefore, it follows that properly executed zobels (and a notch filter for the tweeter resonance) will cause real-world series and parallel crossover networks to behave in an identical manner, with the (relatively) small difference of woofer back EMF applied to the tweeter.

The phase and transient response of both filters will match exactly with impedance equalisation, so in a *properly designed* crossover network, there is nothing to choose between the two. Certainly, the parallel variant is easier to design, and this alone is probably a good reason to stay with a parallel crossover - and probably also explains why the vast majority of loudspeaker designers use parallel rather than serial. In addition, a serial crossover cannot be biamped or biwired (assuming that you consider this important).

It is safe to say that neither crossover is possessed of any magic (only skill), so be very wary of any claims that a particular crossover design is "vastly superior" or "infinitely more transparent" (or any other hyperbole that may be thrust upon you) in advertising material. All crossovers, and indeed, all loudspeakers, are a compromise. While the "form factor" of the crossover is relatively unimportant, the skill and patience required to execute it properly is what really counts.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 30 Jun 2003./ Published 14 Aug 2003

Common Values		Parallel Crossover	Series Crossover
Crossover Frequency	1kHz	C = 14.07uF	C = 28.13 uF
Speaker Impedance	8 Ω	L = 1.8 mH	L = 900 uH

The Audio Pages

Elliott Sound Products **Phase Correction - Myth or Magic**

Copyright © 2004 - Rod Elliott (ESP)
Page Created 02 Mar 2004

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Time Delay
  - 2.1 Phase Shift Networks
  - 2.2 Results Verification
  - 2.3 Description
  - 2.4 Description
- 1.0 Conclusion

1.0 - Introduction

Phase Correction - Myth or Magic? Does it actually work? Although recommended in a lot of designs (especially for subwoofers), you cannot really correct a time delay with a phase shift network, but you can improve performance, albeit marginally some would say. An all-pass filter (aka phase shift network) may be useful in some cases, but it is certainly not the panacea that some would claim.

The best one can hope for (and it does depend a lot on the time delay you are trying to correct) is to convert a deep dip in response into a much shallower ripple at or near the crossover frequency, but it is entirely possible that the end result will not sound as good

as if the dip were just left alone.

This topic has received scant attention at most of the speaker building sites, forum pages and published books on the subject, but is obviously important since it has the potential to make matters worse than they were before you started. A discussion about "time alignment™" has already been done on The Audio Pages - see Phase, Time and Distortion in Loudspeakers for the full discussion.

Unfortunately, it is difficult to achieve true time delay in the analogue domain - it can be done, but requires a large number of components to be placed in the signal path. Since the tweeter is the driver most likely in need of the delay, this means potential degradation of the higher frequencies, something that is commonly quite audible.

2.0 - Time Delay

A time delay is introduced whenever the acoustic centres of any two loudspeakers are different from the perspective of the listener. It is not usually a problem in a three way system between the woofer (or subwoofer) and midrange or mid-bass, but between the mid/ mid-bass and tweeter there is usually considerable room for misalignment in the time domain.

For this exercise, I have used a crossover frequency of 3.0kHz, and the offset between drivers' acoustic centres is 100us (or 35mm, close enough). It is useful to establish the relationship between distance and time, and this may be determined by ...

wavelength = velocity / frequency
period = 1 / frequency

So for any given frequency we can determine the wavelength and period (the time for one complete cycle). For example, at the selected crossover frequency of 3.0kHz, the wavelength is ...

wavelength = 345 / 3000 = 115mm
Period = 1 / 3000 = 333us

It is useful to understand that there is a simple relationship between time delay and effective "phase" shift in terms of wavelengths and the velocity of sound in air (345m/s at sea level, approx 22°C and 50% relative humidity).

For example, if the delay were 166us, a 3.0kHz signal is delayed by exactly 1/2 wavelength, or an equivalent distance of 57mm. This can also be determined by realising that sound travels 0.345mm/us, making calculations rather easier.

2.1 - Phase Shift Vs. Time Delay

While it is possible to make a phase shift network that has a consistent phase shift over a wide range of frequencies, this is not the same as time delay. It can be considered obvious that if two signals are delayed in time by exactly the same amount that the end result is identical to no time delay at all. Likewise, if two signals are affected by identical phase shift networks, the net result is again the same (although there is now a phase shift in the signal which may or may not be audible, depending on many different circumstances).

Having a time delay (in the case of loudspeakers introduced by a physical misalignment of acoustic centres), and attempting to correct this with a phase shift simply does not work as well as one might hope. This should be obvious, but it is still touted as the "proper" way to realign the drivers. This can be shown in the simplest case with a time delay that is exactly 1/2 wavelength. For a 3kHz crossover point, this equates to wavelength ÷ 2, or 57.5mm (which can also be described as a time delay of 166us). The graphs shown below are all done using a 24dB/octave Linkwitz-Riley crossover - the effects with lower orders are very much worse than shown here.

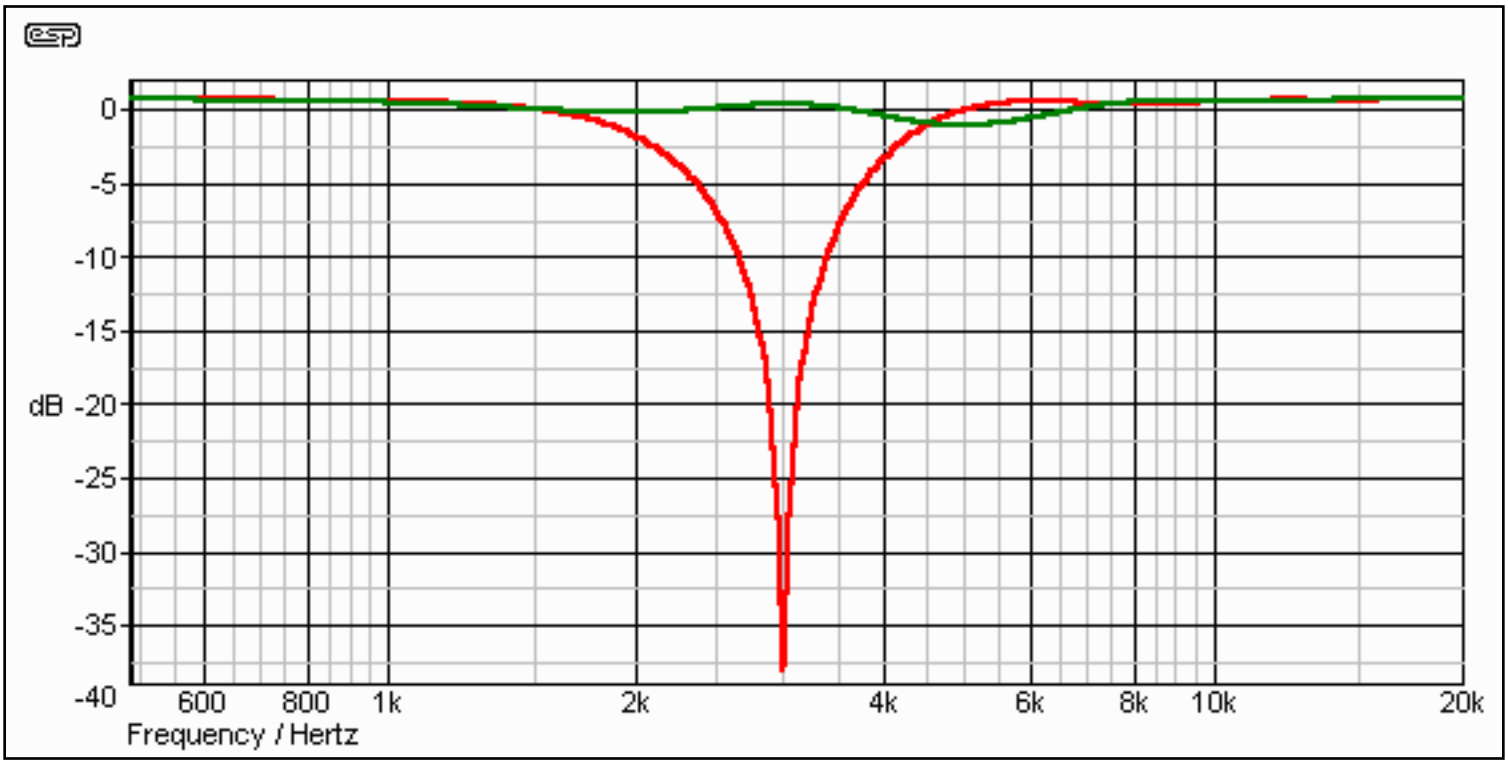


Figure 2.1 - Frequency Response With 57mm Acoustic Centre Offset

The red graph shows the response without phase reversal of the driver, and it is very obvious that there is a significant response dip at the crossover frequency. This was done using electrical summing, which is much more severe than acoustic summing, however the "suck-out" would be very audible indeed.

The green graph is the response with the tweeter reversed in phase, and it looks better than it really is because of the graph scale. However, the response ripple is still within (barely) acceptable limits, at -1.7dB at 4.9kHz. Again, this is with electrical summing - the acoustic response will be somewhat better than shown.

Next, we see the response of the same system, but with a 100us delay. With the tweeter wired out of phase, the response dip is 2.8dB, and when wired in phase it is 5.8dB - the anti-phase connection is better, but neither is acceptable.

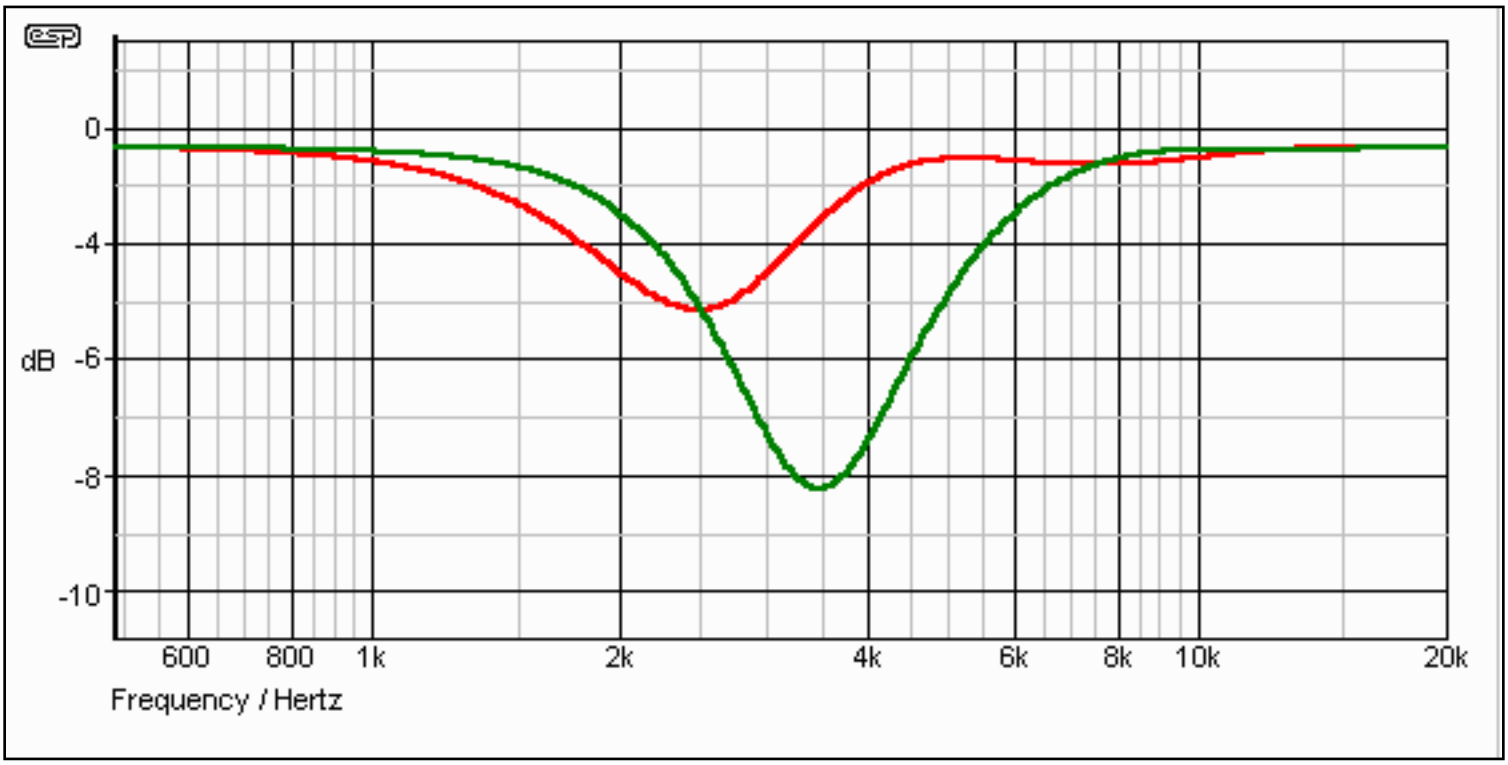


Figure 2.2 - Frequency Response With 35mm Acoustic Centre Offset

As you can see, both graphs show a distinct dip in response (the red graph is the out of phase connection). Acoustic summing will give a result that is slightly better than what you see here, but it will still be audible. The optimum fix is a digital delay that will hold the tweeter signal back by 100us, but this is an expensive option. The next diagram (Figure 3) shows the response using an all-pass filter (phase shift network) that has been optimised to get the best response possible from the combination of drivers and

time delay.

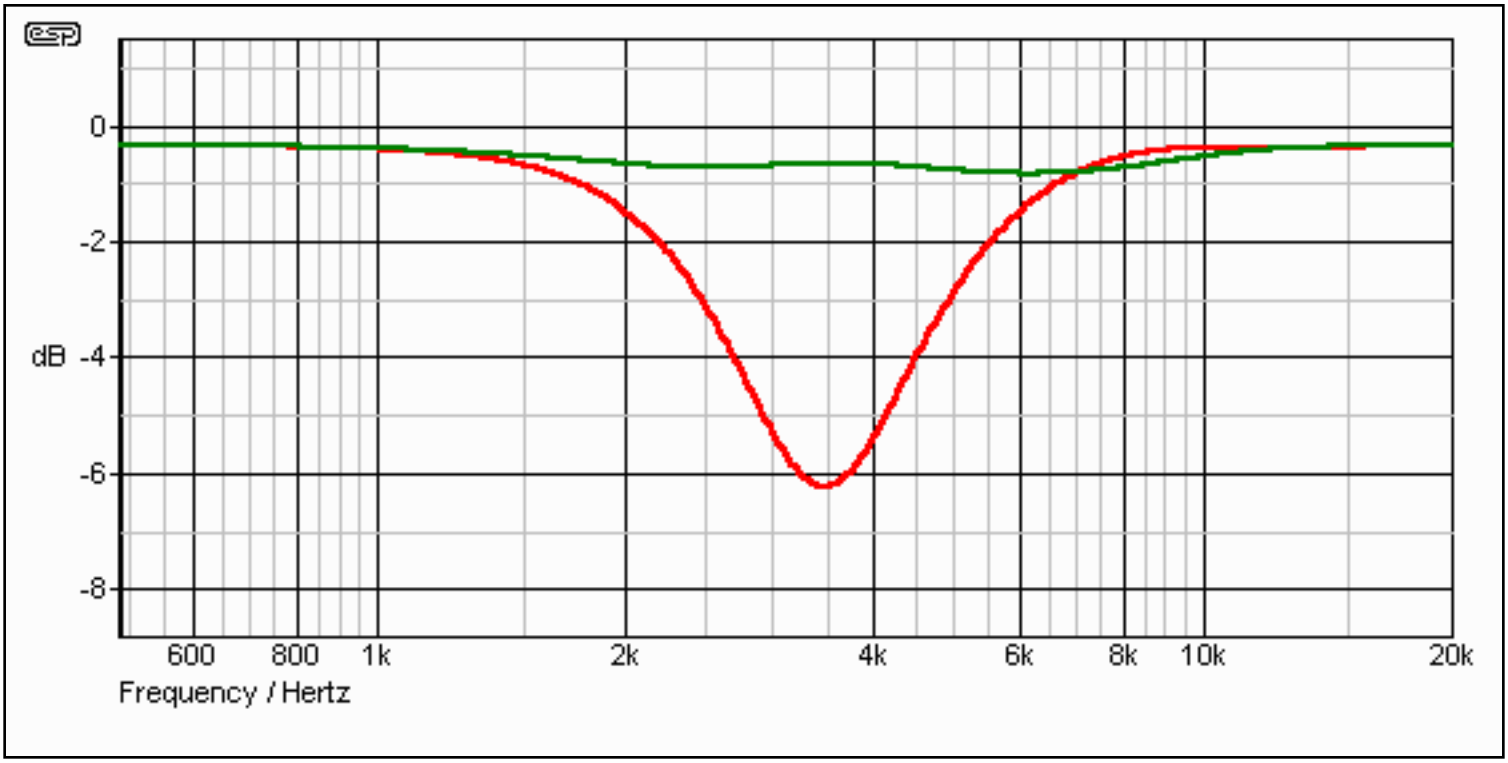


Figure 2.3 - Frequency Response With Phase Compensation

The maximum deviation of the phase corrected version (shown in green) is now 0.51dB, at 6kHz. It is actually possible to improve on this slightly, but in practice this requires accurate measuring equipment. Attempting to perform an adjustment of this nature without measurements will simply lead to confusion and great bafflement, since you have no flat reference for comparison. The setting that sounds the best (to you) may be the result of personal preference or acclimatisation to the sound, and it is very possible that it will not be correct.

In addition to the phase adjustment, it is important that polarity inversion is available. Although the above example uses normal polarity for best results, this will not always be the case. If the polarity is incorrect, no amount of phase adjustment will ever make the situation better, but it can easily make it a lot worse. For example, the response shown in Figure 2.3 is totally changed (and very much for the worse) if the phase is inverted.

2.2 - Phase Shift Networks

The standard and traditional all pass filter (phase shift network) is shown in Figure 3.1 - note that there are two variants, and their behaviour is completely different when phase aligning a loudspeaker. 3.1.a shows a non-inverting configuration, and 3.1.b is inverting.

This is something of a misnomer, since both circuits will have an output whose phase is frequency dependent, but version 3.1.a leaves low frequencies alone and inverts high frequencies, while 3.1.b inverts low frequencies. At the mid frequency of the network, the phase is rotated by 90°, but note that this is not necessarily aligned with the crossover frequency.

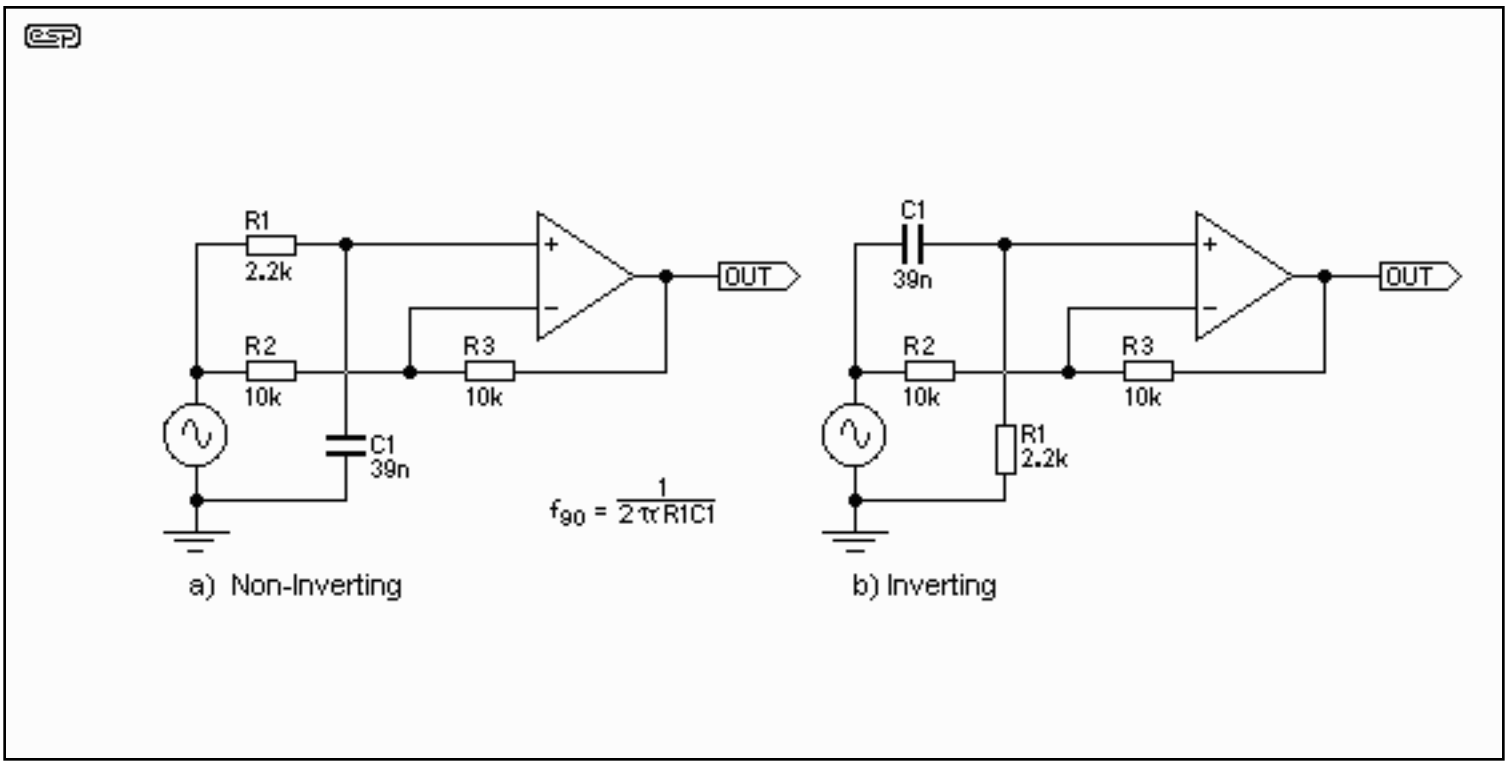


Figure 2.4 - Inverting and Non-Inverting Phase Shift Networks

The networks themselves are virtually identical, with only the position of the resistor and capacitor (R1 and C1) changed. This effectively reverses the frequency at which there is no phase shift - in this case we would arbitrarily select around 10Hz as the 'reference' frequency. At such a low frequency, the capacitor has virtually no effect with either network, and the circuits can be considered to be (almost) conventional inverting and non-inverting opamp circuits.

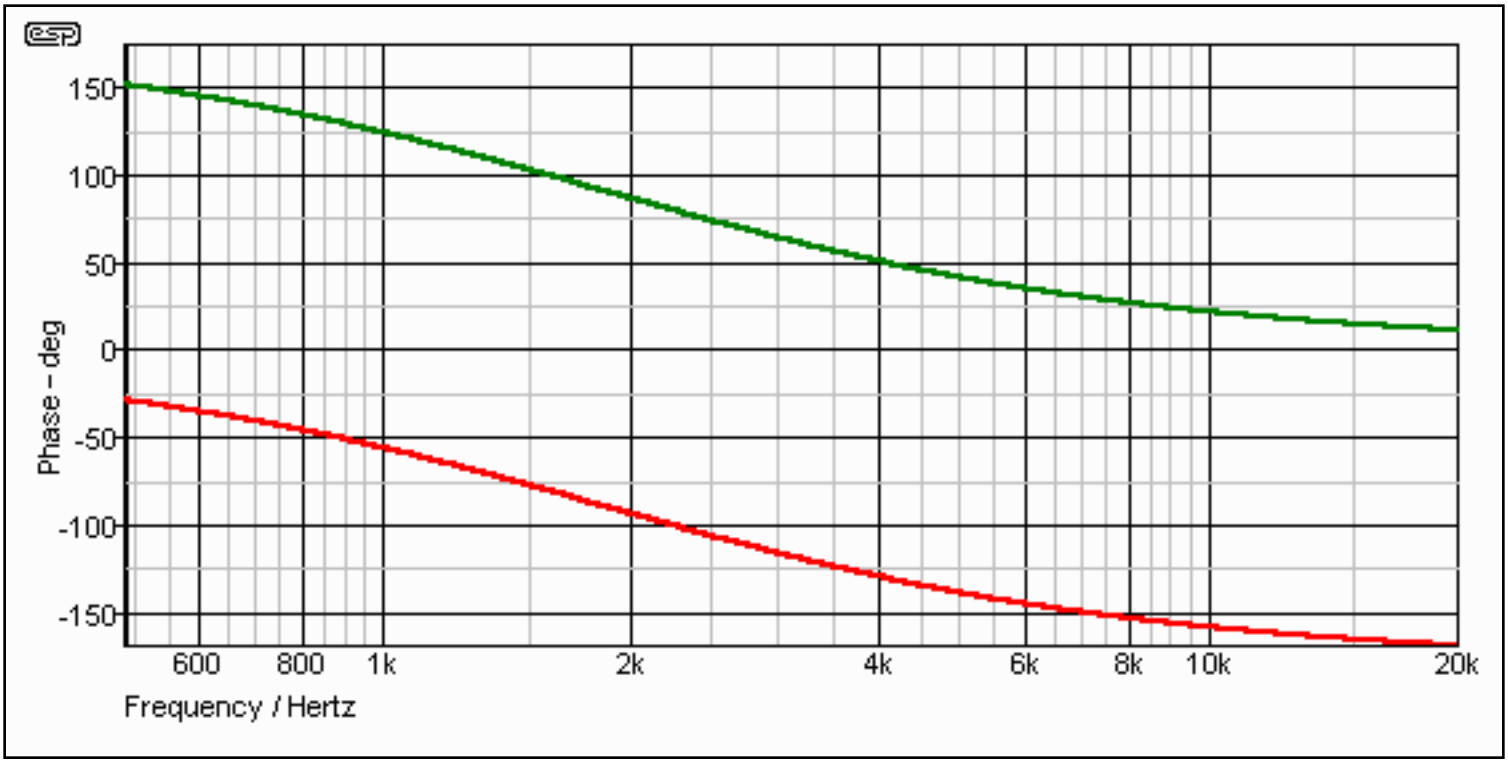


Figure 2.5 - Inverting and Non-Inverting Phase Shift Response

The networks as shown are exactly the same as those used for the alignments shown above. The centre frequency for optimum alignment is not 3kHz as one might expect, but for the example shown is 1,856Hz. Phase shift at 3kHz is 116° or 64°, depending on the phase polarity selected. 116° is the phase shift that gave the best result.

2.3 - Results Verification

If everything has worked as intended, we should get a good correlation between the applied phase shift, time delay and wavelength. Since we know that the time delay used was 100us and frequency was 3.0kHz, we can calculate the effective wavelength that corresponds to the phase shift ...

Wavelength	115mm
Time Delay	100us
Distance Offset	35mm
Phase Shift	116°
Effective distance compensation	(115mm / 360°) * 116° = 37mm

Well, that seems to be very close indeed. Since the non-inverting phase shift network

was used, it has 116° of shift at 3kHz (after careful adjustment), hence the calculation above for the "effective distance compensation". Now, looking at the graphs, it is obvious that they are not perfect, but the end result is a lot better than one could expect without any compensation at all.

Sidenote
There is a relatively simple way to introduce a time delay at any frequency. Coaxial cable has a velocity factor of around 0.7 (typical) - that means that the signal travels at 0.7 of the velocity in a vacuum. Unfortunately, 100us is rather a long time, so given that normal speed of a signal is 3×10^8m/s, for a velocity factor of 0.7 (for example), we would need $3 \times 10^8 * 0.7 = 210$Mm (Mega-metres!) for a one second delay.

Therefore, 210 metres will give a delay of 1us, so you would need 21km of coax to obtain the required delay of 100us. For what that would cost, one could purchase any number of digital delays. In case you were wondering, the idea is not completely silly. High-end oscilloscope manufacturers used to include a coax delay line to delay the signal for the few nano-seconds that it would take for the trigger circuits to operate, so that one saw the waveform without a 'glitch' at the beginning of the trace. This typically only required a few metres of cable (perhaps 10 metres at the most).

Lest one think that knowledge of such technology is restricted to old buggers like me, a web search will reveal that there are still a great many manufacturers of coax delay lines, and they are far from dead. They just happen to be useless for this application.

3.0 - Conclusion

While this discussion has dealt only with 24dB/octave Linkwitz-Riley alignment crossovers, the principles are recommended by many to be applied with other orders as well. First order (6dB/octave) crossovers are likely to benefit the least, because the time delay created by a phase shift is too narrow, and is completely incapable of creating an offset that is broad enough to be useful. This is illustrated below, where the same time delay (100us) was used, and modification of the phase shift network was only able to achieve the response shown.

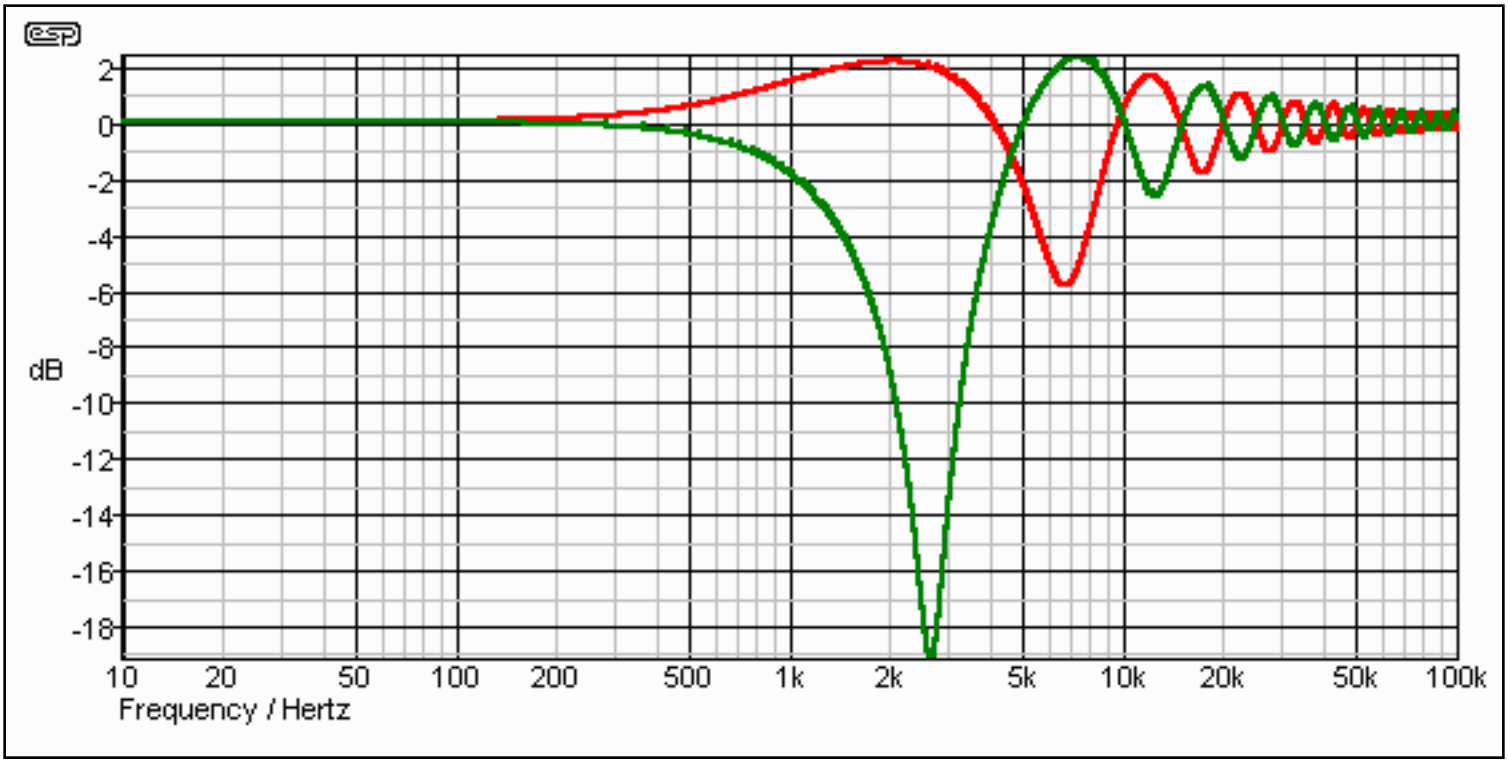


Figure 3.1 - Applying Phase Correction to a 1st Order Crossover

Note that the uncorrected response is a lot worse than the corrected response, but even there, the ripple is over 7dB in peak variation. The *best* one could say for the response is that it is appalling, both before and after 'correction'. There are terms that describe it far more accurately, but they will not be used here

It is fairly obvious that since the range over which an all pass filter can be used is quite narrow, it becomes more effective as higher order filters are used. While it has been shown that the effects are reasonably good with 24dB/octave crossovers and useless with 6dB/octave, it follows that 12dB and 18dB crossovers will fall in the middle, with 12dB types being considerably worse than 18dB networks.

It is worth noting that a phase reversal of the all pass filter changes very little. The first major peak is converted into a dip, but the overall ripple remains the same, at 7dB. As always, these tests were done with summed electrical signals, so the acoustic effects will not be as severe ... but! They will still be there, and *will* be audible.

The next chart shows the response of a corrected and uncorrected 12dB/octave filter. This filter is the same frequency as before (3kHz), has the same 100us delay, and is a Linkwitz-Riley aligned type. A phase inversion is mandatory for 2nd order filters and there is no point showing the result without it. (Hint - a 10dB dip at 3kHz is not a pretty sight.)

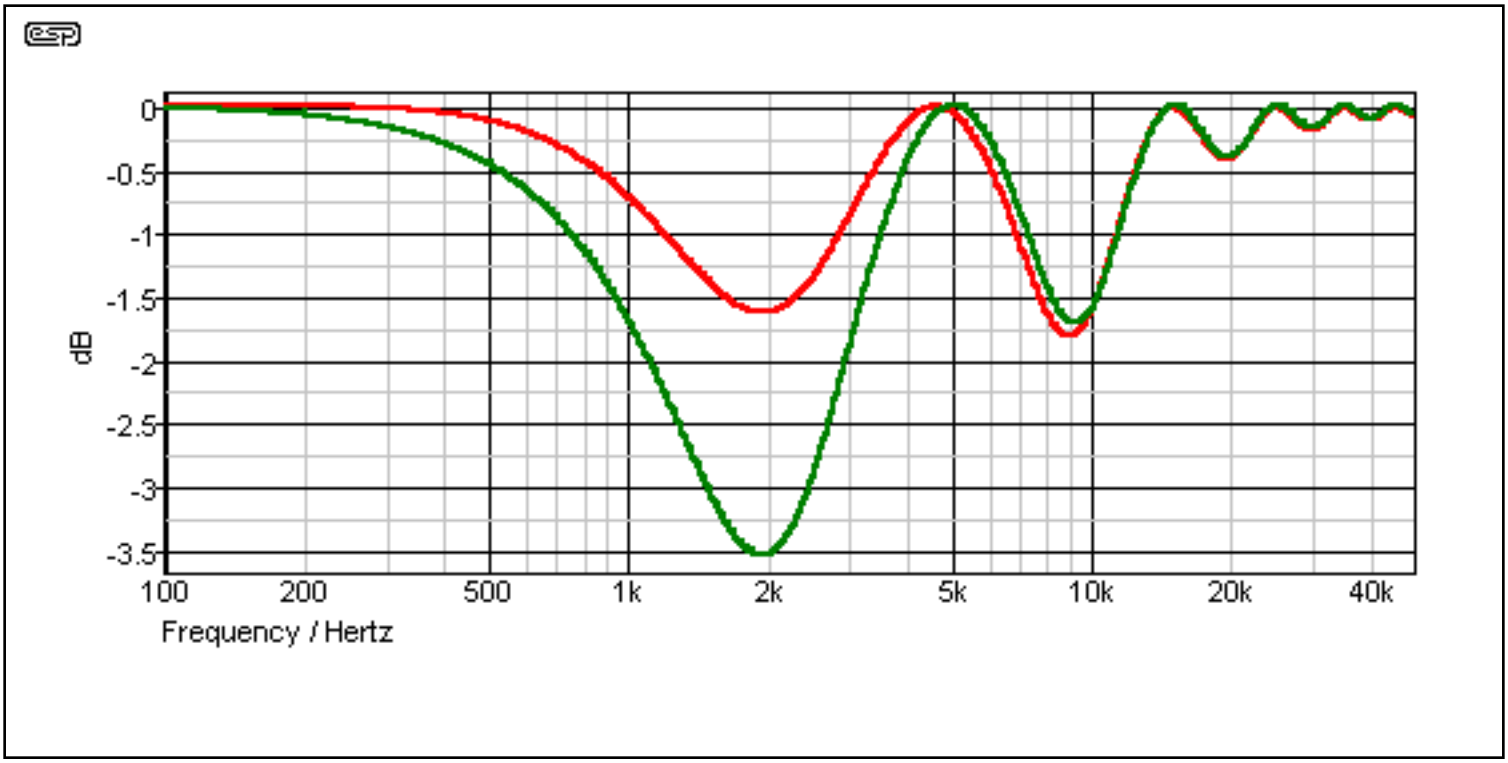


Figure 3.2 - Applying Phase Correction to a 2nd Order Crossover

The result is not completely dreadful, but it is certainly 'sub-optimal' (a kinder and gentler way of saying what I really think). Again, electrical summing is the worst case and the acoustic response will not be quite as bad. A phase shift network will probably rescue an otherwise unusable design in this case.

So, is an all pass filter a panacea for misalignment problems? Quite obviously not, since even at 24dB/octave the ripple is quite visible, and will be quite audible as well to anyone who knows what to listen for. However, it is still a lot better (at least in the amplitude domain) than nothing at all. Stepped baffles used to time align the drivers may look quite high-tech, but the step creates what I refer to as a 'diffraction engine' - a set of internal and external angles that will create interference patterns and response anomalies that are difficult to predict, but will certainly not improve matters.

What of sloping baffles? While there are no untoward diffraction effects (other than those found in any relatively conventional enclosure design), one is forced to listen to both drivers off-axis at all times. While the deviations may be small compared to on-axis response, they are nonetheless audible, especially where the midbass is working at the limits of its frequency range. For those who claim that all baffles with more than one driver should (*must*) be sloping to get time alignment, consider the difference in path length from the top and bottom of the midbass to the listener, and then compare that to wavelengths. Exactly the same effects are at work as with separate drivers, and this is the main reason that off-axis response of large diameter drivers is so poor.

One parameter that is potentially very important is group delay. The uncorrected 6dB filter shows a group delay that peaks at 0.8ms (800us), so in one respect we can be thankful for the fact that there will be a deep notch at that frequency, so the (very) delayed signal will not be audible. The concept of group delay is best explained by a fairly common description on the Net. If the treble were to be reproduced 5 minutes (or even 5 seconds) after the bass and mid, this would be noticed by even the most non-critical listener. Fortunately, this is not common even with the cheapest of speaker systems, but it illustrates the point. In reality, group delay (the delay introduced to any frequency relative to any other frequency), is normally very low, but can still be audible, and the audibility is frequency dependent. In the case of the midbass to tweeter crossover, a delay of 100us would be fairly typical, and is less than 1 cycle of the signal at typical crossover frequencies (a 'periodic time' of 100us represents a frequency of 10kHz). Although this delay in and of itself is inaudible, the phase cancellation effects are very audible indeed.

This is not going to be covered further here, not only because it will confuse the issue and make the processes described seem more complex, but also because I do not have enough data on the audibility of group delay to make an informed comment. This will change, and a follow-up can be expected to explain this further.

For those who might wish to get a little background into group delay, see the discussion at TrueAudio.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © Rod Elliott, 02 Mar 2004

Elliott Sound Products **Measuring Loudspeaker Parameters**

Rod Elliott - ESP
Updated 12 January 2006

Articles Index

Main Index

Please note that the process described here had an error (corrected 12 Jan 2006), and the values for Qms (and therefore Qts) values were wrong in the example. The error has been corrected, and the figures should now be as accurate as it is possible to get them. I apologise for any inconvenience this may cause (or may have caused).

The problem was not with the method as described, but in the accuracy (or lack thereof) of the measurements taken from the speaker model used (I could blame the simulator I used at the time, but that would be a cop-out - even though it is partially true).

The spreadsheet has also been changed (enhanced), and it is recommended that the latest version be downloaded to get accurate results. Interestingly, of four different methods I have been able to locate thus far, they all give (slightly) different results, however, measurement accuracy is the biggest issue facing your endeavours.

Measuring Thiele / Small Loudspeaker Parameters

There are several different ways to measure the Thiele/Small parameters of a loudspeaker driver. The method described here provides a way for the beginner and DIY enthusiast to measure the parameters without any expensive or specialised equipment.

Definitions:

Re	Electrical resistance of voice coil
Fs	Resonant frequency of loudspeaker moving mass (in free air)
Qes	Electrical Q of loudspeaker
Qms	Mechanical Q of loudspeaker
Qts	Total Q of loudspeaker
Vas	Equivalent air volume of moving mass suspension

Measuring Re, Fs, Qes, Qms and Qts

To measure these parameters using the method outlined below, you'll need to have the following items:

- A power amplifier, rated at 1-10 Watts (RMS) or thereabouts
- Audio frequency oscillator (PC based is fine)
- Digital multimeter (with frequency measurement), or PC based instrument
- An accurate test resistor (any value, although I suggest 10 ohms) A ½W component will be quite sufficient.

- Alligator clip leads - you will need 4 sets of leads (leads may be soldered instead if desired)

Figure 1 shows a typical impedance curve for a loudspeaker (see Figure 5 for the equivalent circuit of this speaker, which was simulated for this article). Resonance causes a large increase in impedance, and at some higher frequency, the inductance (or semi-inductance) of the voice coil causes the impedance to rise again. The region for the initial measurements must be within the "linear" region of the impedance curve. In the example below, resonance is at 27Hz, and the linear region ranges from about 100Hz to 400Hz.

At resonance, the speaker impedance is pure resistance. As the frequency increases towards resonance, the impedance characteristic is inductive. Beyond resonance as impedance falls, the impedance characteristic is capacitive. Within the 'linear' region, the impedance is again (almost) resistive, but at slightly below the speaker's nominal impedance (nominal impedance is usually taken as an average value over the usable frequency range). At the frequency where the inductance of the voice coil becomes significant, impedance rises, and is progressively more inductive as the frequency rises. It is common to add a compensation network to maintain an overall resistive characteristic at these higher frequencies, so that the performance of the (passive) crossover network is not compromised. This is not necessary with an active crossover.

Although a 'pure' inductance is shown in the equivalent circuit, this component is often referred to as 'semi-inductance'. Because of losses (primarily eddy current losses within the pole pieces), the impedance typically rises at around 3-4dB/ octave, rather than the expected (and simulated) 6dB/ octave. This has little or no effect on resonance parameters, and can usually be ignored for these measurements.

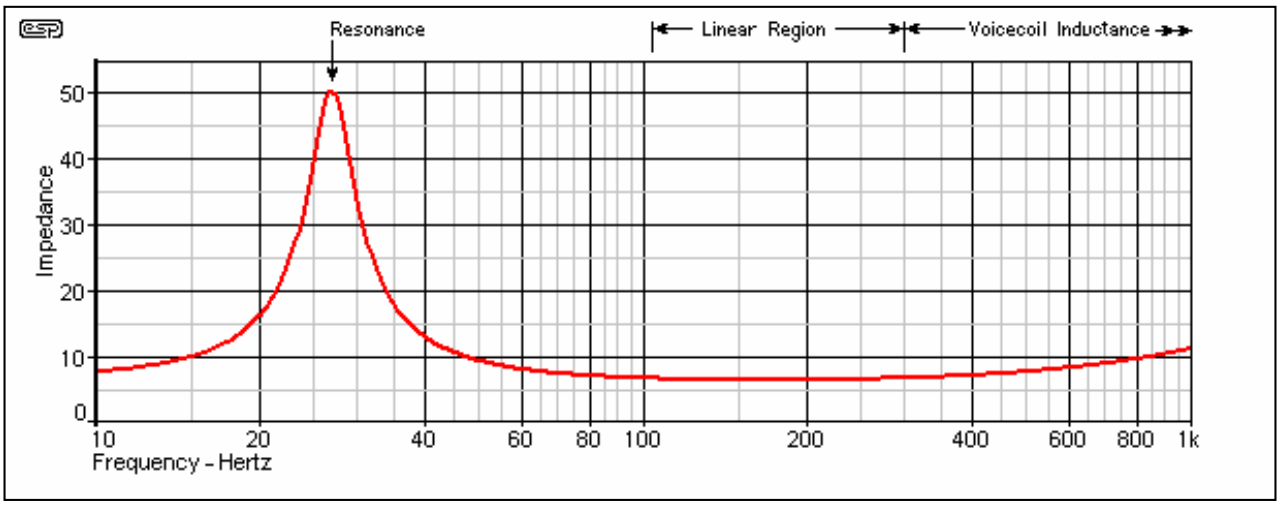


Figure 1 - Loudspeaker Impedance Curve

The multimeter should be capable of measuring frequency, as well as AC voltage and resistance. If it cannot, a frequency counter is highly recommended, since the frequency measurements are critical. The amplifier must be capable of reproducing from 10 Hz to 2 kHz with no variation in output voltage. It is imperative that it is insensitive to any load above 4 ohms. The audio oscillator must also produce a signal with relatively low distortion, and the output voltage must not vary as the frequency is adjusted. If a PC signal generator is used, it will usually display the frequency fairly accurately, but you still need to verify that output level is constant with frequency. Many PC instruments are incapable of fractional frequencies, which may limit the accuracy of the final result.

The need for accuracy cannot be stressed too highly if accurate parameters are expected. It should also be understood that there are many variables and many opportunities for things to go awry - during measurement, construction and normal operation. Loudspeakers are variable beasties at best, and 'perfect' results will never be obtained in practice. The room will usually cause more and greater errors than a small measurement error here.

Measure the resistance across the speaker terminals to obtain *Re*

Measure the exact resistance of the 10 ohm source resistor, *Rs*

The loudspeaker driver should be suspended in free space, with no obstructions or interfering surfaces nearby. Any boundary closer than around 600mm (about 2ft) will affect the accuracy of the measurements

Connect the circuit as shown in Figure 2, and set the oscillator to approximately 100 to 200 Hz (or around 2-3 octaves above resonance) - it must be within the 'linear' range as shown on the graph above.

Set the output of the amplifier to between 0.5V and 1.0V (this is *Vs*). Check that the speaker is nowhere near resonance, by changing the oscillator frequency by 50Hz or so in either direction, and measure the voltage across the resistor. It should not change by any appreciable amount.

You may need to try different voltages, depending on the accuracy of your readings (or calculations). Do not be tempted to use a voltage any higher than around 1V RMS , as the speaker may be driven outside its linear range, which ruins the validity of the measurements. The parameters being measured are 'small signal', and it essential that a small signal is actually used. With an 8 Ohm driver, 10 Ohm resistor and 1V signal, you will typically have a nominal current of around 55mA.

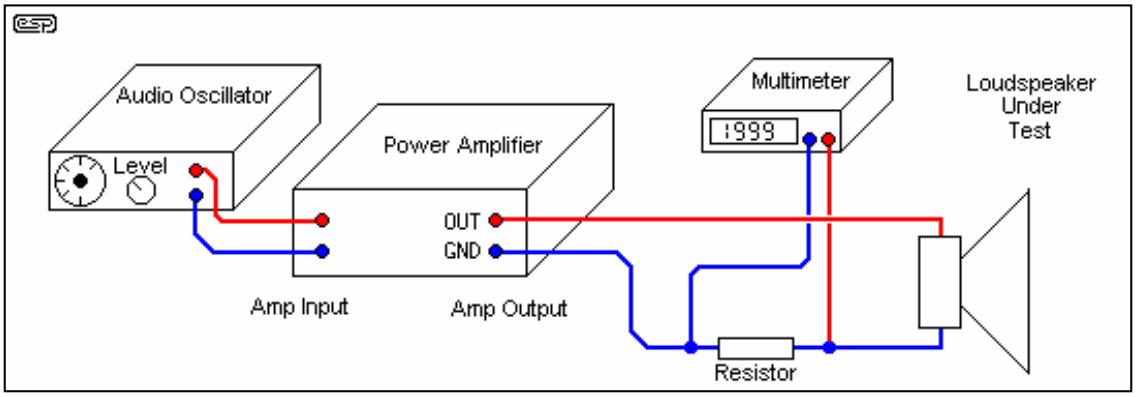


Figure 2 - Measuring Speaker Parameters

The traditional way to measure Q is to measure the bandwidth between the -3dB frequencies, then divide the resonant frequency by the bandwidth. For example, if resonance is at 29.6Hz and -3dB frequencies are at 25Hz and 35Hz, then Q is 2.96. This would be Qms in the calculations. This method may be suitable for low-Q drivers, but you can easily make a tiny error (causing a large final calculation error) with high-Q drivers.

Most of the methods described elsewhere rely on a more complex formula that uses -6dB or even -9dB as the reference point to determine Q. This makes the measurement accuracy slightly less critical. The method below describes the -6dB method, which gives a reasonable compromise between ease of measurement and accuracy.

First, measure the resonant frequency. Adjust the frequency until the voltage across the resistor reaches a null (minimum level). Without changing anything, carefully measure the frequency and voltage across the resistor ...

Frequency	*Fs*
Voltage across the resistor	*Vm*

Calculate the following ...

Speaker current	*Im = Vm / Rs*
Resonance impedance	*Rm = (Vs - Vm) / Im*
r0 (reference value)	*r0 = Is / Im*

-6dB current	*Ir = √(Im * Is)*
-6dB voltage	*Vr = Ir * Rs*

Complete the measurements for *Fl* and *Fh*, for which the voltage across the source resistor is equal to *Vr*, and as a sanity check (to ensure that your calculations and measurements are accurate), calculate the resonant frequency based on these last two measurements. Note that these measurements are critical, and even a small error will cause large deviations in the driver parameters.

Check that ...	*√(Fl * Fh) = Fs*

If the above checks out (within 1Hz or less), then *Qes, Qms* and *Qts* can be calculated as follows ...

Mechanical Q	*Qms = Fs * √r0 / (Fh - Fl)*
Electrical Q	*Qes = (Qms / (r0 - 1)) * (Re / (Rs + Re))*
Total Q	*Qts = Qms * Qes / (Qms + Qes)*

You can use the spreadsheet to perform the calculations automatically for you: **ls-param.zip**

Measuring Vas (equivalent air compliance)

To measure Vas, use a good solid enclosure of known volume that is approximately a cube of the nominal speaker size. For example, a 300mm driver (12") needs a box of about 28 litres (1 cu ft). For reference, a cubic foot is 28.3168 litres, and one litre is contained by a cube of 100mm (10cm) to each side.

HINT: If you make all measurements in centimetres, the result will be in millilitres (cubic centimetres, or cc). This makes it very easy to convert to litres ... simply divide by 1,000. If you work in millimetres (mm), the result is less intuitive, although you can still get to litres by dividing by 1,000,000. (For those who insist on using outmoded measurement systems, I can provide a spreadsheet that uses cubits as the base linear measurement, and firkins for volume. I will part with this upon receipt of a kilderkin of Australian $1 notes or its equivalent <grin>).

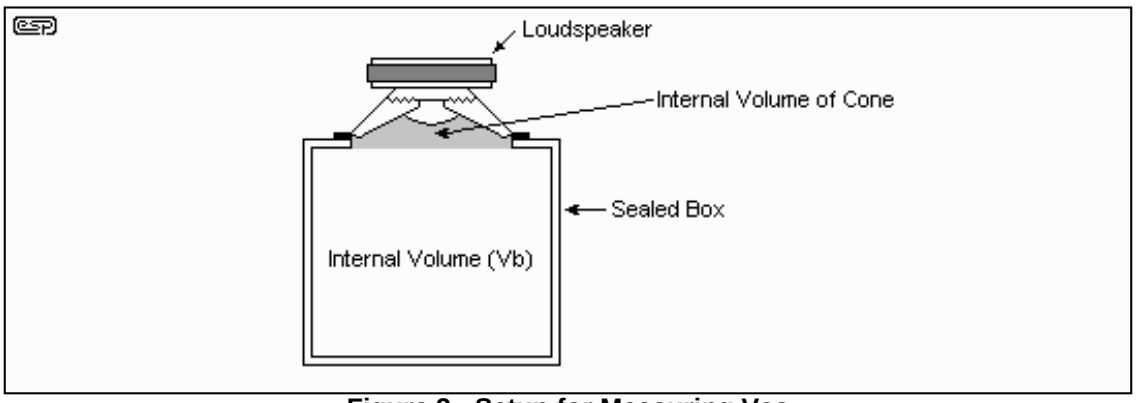


Figure 3 - Setup for Measuring Vas

Determine the total volume, including the speaker cut-out and that trapped by the cone with the speaker mounted on the outside of the box for easy access. Measure the resonant frequency in this situation, and use the free air space resonant frequency determined as shown above. Determining the volume trapped by the speaker cone is slightly tricky.

Use one of the following methods ...

1. Place the driver in a plastic bag, ensuring it is completely sealed. The bag should be loose enough so that it can be pushed easily into the cone area. Place the wrapped speaker on a flat surface, with the cone facing upwards. The cone may now safely be filled with grain (such as rice, wheat, etc), and the grain carefully poured out into a measuring jug. The resulting measurement will be a little greater than the actual volume because the cone will be depressed by the mass of the grain. The area of the speaker cutout in the cabinet must still be added. Don't omit the plastic bag, as fine dust may penetrate the speaker without it.

2. Take a series of measurements. The internal cone area is measured, then divided into sections whose volume may be calculated. For most speakers we will have two basic shapes to deal with, and although this method is not 100% accurate, it will probably give the best result in the majority of cases.

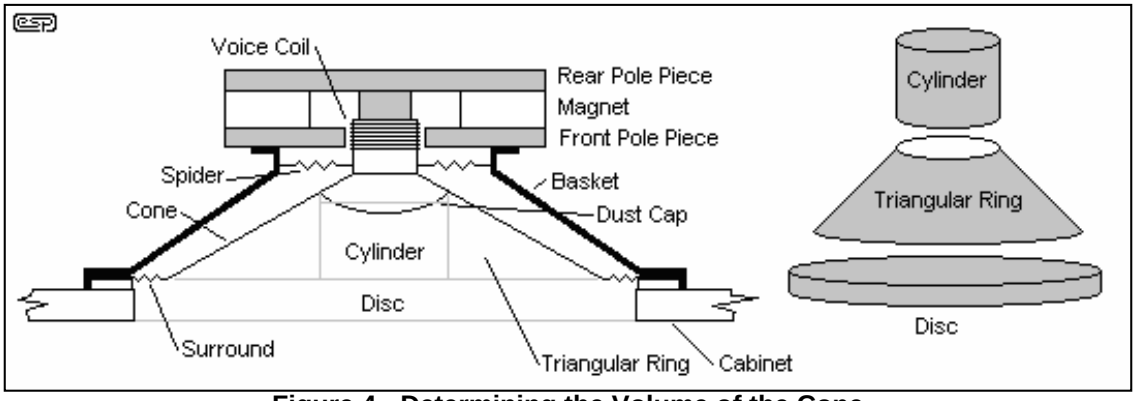


Figure 4 - Determining the Volume of the Cone

There is a flat cylinder (disc) that is formed by the outer area of the basket and the cutout in the enclosure. The triangular ring is formed by the cone itself, as shown above. Finally, the cylinder occupies the area over the dustcap. The depth of the cylinder should be an approximate average of the distance from the mounting surface and the dustcap.

The volume of cylinder and disc is given by the conventional formula ...

$V_{cyl} = \pi * r^2 * h$ (where *r* is the radius and *h* is height), and ...

$V_{disc} = \pi * r^2 * h$

The triangular ring's volume is given by ...

$V_{ring} = ((\pi * r^2 * h) / 3) - V_{cyl}$

The total speaker volume is simply the sum of the 3 volumes calculated above. Box volume is calculated as one normally would, taking great care to ensure that the measurements are accurate. The box may be braced, but must have no fibreglass or other sound deadening material inside. Make sure that the volume occupied by any bracing is accounted for in your calculations. Even a simple box will be sufficiently rigid at the frequencies of interest, so a completely acoustically dead cabinet is not required (although it won't hurt). Do **not** use any speaker cabinet filling material for this test.

$Vas = Vb((Fb / Fs)^2 - 1)$

where *Vb* is the volume trapped by the speaker and box, and *Fb* is the resonance frequency of speaker and box combined. *Fs* is the free air resonance measured previously.

An Example Calculation

A dummy test loudspeaker was used to demonstrate the process, and I have used a simulation of this speaker in the calculations shown. The equivalent circuit is shown in Figure 5. This circuit was also used to create the impedance graph shown in Figure 1.

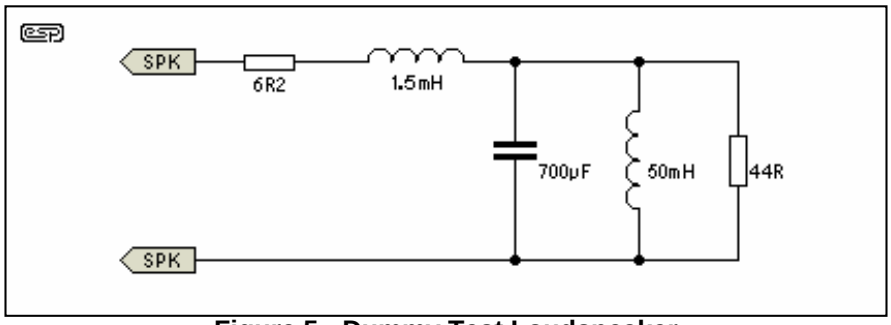


Figure 5 - Dummy Test Loudspeaker

The following screen shot shows the values for the speaker, and the only contrived (i.e. invented) value is for the resonance in the sealed box. It was necessary to invent a number here, as it is not possible to simulate it. The final figure shown is fairly typical of many such drivers, so is not too far from the truth either.

Driver name Demo 250mm

Enter the values in the shaded boxes. Values in bold are calculated

Calculate Qms, Qes and Qts

Source voltage at amplifier terminals	Vs	1000.00	mV
Source DC resistance	Rs	10.00	Ohms
Speaker voice coil DC resistance	Re	6.20	Ohms

Adjust F to at least 2 octaves above resonance. Check that varying frequency causes little or no change in voltage. Compare voltage reading with Vref - they should be within 10%

Voltage across Rs	Vref	**617.28**	mV
Source current	Is	**61.73**	Amps
Voltage across Rs at resonance	Vm	166.66	mV
Resonance frequency	Fs	27.00	Hertz
Speaker impedance at resonance (Rm)	Rm	**50.00**	Ohms
Current flow at resonance	Im	**16.67**	mA
-6dB current	Ir	**32.07**	mA
r0	r0	**3.70**	
-6dB voltage	Vr	**320.74**	mV

Carefully adjust frequency until Vr is observed across Rs, above and below resonance. Enter the frequencies.

Frequency above Fs for stated voltage	Fh	32.28	Hertz
Frequency below Fs for stated voltage	Fl	22.36	Hertz
Fs sanity check (should equal Fs)		**26.87**	Hertz
Mechanical Q	**Qms**	**5.24**	
Electrical Q	**Qes**	**0.74**	
Total Q of driver	**Qts**	**0.65**	

Calculate Vas

Box Volume	Vb	27.00	Litres
Cone Volume	Vc	**1.47**	Litres
Resonance in box	Fb	42.00	Hertz
	Vas	**40.42**	Litres

Cone Volume Calculations

Cutout diameter	200.00	mm
Baffle thickness	25.00	mm
Cone Diameter	200.00	mm
Dustcap Diameter	50.00	mm
Average cone depth	35.00	mm

Approximate cone volume **1.47** Litres

Reference Data

Use this information to create an equivalent circuit of the driver's resonant circuit.

Capacitance	**704.92**	uF
Inductance	**49.29**	mH
Parallel resistor	**43.80**	Ohms

Figure 6 - Calculation Example Using Spreadsheet

As you can see from the screen-shot, the spreadsheet will calculate everything for you, including the cone volume, *Vas* and the values shown in the schematic of the driver. Naturally, you will get figures quite different from those shown, but the principle is exactly the same.

Main Index

Articles Index

My thanks to Brian Steele for allowing me to use the simplified method and formulae he devised for Thiele/Small parameter measurement, and also Vas measurement information. Brian's original data are available **Here.**

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from the author. The formulae shown are used with the permission of Brian Steele.

Page Created and Copyright © 15 Dec 2000./ Updated 18 Aug 2003 - added guestbook suggestion (rice instead of water) + minor reformatting./ 12 Jan 2006 - changed drawings, re-measured figures to get accurate numbers, rebuilt Excel spreadsheet.

The Audio Pages

Elliott Sound Products **Satellites and Subwoofers**

Copyright © 2004 - Robert C. White and Rod Elliott (ESP)
Page Created 26 August 2004
(Updated 17 Jan 2005)

Articles Index

Main Index

Contents

- 1.0 Introduction
- 2.0 Description
  - 2.1 The Basic Principles
  - 2.2 Group I Alignments
  - 2.3 Group II Alignments
  - 2.4 Group III Alignments
  - 2.5 Examples and Charts
- 3.0 Appendix
- 4.0 References

1.0 Introduction

In recent years the satellite sub woofer type of loudspeaker system has become popular, this largely due to the advent of surround sound home theatre. For this and also for music the sat+sub system has many things to recommend it, the use of one subwoofer is made possible because the range bellow 100Hz is, as indicated by most testing, non directional, thus allowing the subwoofer to be put in the best place for bass, and allowing the various other smaller speakers to be arranged in a way most suited to imaging and

surround effect. Another benefit is that the higher cut off frequency of the satellites potentially keeps the non linear distortion caused by large low bass cone excursions at a low level.

The low cost and compact size is however bought at a cost, this is that the sound pressure generated by such a system is limited by the use of a small driver down to a typical 80Hz crossover frequency. This is exacerbated by the current fashion to denigrate the reflex enclosure, with people advised to block up those despised ports, and be transported to the sunlit uplands of sonic bliss thereby.

It is generally true that the transient response of vented enclosures is not as good as that from a (properly designed) sealed box, but we live in a world of compromise. There is little point having a system with perfect transient response if the distortion at realistic listening levels is intolerable. It is also pointless to try to convince someone who has a small listening area that they need large 3-way systems - especially if there is no physical room for such enclosures. So, in this area of compromise, there has to be a solution that will provide a reasonable SPL with minimal excursion created distortion.

While sealed enclosures certainly have their place in the scheme of things, any attempt to use them in a 'full range' 2-way system will almost certainly cause excessive distortion at even a moderate SPL, whereas a solution is available that lowers distortion, raises the SPL limit and if carefully done will give very good performance.

Interestingly, any information about the alignment discussed in this article is very scarce on the Net, to the extent that it is almost non-existent. While the author has recommended the method described for some time (mainly in forum sites), it is an alignment that the editor has also contemplated, but (until now) did not have the information to perform a theoretical design - any attempts had to be empirical. That this is time consuming in the extreme is obvious, and it is probable that the Quasi-Butterworth, 5th order (QB5) alignment will see an increase in popularity in the future.

What follows is a discussion of the issues and methods for dealing with the most vexing problems, but first some discussion about the linear excursion limit is in order.

2.0 Description

Over part of its cone excursion the loudspeaker driver is to a good approximation a linear transducer, if its B*L (magnetic field strength * conductor length) product is constant, then the force generated by the voice coil is linearly proportional to the current

in the voice coil, i.e. ...

$$F = (B * L) * I$$

Olsen, in his 1962 paper [1], identified two non linear elements that produce harmonic distortion as a function of cone displacement. Typically as the voice coil moves out of the magnetic gap, the B*L product is no longer constant but starts to fall as the excursion increases, this can be described mathematically as ...

$$F(x) = a * x + bx^2 \text{ and } F(x) = a * x + bx^3$$

Where, **F(x)** is an expression relevant for force when B*L is falling, **a** & **b** are constants relevant to a particular driver, and **x** is the displacement. The first of these expressions is relevant for the voice coil moving from its rest position asymmetrically, and the second when the motion is symmetrical about the centre position.

When excited by a sine wave the Fourier transform of these yields ...

$$x = A * \cos \omega t + B * \cos 2\omega t \text{ and } x = A * \cos \omega t + B * \cos 3\omega t$$

Where ω represents the angular frequency (equal to $2\pi f$), **x** is the amplitude, ω**t** is the exciting frequency and **A** ,**B** are proportional constants, dependant upon signal level.

The second terms, **B * cos 2**ω**t** and **B * cos 3**ω**t** represent the second and third harmonics respectively.

For small displacements when BL is constant, the second term is vanishingly small, and the distortion produced comes from non linearities in the suspension and surround, and from air trapped in the magnetic gap. The 2nd harmonic component is largely out of our control since its level is largely due to the design and construction of the driver.

The third harmonic is in our control however because it is directly related to the amount of signal we put in. This where the linear excursion comes in....

The peak linear excursion limit is derived, unless it is specified by the manufacturer, by using the rule of thumb that it is the voice coil length, minus the magnetic gap height over two, this is true for the usual overhung voice coil. Gander, [2] measured a wide selection of drivers and came up with the findings that calculating the linear excursion limit in this way is a very reliable indicator of the driver producing 3% third harmonic

distortion, when driven to this limit. In what follows this is taken as the maximum peak distortion that is acceptable.

As an example of what this means in practical terms, let us take a typical good quality satellite mid/woofer, the Vifa P17WJ, and put it in a sealed box.

Nominal impedance [ohm]	8	Air gap height [mm]	6
Voice coil resistance [ohm]	5.8	Voice coil inductance [mH]	0.55
Nominal power [W]	40	Eff. diaphragm Area [cm²]	136
Short term max power [W]	350	Moving mass [g]	14
Long term max power [W]	150	Magnet weight [g]	415
Operating power [W]	6.3	Force factor [Bl]	6.5
Sensitivity [dB]	88	VAS [litres]	34.7
Frequency range [Hz]	37-5000	Qms	1.55
Free air resonance [Hz]	37	Qes	0.45
Voice coil diameter [mm]	32	Qts	0.35
Voice coil height [mm]	14		

P17WJ-00-08 Parameters

In a box of 11.5 litres it will have a Q of 0.707 and an fs around 78Hz by the manufacturers specifications. it has a linear cone excursion of around 4mm peak. Using Small's expression for the acoustic power we can expect in an average room [3]

$$\mathbf{Pa = kp * Vd^2 * f3^4}$$

Where **kp** is the power output constant, it has a value of around 0.8 for a sealed box

Vd is the volume displacement (cone area * peak excursion) and **f3** is the system's lower -3dB frequency. This gives...

$$\mathbf{kp = 0.8 * (0.0136 * 0.004)^2 * 78^4} = 0.088 \text{ Watts}$$

Converting this to decibels we have...

107.5 - 10 * log10 (1 / 0.088)= 96.9dB

This is the peak sound level we can reproduce before exceeding the excursion limit. Taking into account that "normal" program is said to have a 20dB dynamic range then our average level is around 77dB. This falls far short of the standard 90dB with 20dB headroom requirement for a hi fi system. Although it can play a lot louder, this will not be without producing large amounts of harsh sounding non linear distortion - the amount of distortion increases very rapidly beyond our excursion limit.

For instance the 110dB peak output causes a 4mm excursion at 189Hz. At the 78Hz region, the driver needs an excursion of 16mm.

From the above it seems that the position is hopeless, crossing small satellites of this sort over to a single subwoofer causes large amounts of distortion.

There are of course several alternatives, these include three way satellites, stereo subwoofers crossing over at a higher frequency and so on, but all of these lack the basic simplicity and low cost of the two way satellite plus single subwoofer system.

Luckily there is a solution, this is in the form of the QB5 reflex alignment.

I know that immediately I say that word 'reflex', many people will throw their hands up in horror, make the sign of the cross, or other gestures that are best not considered. Swimming against the tide of fashionable anti-reflex sentiment is difficult, but I nevertheless present the following for your consideration ...

2.1 The Basic Principles

Two people named Kreutz and Panzer derived the QB5 reflex alignments, and published their work in 1993, [4] the class II group of these have the characteristic that they optimise power handling by means of reducing the cone excursion needed for a given acoustic power output, and these are the most useful for our purpose.

In essence, what we will do is create a box that is specifically designed to have a peaked response at the low end. This means that the box tuning (rather than the loudspeaker cone) is producing the larger proportion of the LF energy, so less power is needed to achieve a given SPL. In order to prevent the system sounding like a boom-

box, a 2nd order filter (whose Q is also obtained from the table) is placed before the power amp, tuned to the frequency obtained from the charts below. The system as a whole will now have a vastly higher power output at the low frequency end, without the problem of excursion limiting and subsequent distortion. On the down side (and like all things, there are compromises), the box will be larger than a conventionally tuned system, but this would be the case regardless if a larger woofer were to be used to try to maximise the excursion limited SPL.

The details of these derivations are as they say 'beyond the scope of this article', and for those interested the paper cited in the references is available from the AES. Suffice to say however, that the benefits of filter assisted reflex boxes have been known and exploited for many years, but the derivation of these particular alignments was not achieved formally until Panzer came up with a solution to some very difficult mathematics. Luckily for us this gave us three alignment tables, these are reproduced below.

[Reproduced from:- "***Derivation of the Quasi-Butterworth 5 Alignments***" , by J.J.M. Kreutz & J. Panzer, Journal of the Audio engineering Society, Vol. 42, No. 5, May 1994]

Qt	**Vas/Vb**	**F3/Fs**	**kp**	**Fb/Fs**	**1/Qa**	**T_{60}**	**Fa/Fs**
Driver Qts	Box Volume	-3dB Freq.	Power Ratio	Box Tuning	Filter Damping	Settling Time	Filter Freq.

The three tables stem from the value of a parameter used in the solutions of the equation set. In practical terms, Group I is the optimum power handling vs. bass extension for drivers in the Qt range specified. The Group II alignments are for maximum power handling, again within the Qt range covered. Group III are included for completeness, as they represent the edge of usefulness and can be largely ignored.

2.2 Group I Alignments

Qt	**Vas/Vb**	**F3/Fs**	**kp**	**Fb/Fs**	**1/Qa**	**T_{60}**	**Fa/Fs**
0.324	1.989	0.935	3.911	0.995	0.467	4.778	0.944
0.318	2.103	1.000	6.436	1.00	0.518	4.315	1.000
0.311	2.239	1.059	6.782	1.008	0.549	4.074	1.055
0.303	2.396	1.114	7.237	1.018	0.569	3.933	1.110
0.295	2.574	1.169	7.811	1.030	0.582	3.844	1.164
0.287	2.773	1.223	8.512	1.046	0.592	3.785	1.219

0.279	2.991	1.277	9.349	1.063	0.598	3.746	1.274
0.271	3.229	1.332	10.331	1.083	0.603	3.717	1.329
0.263	3.486	1.387	11.469	1.104	0.606	3.697	1.385
0.255	3.762	1.443	12.776	1.127	0.608	3.682	1.441
0.247	4.056	1.499	14.265	1.152	0.610	3.671	1.498
0.240	4.368	1.556	15.951	1.178	0.612	3.663	1.555
0.233	4.698	1.614	17.849	1.205	0.613	3.656	1.613
0.226	5.046	1.672	19.977	1.234	0.614	3.651	1.671
0.219	5.412	1.731	22.352	1.263	0.614	3.647	1.730
0.213	5.795	1.790	24.992	1.293	0.615	3.643	1.789
0.207	6.196	1.849	27.919	1.324	0.616	3.641	1.848
0.201	6.614	1.909	31.153	1.356	0.616	3.639	1.908
0.196	7.050	1.969	34.716	1.389	0.616	3.637	1.968
0.190	7.503	2.030	38.631	1.422	0.616	3.635	2.029
0.185	7.973	2.090	42.921	1.456	0.617	3.634	2.090

2.3 Group II Alignments

Qt	Vas/Vb	F3/Fs	kp	Fb/Fs	1/Qa	T_{60}	Fa/Fs
0.445	0.550	1.000	6.436	1.000	1.414	4.316	1.000
0.425	0.751	1.152	10.152	1.115	1.527	3.919	1.107
0.415	0.856	1.231	12.587	1.174	1.558	3.839	1.187
0.405	0.960	1.303	15.176	1.228	1.576	3.790	1.264
0.394	1.072	1.373	18.090	1.281	1.588	3.756	1.339
0.384	1.193	1.444	21.426	1.335	1.595	3.729	1.415
0.373	1.325	1.515	25.264	1.389	1.601	3.709	1.491
0.362	1.467	1.588	29.678	1.444	1.604	3.693	1.567
0.352	1.620	1.661	34.742	1.500	1.607	3.681	1.643
0.342	1.784	1.736	40.528	1.557	1.609	3.671	1.720
0.331	1.959	1.811	47.115	1.615	1.611	3.663	1.797
0.322	2.144	1.887	54.580	1.674	1.612	3.657	1.875
0.312	2.339	1.963	63.009	1.733	1.613	3.652	1.953
0.303	2.545	2.040	72.486	1.793	1.614	3.648	2.031

0.294	2.761	2.118	83.102	1.854	1.615	3.645	2.110
0.286	2.988	2.196	94.952	1.916	1.615	3.642	2.188
0.278	3.224	2.274	108.134	1.978	1.616	3.640	2.268
0.270	3.470	2.353	122.749	2.04	1.616	3.638	2.347
0.262	3.726	2.432	138.903	2.103	1.616	3.636	2.426
0.255	3.993	2.511	156.706	2.166	1.616	3.635	2.506
0.249	4.269	2.590	176.216	2.229	1.617	3.634	2.586

2.4 Group III Alignments

Qt	Vas/Vb	F3/Fs	kp	Fb/Fs	1/Qa	T_{60}	Fa/Fs
0.514	0.533	1.029	4.976	1.018	2.020	3.833	0.794
0.517	0.527	1.030	5.136	1.019	2.008	3.852	0.813
0.520	0.520	1.031	5.295	1.02	1.996	3.871	0.832
0.523	0.513	1.032	5.451	1.02	1.987	3.892	0.850
0.526	0.505	1.032	5.603	1.020	1.978	3.914	0.867
0.530	0.496	1.032	5.749	1.020	1.970	3.939	0.884
0.534	0.487	1.031	5.887	1.020	1.963	3.966	0.901
0.538	0.476	1.030	6.016	1.019	1.957	3.997	0.917
0.543	0.465	1.028	6.133	1.017	1.951	4.031	0.932
0.549	0.452	1.024	6.234	1.015	1.947	4.070	0.947
0.555	0.438	1.020	6.319	1.013	1.942	4.116	0.961
0.562	0.422	1.015	6.383	1.009	1.938	4.171	0.975
0.569	0.405	1.008	6.423	1.005	1.935	4.236	0.988
0.577	0.386	1.000	6.434	1.000	1.932	4.317	1.000
0.587	0.364	0.990	6.251	0.994	1.929	4.417	1.011
0.598	0.339	0.977	6.035	0.986	1.927	4.545	1.022
0.610	0.312	0.962	5.780	0.977	1.924	4.713	1.031
0.625	0.280	0.943	5.482	0.966	1.922	4.939	1.039
0.643	0.244	0.921	5.132	0.952	1.921	5.254	1.045
0.664	0.202	0.894	4.719	0.936	1.919	5.716	1.049
0.691	0.150	0.861	4.224	0.915	1.918	6.449	1.050

* T_{60} is the time needed for a pulse input to settle by 60dB

2.5 Examples and Charts

Plugging our Vifa P17WJ driver into the relevant class **II** alignment, i.e. that for a Qt of 0.35, yields (after calculation) the following ...

Box volume	21.4 litres
f3	61.5Hz
Box tuning frequency	55.5 Hz
Fa (Aux. filter 12dB/Octave high pass)	60.8Hz
Qa	0.622
kp	34.74

Note that in the above the power handling constant **kp** has gone up from 0.8 to an impressive 34.74, the practical effect of this is shown below...

$Pa = 34.74 * (0.0136 * 0.004)^2 * 61.5^4 = 1.47$ Watts
dB SPL = 10 * log10(1.47) + 107.5 = 109.17dB

Our peak output before we reach our excursion limit has gone up by 12.3dB even with a lower f3, if we use the high pass filter from our subwoofer amplifier this should be set at the 61Hz frequency as a minimum, setting it higher results in an even greater increase in power handling.

Two of these speakers in a room can easily meet the 110dB SPL peak level that is the standard for domestic hi fi systems without exceeding the 4mm peak cone excursion limit, this results in a significant reduction in non linear distortion. One must be careful not to exceed the long term power handling of the drivers - since there will be little loudspeaker distortion at even high levels, there will be no warning that the average electrical power is too high, so some common sense is required to prevent voicecoil damage.

Another fully worked example (similar speaker, but different parameters) will help understanding ...

Nominal impedance [ohm]	6	Air gap height [mm]	-
Voice coil resistance [ohm]	4.0	Voice coil inductance [mH]	0.4
Nominal power [W]	50	Eff. diaphragm Area [cm²]	140
Short term max power [W]	-	Moving mass [g]	-
Long term max power [W]	-	Magnet weight [g]	-
Operating power [W]	-	Force factor [Bl]	4.1
Sensitivity [dB]	87	VAS [litres]	49.4
Frequency range [Hz]	-	Qms	1.34
Free air resonance [Hz]	35	Qes	0.58
Voice coil diameter [mm]	-	Qts	0.40
Voice coil height [mm]	-		

P17WG-00-08 Parameters

The above data were obtained from WinISD, and some parameters are missing, or may be in error. However, this was used for the simulations that follow, and the general idea is the same for any speaker.

Driver	Vifa P17WG-00-08		
Qt	0.40		
fs	35 Hz		
Vas / Vb	0.960	Vb = 49.4 / 0.96	Vb = 51.45 litres
f3 / fs	1.303	f3 = 35 * 1.303	f3 = 45.6 Hz
fb / fs	1.228	fb = 35 * 1.228	fb = 42.98 Hz
1 / Qa	1.576	Qa = 1 / 1.576	Qa = 0.635
fa / fs	1.264	fa = 35 * 1.264	fa = 44.24 Hz

The box in this case is a little over the normal volume for a critically aligned vented enclosure, and will have a response as shown below. The normal reaction to this would be *"Yuck!"* because of the large peak of 2.45dB occurring just before rolloff. This is where the electrical filter comes into play, and this filter should have a frequency of

44.24 Hz (**fa**) and a Q of 0.635 (**Qa**) to match this alignment.

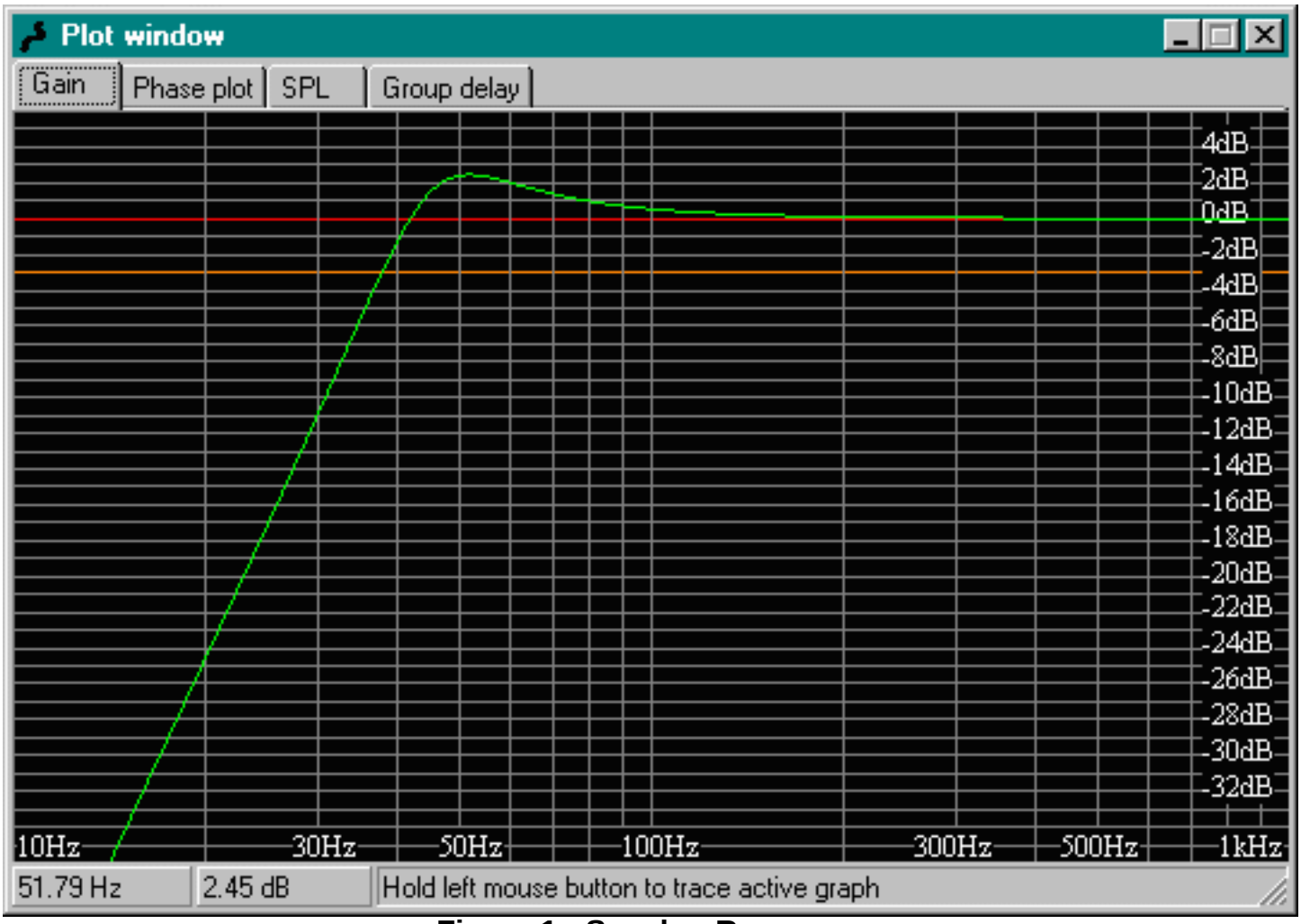


Figure 1 - Speaker Response

The above graph is the output from WinISD, with the box size and tuning frequencies adjusted according to the calculations above. The actual f3 point of the speaker is 37.8Hz, but this will be raised by the electrical filter to the figure of 45.6Hz as obtained from the table. The simplest filter to use for this is the equal component value Sallen-Key, and the schematic for this is shown below. As noted above, the exact Q will not make a great deal of difference - it may look bad, but will be completely overshadowed by room nodes. For the time being, we'll look at the system response with the correct Q, especially since it is quite easy to do.

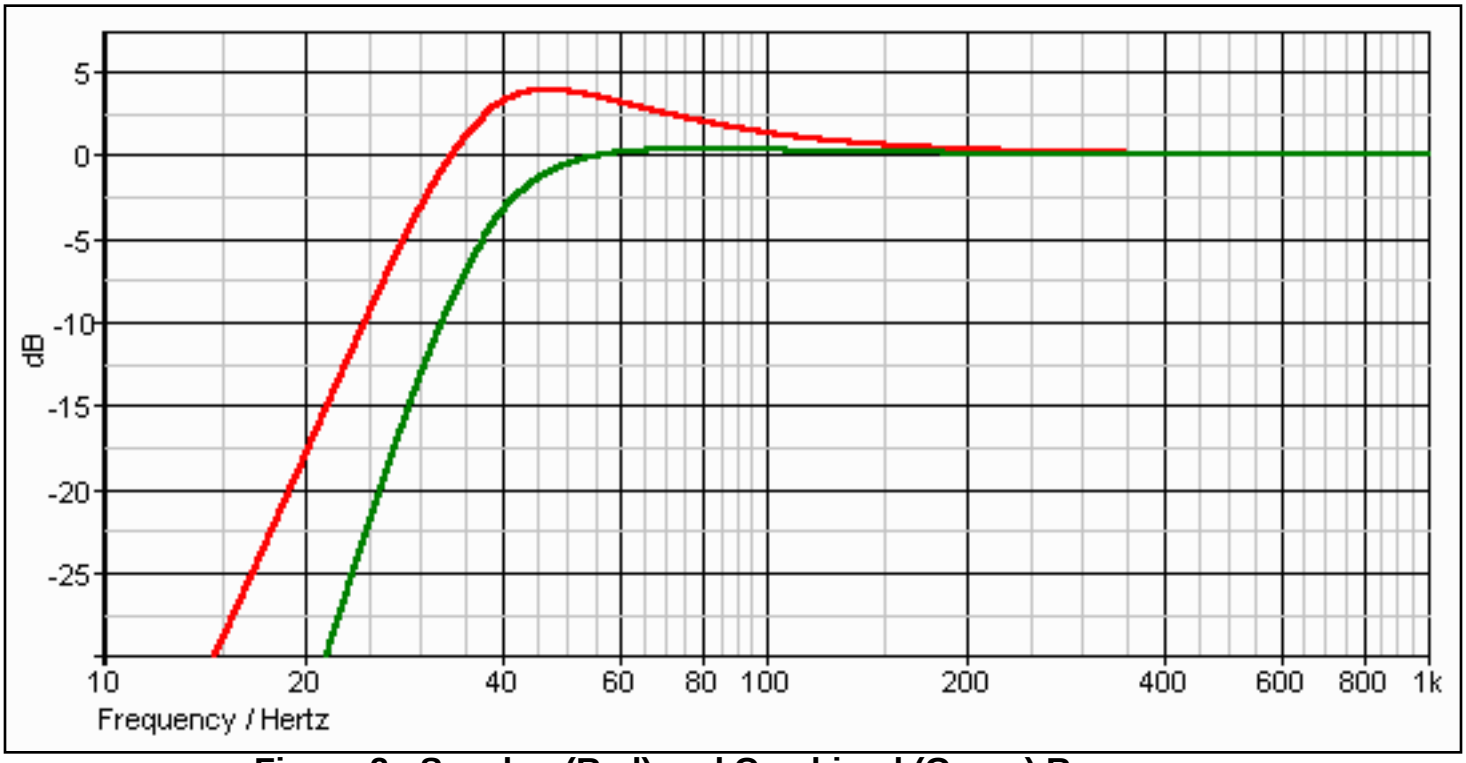


Figure 2 - Speaker (Red) and Combined (Green) Response

I used the Simetrix simulator to produce the above, and although it is not a perfect match for the loudspeaker response in Fig. 1, the result is very close. The red trace is the uncorrected simulated loudspeaker response, while the green trace shows the result after correction with the auxiliary filter. The simplest way to make the auxiliary filter is to use the 'equal component value Sallen-Key' filter topology. This allows independent selection of frequency and Q, but the gain of the circuit will change. The gain obtained from a low Q filter is small, and is unlikely to cause any problems in typical systems.

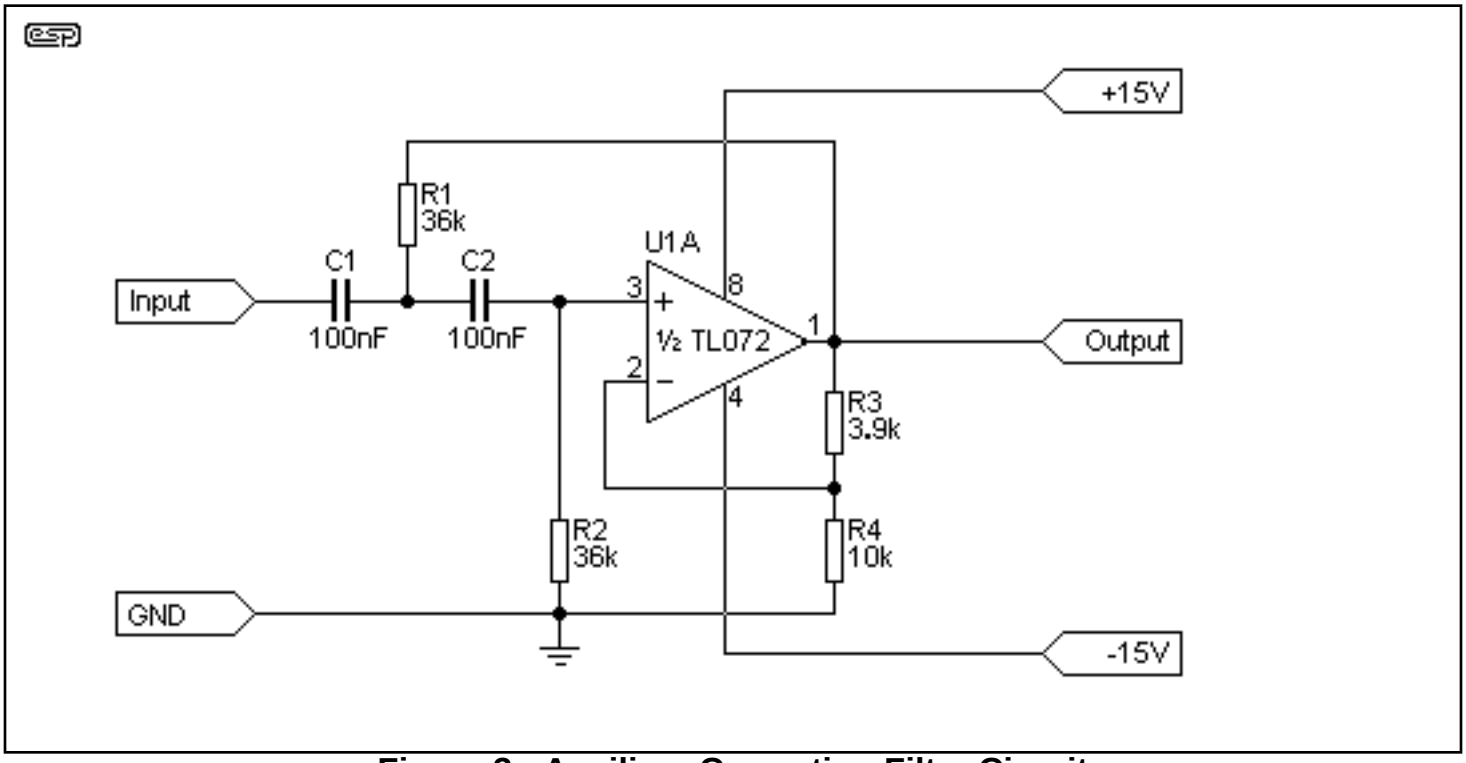


Figure 3 - Auxiliary Correction Filter Circuit

Referring to the schematic above, you can see that R1=R2 and C1=C2. The second half of the opamp would be used for the other channel. Note that this circuit must be fed from a low impedance source (typically less than 1k). The values needed are given by ...

f = 1 / (2 * π * R * C)

This is not affected by the gain setting (which alters the Q), so it is a flexible solution with no bad habits. For the filter response we calculated above, first we shall select a convenient value for C - say 100nF. The value for R is obtained by ...

R = 1 / (2 * π * C * f)= 1 / 2 * π* 100E-9 * 44.24 = 35.98kΩ

It has to be said that the calculated resistance value is not going to be easy to get, but a 50k dual-gang pot will allow you to set the frequency wherever you like. Alternatively, you could use a 33k and a 2.7k in series, or even just 33k with a still insignificant error. Now that the frequency is decided, we need to calculate the Q. This is not even a little bit painful, since the tables actually give the figure for damping (1/Qa), which simplifies things even more.

Q = 1 / (3 - K)or **R3 = (2 - d) * R4**

where **K** is the circuit gain (determined by the ratio of R3:R4), and **d** = 1/Q. For minimum DC offset, the parallel combination of R3 and R4 should equal R1, but this is of no consequence if the power amplifier has a capacitor input as offset will be very low. A value of 10k will be quite acceptable for our purposes. From the table, we determined that d (1/Qa) is 1.576, so substituting ...

R3 = (2 - 1.576) * 10k= 0.424 * 10k = 4.24k

Alas, we have another non-standard value, but a 3.9k resistor in series with 330Ω gives 4.23k which will be quite acceptable. In fact, you could just use the 3.9k resistor, since the error is very small indeed, and will be completely un-noticed in practice. The values calculated and shown on the schematic were used in the simulation, and as is obvious, the error is small.

Using the same formula as before, we can calculate the acoustic power increase over a sealed box. 5mm linear excursion is assumed, from the specifications for the P17WG ...

$Pa = kp * Vd^2 * f3^4$

$Pa = 0.8 * (0.0136 * 0.005)^2 * 62^4$= 34mW

and for the QB5 alignment we calculated ...

$Pa = 15.176 * (0.0136 * 0.005)^2 * 40^4$= 179mW

Not quite as impressive as the previous example, but note that we have a -3dB frequency of 40Hz instead of 62Hz, and still have a 7.2dB increase in effective acoustic power output. Converting to dB SPL, we get ...

dB SPL = 10 * log10(0.179) + 107.5= 100.03 dB

The is a substantial improvement over the 92.8dB SPL obtainable from a sealed box, plus we have over ½ octave more low frequency extension. As before, if we were to raise the crossover frequency, there will be an equivalent increase in level before the excursion limit is reached. As it stands, the arrangement just calculated would outperform the majority of systems using conventional vented boxes and the same driver. The enclosure is only marginally larger than a 'critically aligned' system, but has an f3 of 40Hz rather than 34Hz. Another plus is that there is a filter that will help reduce subsonic excursions that produce no useful output, but can introduce considerable

distortion because the cone is unloaded below the cutoff frequency.

As previously mentioned, the 2nd order high pass output from your sub amplifier can be used as the filter. While it is true that this is usually a Butterworth type, with a Q of around 0.7, overall this does not make much difference at this low crossover frequency, as room modes and cancellation/reinforcement from walls predominate over any exact crossover characteristic. Be aware that many 'plate' sub amps have a fixed high-pass filter, and only the low-pass section is variable. Such units will almost certainly need the filter circuit described.

If the high-pass section of the crossover is variable this is also useful, as this allows an optimum setting for a smooth response to be found more easily. It also increases power handling even more. Referring to the original driver (the P17WJ), if kp remains the same and f3 is increased to 80Hz, the excursion limited peak output goes up to 113.7dB - this represents an increase of 16.8dB in peak output over a sealed box before excursion limit is reached. One can use a peak input of 263 Watts in the case of the P17WJ, which has an unspecified 'short term' maximum power input rating of 350 Watts.

The method outlined here allows the maximum performance to be extracted from the single subwoofer plus two way satellite configuration, and in terms of 'bang for your buck' is hard to beat.

It does however use the now demonised reflex enclosure, but it must be emphasised that any negative effects from a port are more than compensated for by the very significant reduction in non linear distortion this technique confers.

In addition, it could be argued that the electrical filter brings the box back into alignment, and not only relieves the loudspeaker of undue stress and distortion, but by removing the low frequency component from the amplifier, the power actually needed could be reduced by perhaps 3dB or more. While this may not sound like much, it is the difference between a 100W amp and a 50W amp, so the savings all round could be very significant.

Appendix

Curve fit solutions for the first two alignments are presented here. These may be used to calculate the values, rather than using the lookup tables.

Class I Alignments

All of these have a correlation coefficient of at least 0.98, except for 1/Qa, with 0.91

Vas / Vb = 0.126 * (Qt^-2.4731)
F3 / Fs = -1.2576 - 1.9761 * ln(Qt)
kp = 221.419 - 1.401.7 * Qt + 2283.34*Qt^2
1 / Qa = 0.4437 + 10.2652 * Qt^2 - 29.982 * Qt^3
Fb / Fs = 3.252 - 13.448 * Qt + 20.0398 * Qt^2
Fa / Fs = 4.4786 - 15.94 * Qt + 15.7596 * Qt^2

Class II Alignments

1/Qa has a correlation coefficient of 0.871

Vas / Vb = 16.883 - 69.941 * Qt + 75.2755 * Qt^2
F3 / Fs = 5.6721 - 15.168 * Qt + 10.6884 * Qt^2
kp = 9357.22 * e^(-15.987 * Qt)
1 / Qa = 1.3873 + 7.3126 * Qt^2 - 15.584 * Qt^3
Fb / Fs = 4.7413 - 12.556 * Qt + 9.4877 * Qt^2
Fa/Fs = -1.173 - 2.694 * ln(Qt)

References

[1]	H.F. Olsen	*"Analysis of the effects of non linear elements upon the performance of back enclosed direct radiator loudspeaker mechanisms"* AES journal, Vol 10, pp.156-162, (Apr 1962)
[2]	M.R. Gander	*"Moving coil loudspeaker topology as an indicator of linear excursion capability"* AES journal, Vol 29, No ½, (Jan./Feb 1981).
[3]	R.H. Small	*"Vented box loudspeaker systems, Part II: Large signal analysis"* AES journal, Vol 21, No 6 (Jul/Aug 1973).
[4]	J.J.M. Kreutz & J. Panzer	*"Derivation of the quasi Butterworth 5 alignments"* AES journal, Vol. 42, No 5 (May 1994)

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of the reference authors quoted (in particular the AES), Robert C. White and Rod Elliott, and is Copyright © 2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The authors (Robert C White and Rod Elliott) grant the reader the right to use this information for personal use only, and further allow that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Robert C. White and Rod Elliott.

Page created and copyright © 26 August 2004./ Updated 17 Jan 05 - corrected errors in formulae.

The Audio Pages

Elliott Sound Products **The Subwoofer Conundrum**

Copyright © 2004 - Rod Elliott (ESP)
Page Created 01 Mar 2004

Articles Index

Main Index

Contents

- 1 - Introduction
- 2 - Timing is Everything
- 3 - Optimum Crossover Frequency
- 4 - Cone Area Vs. Displacement
- 5 - Soundwave Vs. Pressure Mode
- 6 - Subwoofer Positioning
- 7 - Conclusion

---

1 - Introduction

Conundrum *(noun): A paradoxical, insoluble, or difficult problem. A dilemma.* Well, that just about sums it up for subwoofers, doesn't it? Of all the add-ons that can be applied to a system, the sub often gets the short straw - they are routinely shoved into a corner, or behind the couch, or even made into coffee tables. The problem is that none of these treatments will work properly unless you have some idea of the physics behind it. The vast number of compromises needed only makes things harder.

The use of subwoofers in any audio system that is intended mainly for music is often the cause of much soul-searching, not to mention web searching, trying to find out what will

work and what won't. Home theatre is easier, since the primary requirement is for satisfyingly deep rumbles at the appropriate times, and to give the sense of "depth" to the soundtrack, and accuracy is not required to the same degree. Yes, I know that many consider film soundtracks to warrant accuracy too, but most listeners/ viewers don't care that much.

In the case of a home theatre system, there is usually no concept of "speed", since the sounds do not usually have the tight connection found in music, and a few milliseconds here or there is of little consequence. Not so if the system will be used for music as well, since now there are distinct notes that should be reproduced with as little time delay as possible. If this is not the case, the bass becomes blurred and indistinct, and has no proper connection to the music overlaying the lowest notes.

People often describe a bass reproducer as "slow" or "fast", but in fact neither term is applicable. A 40Hz bass fundamental cannot be fast or slow - it is simply a 40Hz (transient) tone, and our hearing is depressingly bad at even hearing such frequencies until they have been present for several cycles.

So, do the terms actually mean anything, or is this more fluff and bluster from the marketing departments of the larger manufacturers?

---

2 - Timing is Everything

When used in any advertising, you have basically fluff and bluster, and little else. Bass *cannot* be slow or fast, but it *can* be reproduced and heard when you are supposed to hear it, or it can be out of time with the rest of the music. The timing in *itself* is usually not the problem, but there are many things afoot that make a huge difference. These (not surprisingly) are the subject of this article.

Consider the setup shown in Figure 1. The sub is off to one side, with the listener placed directly in front of the main speakers. With spacings as shown, the distance from the main speakers to the listener is 2.23 metres to each box. Subwoofer to listener is 3.6 metres, a path length difference of 1.37 metres. This is not an uncommon setup, and as is often the case, the WAF (Wife Acceptance Factor) must be considered in the final placement. It is also non-sensible to have a subwoofer in the middle of the floor for people to trip over

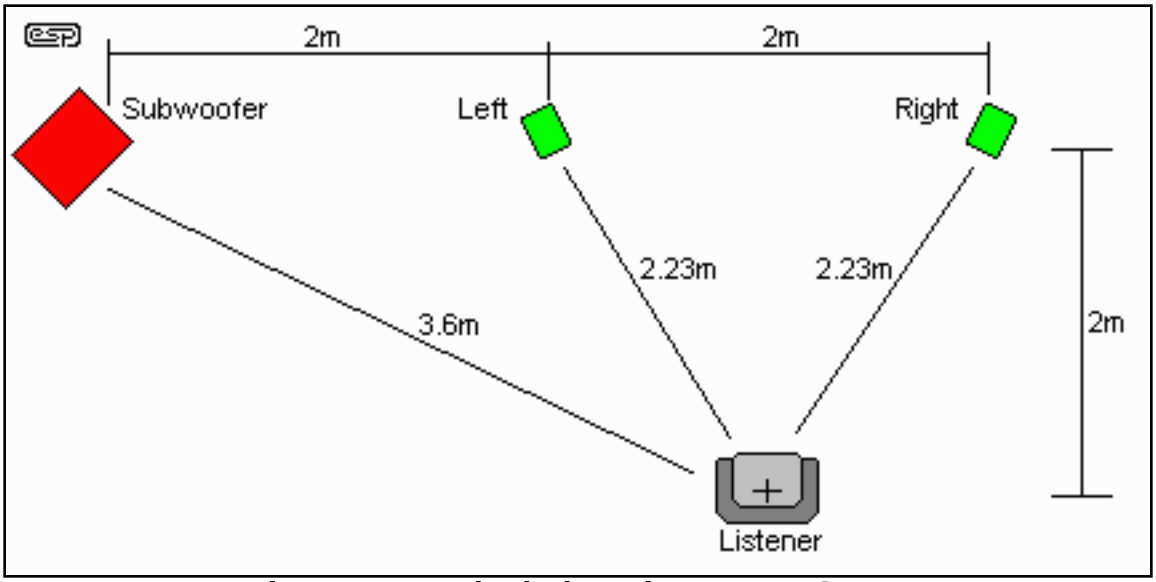


Figure 1 - Typical Listening Room Setup

At higher frequencies, a path difference of 1.4 metres (near enough) would be completely unacceptable, but it can also be unacceptable for bass, too. It depends on the crossover frequency. The following table shows the important frequencies and wavelengths at various frequencies.

Frequency	1/4 Wave (-3dB)	1/2 Wave (cancels!)	3/4 Wave (-3dB)	Wavelength
120Hz	0.718	1.44	2.16	2.875
100Hz	0.8625	1.725	2.875	3.45
80Hz	1.078	2.156	3.23	4.3125
60Hz	1.4375	2.875	4.3125	5.75
40Hz	2.156	4.31	6.47	8.625

Table 1 - Frequency Vs. Wavelength (in metres)

For the purposes of this exercise, I have no choice but to assume a perfect (anechoic) listening space, devoid of reflections and standing waves. Unfortunately, these always exist, and rarely improve matters, but may mask some of the effects that will be described. No matter, since the basic principles stand and should be considered - the effects of standing waves (in particular) are a separate topic altogether.

Over the entire crossover range, if the difference between the listener and main speakers vs. the distance from listener to sub is 1/2 wavelength, then the sub's output will partially cancel the output from the main speakers. A phase switch is often included to fix this, and this is the most obvious of all problems. In the case shown in Figure 1, 120Hz is a crossover frequency to be avoided at all costs, since there is almost a

complete cancellation due to relative path lengths. Remember that the crossover range should be considered to be a full octave with typical crossovers, and much higher slopes are needed to minimise this. 12dB/Octave is an absolute minimum requirement, and 24dB/Octave should be considered the standard.

But, why can't we just operate the phase switch and make it right again? Well, we can, but then there are other problems. What of the path length between the sub and the speakers themselves, and what of the frequencies either side of crossover? Since these are all different, there will be interactions at the main speakers, at the sub, and at various places throughout the room. Note that in an ideal case, 1/4 or 3/4 wavelengths will cause a dip of 3dB, since they have a 90° and 270° degree phase angle respectively. If at all possible, the relative path lengths should be no more than 1/4 wavelength for at least 1/2 octave above crossover - this effectively rules out the layout shown in Figure 1, but in reality it *will* work provided the crossover frequency is no higher than 60Hz or so.

Every multiple of 1/2 wavelength throughout the room will cause a dip in response, which may be partial or total (no signal at all), and this will be effective to a greater or lesser degree for the whole range of frequencies across the crossover region. With a setup such as this, the crossover frequency should be no higher than 60Hz, or the subwoofer must be moved to a more central location, closer to the main speakers.

At any crossover frequency higher than around 60Hz, standing waves, which are influenced by the size of the room, the wall, floor and ceiling materials, (large) soft furnishings, etc., will confuse an already confused bass image, and may well give rise to the impression that it is 'slow'. Bass signals around the crossover frequency will be of variable amplitude, and may change from one note to the next. Despite claims to the contrary in some areas, you will also probably find that you *can* tell where the bass is coming from as well - there will be enough higher frequency energy from a 120Hz crossover to enable you to localise the bass as being well off-centre.

---

3 - Optimum Crossover Frequency

Generally, a sub should be crossed over at the lowest practicable frequency. Based on the information in Table 1, you may need to re-think the idea of just shoving it behind a lounge chair, based on the crossover frequency and the relative distances from main speakers (*both* of them!) to the listener and the sub itself.

An ideal solution would be to use two subs, each located near to the speakers.

Unfortunately, cost, size and even the ability to place them optimally often (usually) rules out this approach. The next best thing is to try to arrange it so that the path length between the sound sources and the listener are within at least 1/4 wavelength at the crossover frequency. Unfortunately, this will often mean a location that is already occupied by other furnishings, the TV or even the cat

The limitations on crossover frequency are alleviated somewhat by positioning, but the optimum position for the sub to minimise phase and timing issues may well turn out to the worst possible location for standing waves and good bass at the listening position. As shown in Figure 2, by re-positioning the sub so that it is between the two main speakers will allow a crossover frequency that is much higher than the previous example, but the room shape and size may make this position unacceptable for standing waves.

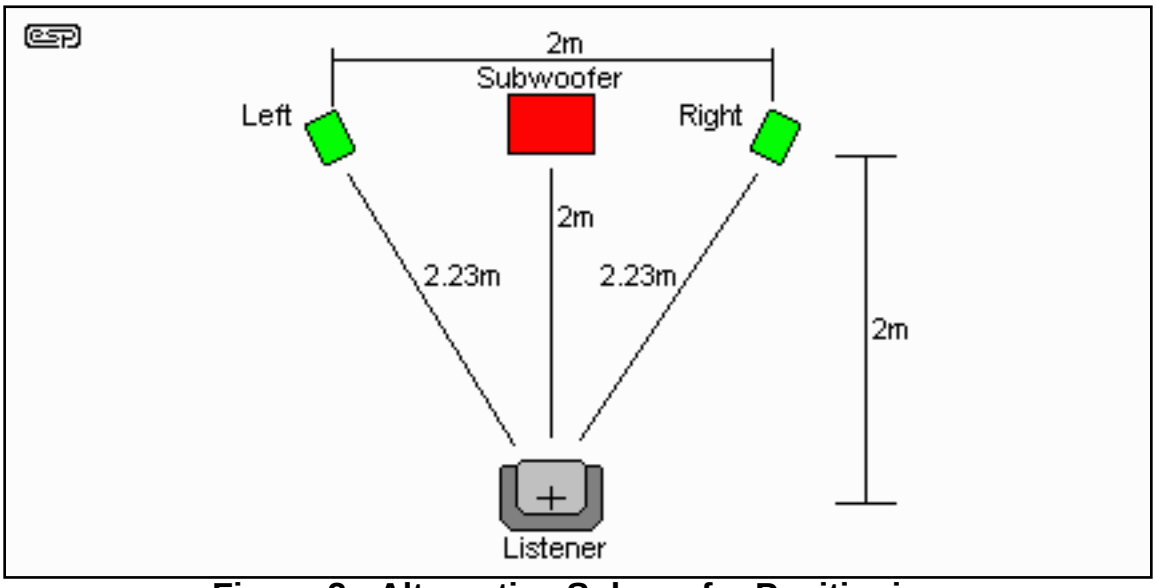


Figure 2 - Alternative Subwoofer Positioning

If this works well for bass at the listening positioning, then based on timing errors, the maximum phase shift is negligible, and well under 90° (path length difference is reduced to 230mm) even with a 100Hz crossover frequency. If this cannot be achieved, then the alternatives are few.

Adjustable phase controls usually do more harm than good, since the final phase relationships are something of a lottery. You might win, but the odds are stacked against you. Digital delay will work too, but it cannot correct off-centre positioning where the distances from each box to the listener and between each other are all different. One of the few tools at your disposal (for a sensible price at least) is an equaliser (for example the Sub-Woofer Equaliser shown in the projects section. It is not a panacea, but can help if things aren't quite right, and can correct many problems that cannot be fixed any

other way. The other thing (of course) is to optimise the crossover frequency, getting it as low as you can - the limiting factor is often the main speakers (especially if smaller bookshelf types), but most speakers will get down to 70Hz, and many a lot deeper.

There is always an inevitable tradeoff between the minimum frequency of the main speakers and the likelihood of intermodulation distortion caused by excessive cone excursions - this is relieved somewhat by using a 3-way main system, but this is not always practical. Additional phase anomalies are introduced as any loudspeaker approaches resonance, again, there is little you can do about it, but careful placement is a better solution than variable phase controls in most cases.

---

4 - Cone Area Vs. Displacement

A common misconception is that cone area can be reduced if displacement (Xmax) is increased. While it is true up to a point, it is only possible to increase Xmax up to a few millimetres of linear travel without a severe sacrifice in loudspeaker driver efficiency. Consider the two options shown in Figure 3 - the voicecoil may be overhung or underhung, but one either loses efficiency by having only a part of the coil in the gap, or by having much of the magnetic field bypassing the voicecoil altogether. Low efficiency (less than 90dB/m/W) requires more amplifier power and more heat. Remember that each 3dB fall in efficiency requires double the amplifier power for the same SPL.

Also, many manufacturers rate efficiency at 2.83V instead of 1W. At 8 ohms, the figures are the same, but for a 4 ohm driver, the efficiency is artificially inflated by 3dB, and you will need twice as much power as you may have thought.

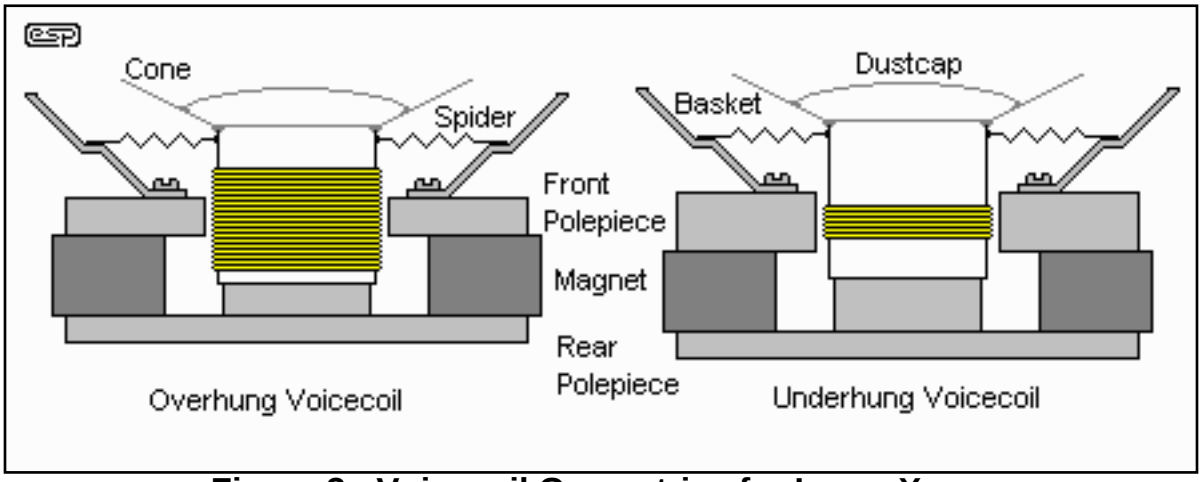


Figure 3 - Voicecoil Geometries for Large Xmax

In each case, *linear* Xmax is defined as the distance the voicecoil can travel, while

remaining within the magnetic field. Xmax is often stated as the maximum cone travel before the suspension prevents further movement (or the coil former hits the rear polepiece, which will damage the former). IMO this is very misleading, and while the cone may well be capable of travel to the distances (Xmax) claimed, the distortion becomes unacceptably high. Distortion can easily exceed 20% in many cases.

This is only a part of the problem however. For the sake of easy explanation, imagine a cone 25mm in diameter, but with a linear Xmax of 350mm. It would reproduce little or no bass at all, despite its massive excursion. At the other end of the scale, a 350mm diameter loudspeaker cone with an Xmax of 25mm will be an efficient reproducer of very low frequencies. The problem with generating bass using excursion to 'replace' diameter is the radiation impedance of the cone area. Cone loudspeakers are inefficient at the best of times, not (only) because the coil and magnetic circuits are not always optimised, but because there is a huge mismatch between the impedance of the cone and that of the air which must carry the soundwaves. This is but one of many compromises that must be made in any loudspeaker driver design.

An ideal bass radiator has a cone that is large, very light (or at least a lot lighter than most of the current crop of subwoofers), is still strong, and has a low resonant frequency. Unfortunately, this is very difficult to achieve unless the suspension is made very "floppy" indeed, and this will cause problems since it is not possible for such a suspension to keep the voicecoil properly centred when the speaker is driven hard. A low resonant frequency with a light cone also means that the driver will be very sensitive to the enclosure size (large Vas), and will need a big box to work well. This all goes against the current trends (and the need for the sub to be "acceptably invisible" in the listening room).

Horn loudspeakers minimise the impedance mismatch between diaphragm and the air, by acting as an acoustical transformer. The horn flare provides a controlled expanding wavefront to transform relatively small diaphragm movement into a large air movement, and with careful design efficiencies of well over 100dB/W/m are easily achieved. Unfortunately, bass is again a problem, since the circumference of the mouth should be equal to the wavelength of the lowest frequency to be reproduced. At 20Hz, this means a circumference of over 17 metres, and for a square horn (less than ideal, but smaller than circular) that means 4.3 metres per side! That is very large indeed, and the length hasn't even been considered yet. To avoid impedance mismatch (which causes "ripples" - peaks and dips in the frequency response), a horn should have a minimum length of 1/4 wavelength , and preferably more. This is not a practical solution for most listening rooms.

Radiation impedance is a complex area of acoustics, and I'm not going to even try to explain the maths involved. Suffice to say that small cones (in small boxes) will be incapable of reproducing spectacular bass, regardless of the claimed Xmax of the driver. As an example (and please note that this is a very rough estimate only) the following table shows the expected minimum frequency of various sized drivers in a sealed enclosure having a small baffle, and radiating into ½ space.

Driver Diameter	Minimum Frequency (-3dB)
200mm (8")	40 Hz
250mm (10")	32 Hz
300mm (12")	27 Hz
380mm (15")	21 Hz
450mm (18")	18 Hz

Table 2 - Diameter Vs. Minimum Frequency (½ Space)

The above figures should be taken as a guide only, and are based on wave propagation (as opposed to pressure mode - see below). The figures are independent of Xmax! When the subwoofer is located at a room boundary (e.g. close to the wall and sitting on the floor) there is an increase of low frequency energy, and a driver will reproduce to a lower frequency than indicated, but with poorer frequency linearity. Corner location gives a further boost, but at the expense of further frequency non-linearities. Vented (or passive radiator) boxes perform differently, and it is not possible to cover all the permutations here.

It has been claimed (and it is true up to a point) that distortion of extreme bass is not a problem, since it is masked by the main system, and may be rendered inaudible. While this is fine in theory, remember that the distortion will be primarily 3rd and 5th harmonics. If a 50Hz signal has significant distortion because Xmax has been exceeded, then you *will* hear 150Hz and 250Hz harmonics. Further, these will allow precise location of the sub, since 250Hz is easily high enough to allow our ears to localise the sound source.

The same thing happens with vented subs - if vent noise becomes audible during low frequency programme transients, then you will again be able to pinpoint the sub's location quite easily. This detracts greatly from the listening experience. In all cases, a hi-fi system should be as transparent as possible, and do nothing to draw attention to itself. Only the programme material is important, and it is this - and this only - that you should be listening to.

5 - Soundwave Vs. Pressure Mode

One claim that is very popular is that you cannot reproduce bass in a small space. While this is certainly true of soundwave propagation, it is completely false if pressure mode is excited within the space. In most rooms, there is a transition point between soundwave propagation and pressure mode (sometimes referred to as 'Room Gain'). That low frequencies can be reproduced in small spaces is clearly shown by headphones and car subwoofer systems. Both of these rely on pressure mode, and the low frequency -3dB point is defined by 'leakage' because the space is not perfectly sealed. In a completely sealed environment, bass reproduction is attainable right down to DC - not that this is actually useful. Most subwoofers ultimately rely on pressure mode to obtain the lowest frequencies in typical rooms, and it is notable that vented systems in particular are unable to excite the pressure mode properly in many rooms.

One of the reasons for this could be that there is a vent that allows the pressure to equalise. In theory, this is meant to be a resonant system, where the back wave of the loudspeaker is inverted in phase and augments the main cone wavefront. That the principle works is demonstrated by many large systems in auditoria, theatres and even outdoor venues, but all of these are large spaces where sound*wave* propagation is dominant.

When room dimensions become small compared to wavelength, soundwave propagation will not work, and bass can only be reproduced by pressurising (and de-pressurising) the listening space ... pressure mode. In tests I have performed in my workshop, a vented subwoofer seemed to make a lot of noise, but completely failed to produce bass that could actually be felt. A similar driver in a sealed box causes the whole house to vibrate, something that I have not been able to achieve to the same (or even similar) levels using any vented subwoofer system.

To some extent, this section is comprised of musings on my part, and at this stage I have no measurements to back up my claims. The effects I have reported are (or at least seem to be) real, but there is no clear point of delineation - I don't know exactly where soundwave propagation ceases to be the dominant force, and where pressure mode takes over. This will be covered in more detail when I get the chance to do some measurements and research.

6 - Subwoofer Positioning

One of the easiest ways to position a sub is to place it in the listening position - in the chair. Play material that has significant low frequency material, and then crawl around the room, placing your head in the most desirable potential locations. Listen carefully to the bass - it should be smooth and extended, with a minimum of large peaks or dips. The optimum position for the sub is now the location where you heard the best response.

It is very likely that the position of best response is completely undesirable for other reasons, so be prepared to spend a fair bit of time moving around, and listening carefully. There are always compromises, but with care you can still find a location that is acceptable aesthetically, is not inconvenient (e.g. the middle of the lounge room doorway), and does not cause howls of protest from SWMBO (She Who Must Be Obeyed).

It is a given that the other members of your family will naturally assume that you have finally lost it completely during this exercise, but it *may* be possible to get their assistance - or at least a second opinion. Involvement in the process could make it a lot easier to explain why the china cabinet really should be moved - preferably to another room if you have a powerful subwoofer

---

7 - Conclusion

It is probable that this article has not helped a great deal, and may even add to the confusion. Apart from the more common sealed and vented subwoofers, there are also dipole subs (of various types and designs), horns (which are generally too small, but often work well despite this), and even whole walls of low frequency drivers. No one system type can be recommended for all applications, and most drivers will function at their best in only one box type. Sub drivers that are designed for small sealed boxes will usually not work well in a larger vented box, and vice versa.

This should be considered a work in progress, and more information will be added as soon as additional tests can be run, and more details obtained.

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © ESP 01 Mar 2004

Elliott Sound Products — **The Audio Pages**

Professional results for the Do-It-Yourself enthusiast

Total Visitors since Jan 2001 - 5322199

Pages Last Updated 26 May 2006

Welcome to ESP, the home of DIY audio articles, projects and shortform kits. Audio is a wonderful hobby, and with DIY there is a huge scope to get the very best performance for a lot less money than otherwise. DIY audio is fun and educational, and by building your own audio equipment, you get to learn electronics as well as build equipment that would cost a great deal more in the shops. ESP is committed to providing high quality projects that are known to work, printed circuit boards (PCBs) for the most popular projects, and to assist you if necessary to complete or debug your project. There is a very wide range of power amplifiers, preamps, guitar/ bass amplifiers, studio equipment, effects, and other projects to build, including loudspeakers, subwoofers, and many others. Purchase PCBs to gain access to even more information.

IMPORTANT: Please see the disclaimer for information about these pages.

		Update
Navigation		
Site Map	Complete index of all available pages on this site	
Search	Search the ESP pages to find what you are looking for	
Forum	Discuss ESP projects and articles with other readers on the new ESP Forum.	
Products & Services		
Price List	Complete Pricelist for all Products	*02 Feb*
Purchase	Project Printed Circuit Boards and other ESP merchandise	*02 Feb*
Articles	Amplifier and power supply design, loudspeakers, basic theory, etc. new	*11 May*
Projects	Complete designs and ideas, with schematics and construction details	*21 Apr*
App. Notes	Application notes, covering a wide range of different applications, not necessarily audio related	*Jul 05*
Education	Customised projects for schools, technical colleges and universities	*Apr 03*
Consulting	ESP also offers custom design and consulting services	*2002*
Privacy, Ethics & Security		
Privacy	ESP's privacy and ethics policy	*2002*
Disclaimer	The fine print. Please make sure that you read it thoroughly.	*2002*
Spam/Scams/Security	Some things for you to consider, regarding the threats posed by the Internet	*Mar 06*
Editorial		
ESP Column	I'm As Mad As Hell, find out why	*Mar 04*

Useful & Miscellaneous

Contribute	If you have an article or project you would like published, let me know	*Jun 05*
Downloads	Various useful programs and spreadsheets to make your life easier	*Dec 04*
FAQs	Frequently Asked Questions. Check here first	*01 Feb*
Links	to my favourite sites (plus additional and miscellaneous links)	*Apr 05*
Humour	Jokes and other humour. Contributions most welcome	*2002*
Comments	Links to readers' comments and responses	*2002*

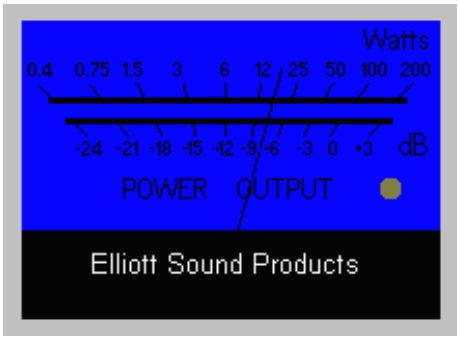
General Information

About ...	The history of The Audio Pages	*2002*
The Author	For those brave or foolish enough to want to find out about me	*2002*
Philosophy	The general philosophy of The Audio Pages and the author.	*Mar 04*
Why DIY?	There are many reasons that people DIY, and there are also many benefits.	*Mar 05*
Contact	Send e-mail to ESP	

Dates shown in ***RED*** indicate new or updated pages, typically within the last two months.

ESP projects and articles are free for personal use only. Any commercial use is strictly prohibited by international copyright laws without the express written consent of Rod Elliott. Images and text are copyright © 1999-2006 Rod Elliott, unless otherwise stated. Web and all graphics design by Rod Elliott except for specific (non ESP) logos where used, whose copyright is the property of the respective owner(s).

Links - Anyone wishing to link to my pages may do so freely, but please make the main index page (sound.au.com) the sole link. Do not link to individual pages or image files without asking first. This ensures that visitors get to see all that the site has to offer, and not just a part of it.



Anyone wishing to do so may use this image for linking to the ESP site.

Sign My Guestbook View My Guestbook

Elliott Sound Products
PO Box 233
Thornleigh NSW 2120
Australia
ABN 77 069 281 668

This site comes to you from Sydney, Australia.

Thanks for visiting, and please sign my Guestbook.

The Audiophile Webring

[Join Now | Ring Hub | Random | **<< Prev** | **Next >>**]

Visit a complete list of WebRing memberships here

Problems with this site? Please e-mail webmaster@sound.au.com

The ESP logo is a registered trade mark of Elliott Sound Products.

Best Viewed With Any Browser

Best Viewed at 1024 x 768 Screen Resolution

Elliott Sound Products **Project 09**

24 dB/Octave 2/3-Way Linkwitz-Riley Electronic Crossover

Rod Elliott - ESP
Last Updated 06 Sep 05

Please Note: PCBs are available for this project. Click the image for details.
Click here to see a photo of the board in 12dB/Octave format (See Project 81 for details of this modification).

Introduction

The Linkwitz-Riley filter featured here has (almost) perfect phase-coherency, with no peaks or dips at the crossover frequency. The design is adaptable to 2-way or 3-way operation, and all formulas are provided below (or use the ESP-LR component calculator program).

The photo shows the standard P09 circuit board, but wired for 12dB/octave operation as described in Project 81.

2-Way Linkwitz Riley Crossover

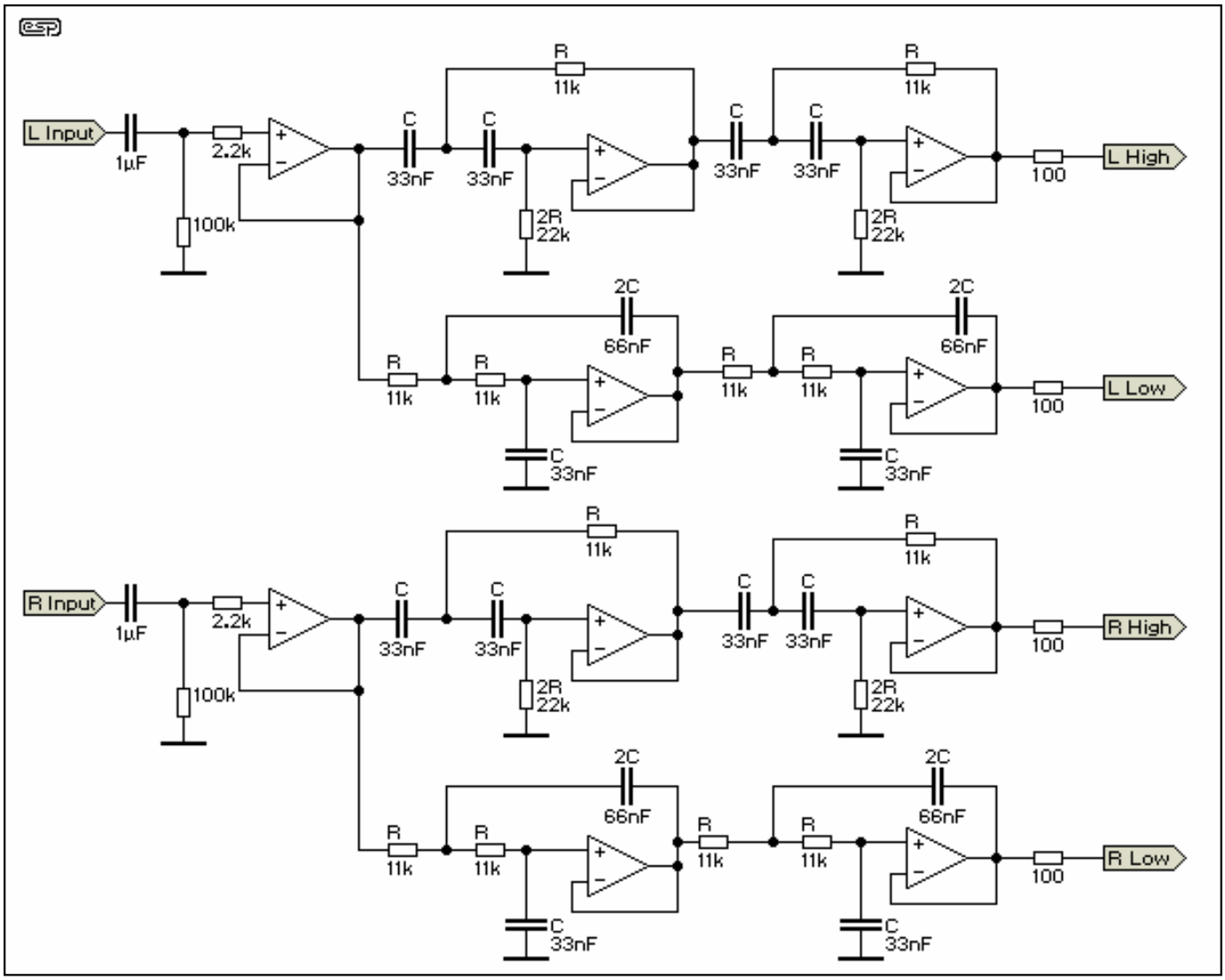


Figure 1A - Stereo Version of a 2-Way LR Crossover

Figure 1A shows a full stereo version, with two identical filter sections. With the component values shown, these have a crossover frequency of 310Hz (refer to the article on Bi-Amping to see the reason for my choice of frequency). This unit will provide a completely flat frequency response across the crossover frequency, with the signal from both filters remaining in phase at all times.

The 2-Way unit is separated into 3 sections per channel ...

- **Input Buffer** - ensures that all filters are driven from a low impedance source, to prevent frequency and phase shifts
- **High Pass** - as shown, frequency is approx. 310Hz
- **Low Pass** - as shown, frequency is approx. 310Hz

It is important with both versions that the filters are properly matched, both within the individual filters, and between channels. While small variations between channels will not be audible, if the high and low pass sections are not accurately matched, then phase and amplitude errors will result.

3-Way Linkwitz Riley Crossover

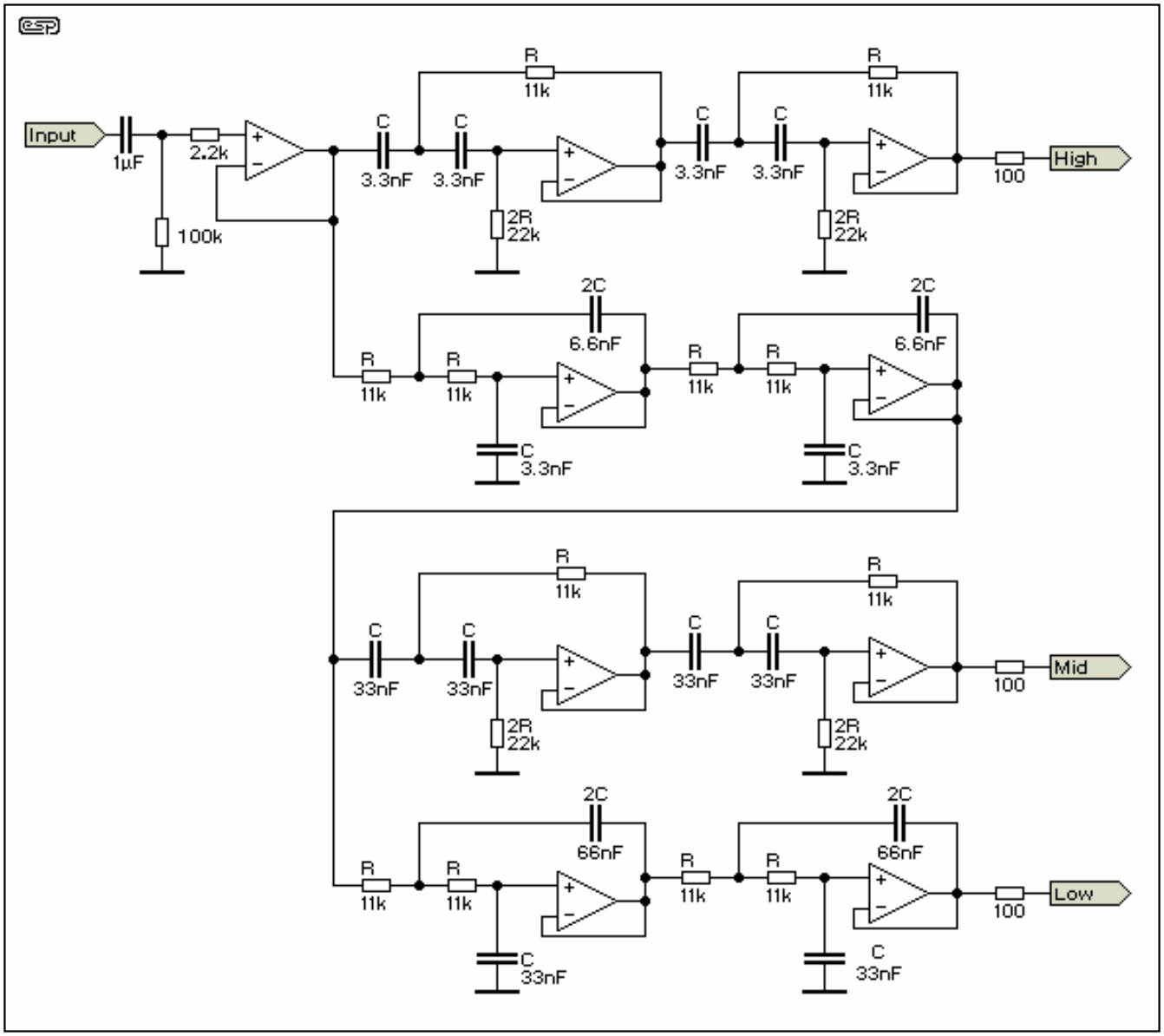
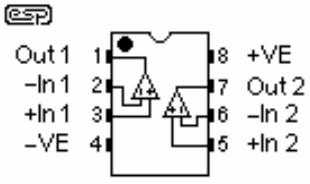


Figure 1B - 3-Way Mono LR Crossover (2 Needed for Stereo)

Figure 1B shows the way to connect a 3-Way crossover. This unit will produce excellent results, with good phase coherency and a flat response across the entire frequency band.

I know the circuits look complicated, but each is basically repetition of a common circuit "block" - the filter section. Since the opamps are all used as unity gain buffers, the use of premium devices is not really essential, so the TL072 type would be quite servicable in this role. Needless to say, if you want to use better devices (even discrete opamps) you can easily do so. Make sure that any device used is stable for unity gain - this is not always the case with some devices, especially when external compensation is used. In this case, use the manufacturer's recommended value of stability cap for unity gain operation.



Power supply connections (and bypass capacitors) have not been shown, but the diagram shows the standard connections for a dual opamp. The IC is viewed from the top. The +/-15V power supply described (see Project 05 - Power Supply For Preamps) is suitable for this crossover as well. For dual opamps, power is connected to Pin 4 (-ve) and Pin 8 (+ve).

NOTE: Only one channel is shown for the 3-Way - for a stereo setup, two identical filter circuits are required.

As can be seen, the 3-Way unit is separated into 4 sections ...

- **Input Buffer** - ensures that all filters are driven from a low impedance source, to prevent frequency and phase shifts
- **High Pass** - as shown, frequency is approx. 3100Hz
- **Band Pass** - as shown, frequencies used are high pass at 3100Hz and low pass at 310Hz
- **Low Pass** - as shown, frequency is approx. 310Hz

In 3-Way mode, the bandpass section must have a highpass section whose frequency is exactly equal to that of the main lowpass filter, and a lowpass section whose frequency is equal to the main highpass filter. (No, this isn't confusing, it just looks that way.) See the chart above for clarification if this doesn't seem to make sense.

If it helps, I have added a block diagram that may make things clearer. This is shown below, and has all the sections for a 3-way crossover network. Again, this is mono, so two complete blocks are used for a stereo system.

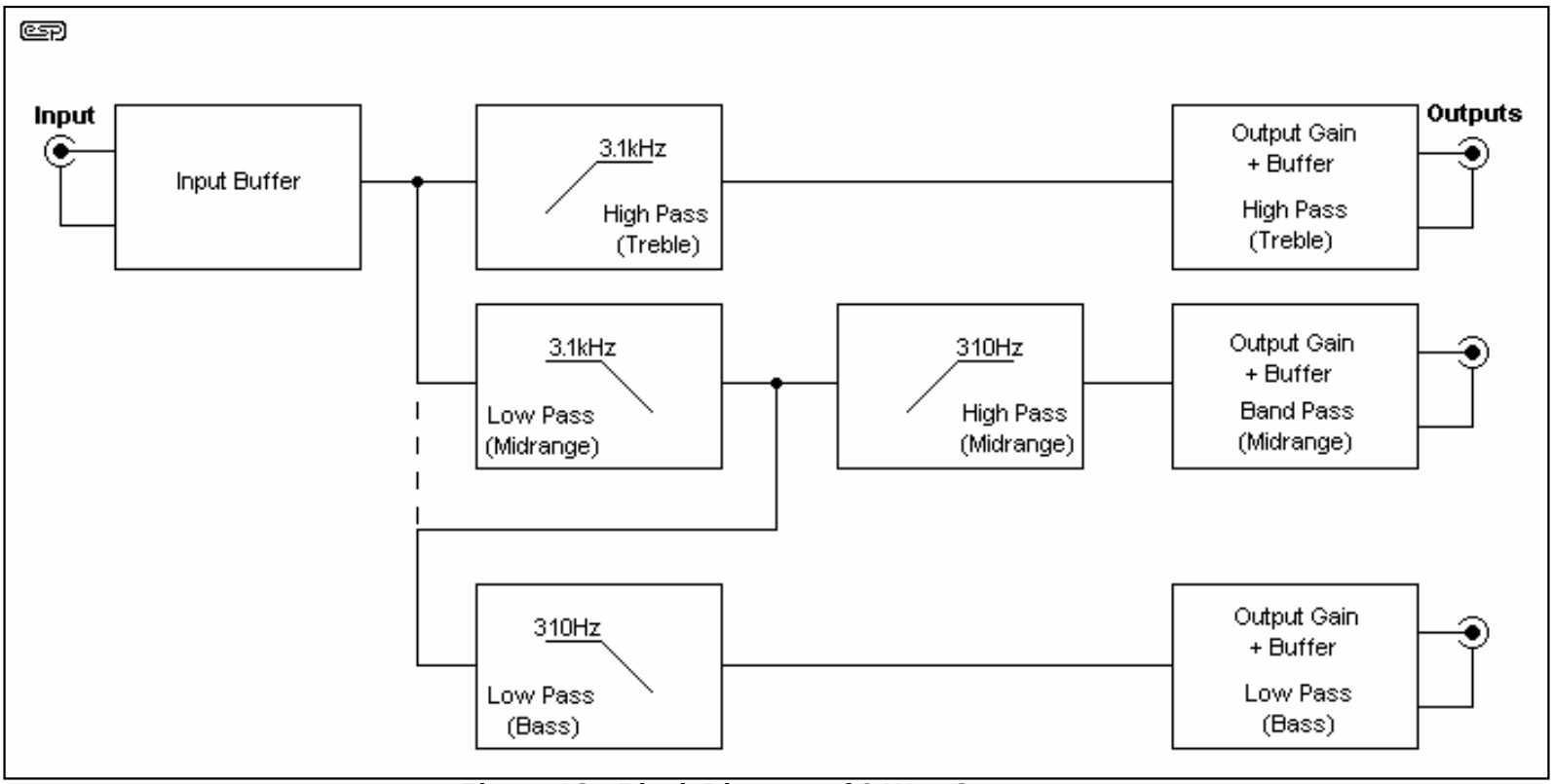


Figure 1C - Block Diagram of 3-Way Crossover

Frequencies shown are for reference only, and are the same as described above. Naturally, these may need to be changed to suit your application.

The frequency responses of each section are shown below, note that the crossover frequency is at the -6dB point, and not at the traditional -3dB frequency. This is an important difference between a Butterworth and Linkwitz-Riley filter, and allows the signals to be in phase across the audio band, regardless of which

filter section they are being passed by. The summed output of this filter is flat, there are no peaks or dips, and no phase reversals are produced (unlike Butterworth filters).

A simple test with any electronic crossover is to connect a 10k resistor to each output, and join the other ends together. Run a frequency sweep from an audio oscillator into the input, and observe the output level at the output of the resistor summing network. Most crossovers exhibit a 3dB increase at the xover frequency, and drop back to the reference level about an octave or so each side. This is a less than ideal situation, since in most cases a similar effect will occur from the speaker's summed acoustical output - assuming that the drivers are "time aligned" so the output of each is in phase (acoustically speaking) at the crossover frequency. If time alignment is not done, and the physical distance difference between speaker voice coils is large (more than 0.1 wavelength of the frequency concerned), then other acoustical differences caused by phase will tend to overshadow any anomaly in the crossover network.

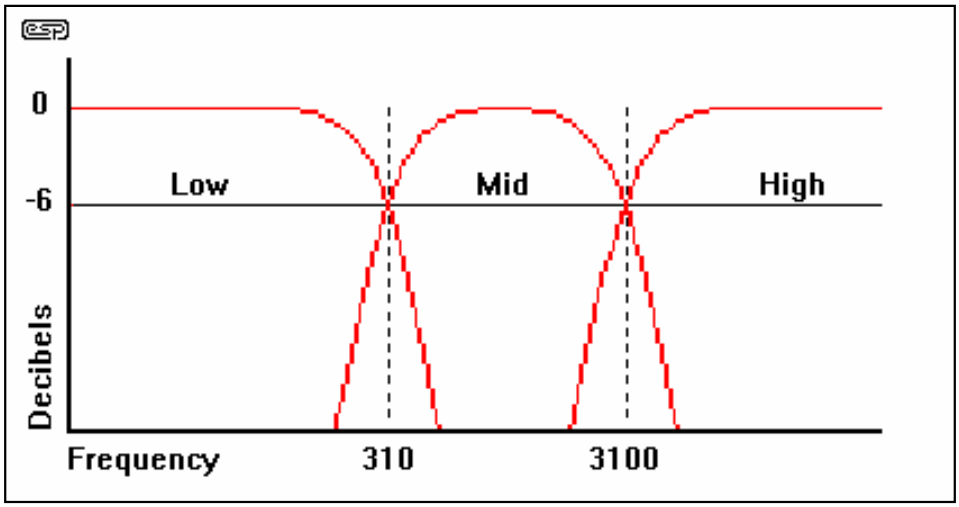


Figure 2 - Frequency Response of 3-Way Linkwitz-Riley Crossover Network

When the original graphs were produced, frequency response ranged from 10Hz to 50kHz, and the influence of the 1uF capacitor in the input is negligible (the -3dB point is 1.59Hz). Insertion loss is 0dB, since there is no gain or loss introduced by the filters in their pass-band.

Output Buffers (and ....)

If you are going to use the crossover, you will need some way of equalising the levels from each output to match the power amp sensitivity and speaker efficiency. I have had several requests for this, so the circuit for a suitable buffer is shown in Figure 3. There is nothing special about it, but it is designed to give a gain of 2 to allow maximum flexibility, and ensures that the impedance of the pots does not cause any high frequency loss with long nterconnects.

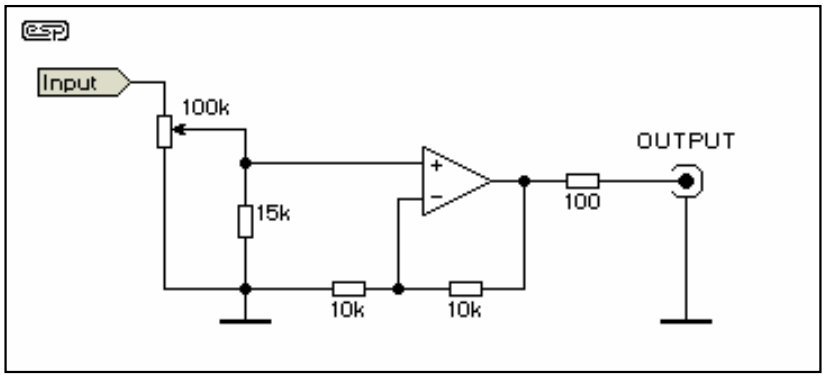


Figure 3 - Buffer Stage. One Per Output Needed

These buffers should use high quality opamps, and can be mounted on a separate board if desired. Construction is not critical, but proper bypassing must be used to ensure stability.

Several people (including me) have found that the crossover unit has a short 'chirp' or 'snap' (depending on the opamp characteristics) as power is removed, and this may be accompanied by some DC swing. If you use the new version of the P05 preamp power supply, the auxiliary output can be used to activate a 6-pole relay (or as many smaller relays as needed) to short all outputs to earth when there is no power. The normally closed contacts simply short the outputs to ground, and when power is applied the short is removed. P05 (Rev-A) boards use a loss of AC detector that will mute the crossover almost immediately when power is turned off.

Because all common opamps have short circuit protection, this will not cause any damage, and current is limited further by the 100 ohm output resistors.

My thanks to Chris Hoppe for initially alerting me to this problem.

Variable Frequency Crossover

As you can see from the main circuit diagram, a 4th order Linkwitz-Riley would be difficult to make into a variable network, due to the large number of resistors which need to change. Use of multi-ganged potentiometers is discouraged, because of the matching requirements. Sufficiently accurate 8-gang pots are unlikely to be readily available!

One possibility is to build modules, containing the tuning components. These may be plugged into suitable sockets on a printed circuit board, which otherwise contains the rest of the circuitry. The modules could be made quite small. A possibility for this option is shown in Figure 3, showing the connections which need to be made for each unit. It would not be possible to make these "hot-pluggable" (i.e. with power on), since the opamps will be deprived of their biasing resistors, and will swing to the +ve or -ve rail. Power amplifiers and speakers are not expected to enjoy the experience!

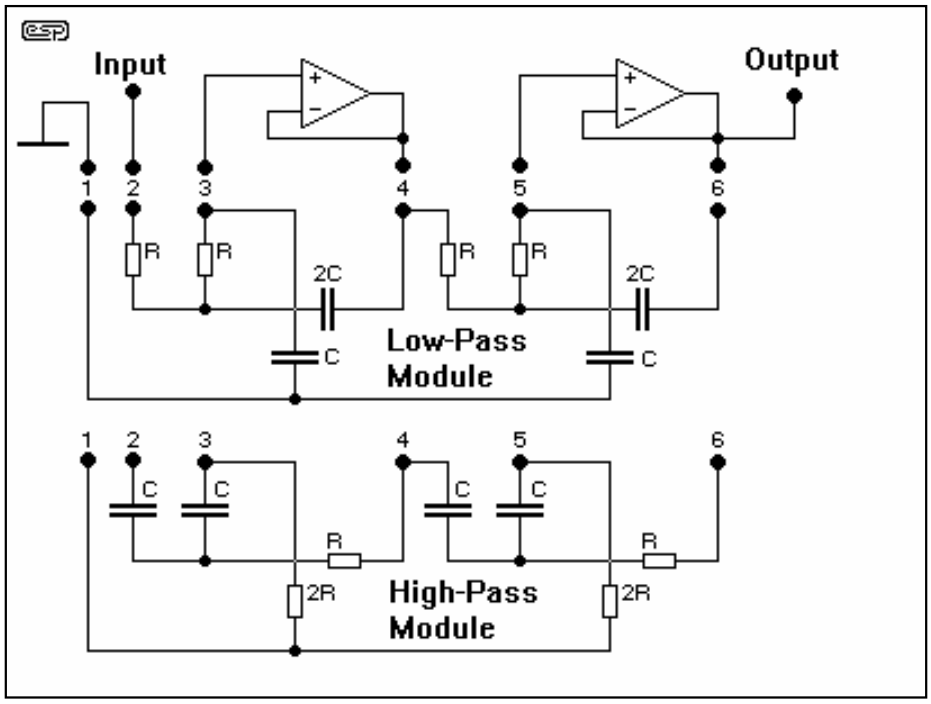


Figure 4 - Modular Connections for L-R Crossover

With this arrangement, the main PCB will simply contain a whole bunch of opamps (and their power supply bypass components), and some suitable sockets for the tuning modules. Each plug/socket combination needs 6 pins (as shown), with one for the earth (ground) connection, and the others for connection to the opamps. The plugs and sockets should be of high quality (preferably gold plated) to ensure reliability. The standard 0.1" pitch PCB pins (often used for IDC connectors) would make a good choice - they are quite cheap and reliable. Sockets are also available, but you might have to search for them - I have not seen them in any of the local retail electronics shops in Australia.

It might be an idea to use a 10M Ohm resistor (which will have virtually no effect on the tuning resistance) from the opamp's +ve input terminal to earth, just to ensure that bias cannot be lost. This will cause some offset, but it will be a lot less than 15V.

Tuning Formula If you absolutely insist on performing the calculations yourself, the formulae are ...

(1) $R = 1 / (2 * \pi * 1.414 * f * C)$
(2) $C = 1 / (2 * \pi * 1.414 * f * R)$
(3) $f = 1 / (2 * \pi * 1.414 * R * C)$

Where R = resistance in Ohms, π = 3.14159, 1.414 is $\sqrt{2}$, f = frequency in Hertz and C = capacitance in Farads

(1) This assumes that you have selected the capacitance first, which is the most sensible, since they are available in fewer different values in each decade than resistors. Capacitors follow the "E12" series, which has 12 values per decade, so:

1.0, 1.2, 1.5, 1.8, 2.2, 2.7, 3.3, 3.9, 4.7, 5.6, 6.8, 8.2, 10

These are multiplied by 10, 100 (etc), to obtain all the values from 1nF - 10nF, 10nF - 100nF, and 100nF -

1uF. Values above 1uF and below 100pF are generally not as readily available in all values, and should be avoided for this design, since very large or very small values will create impedances which are too difficult to handle. Very low values of capacitance mean that even small amounts of stray capacitance on PCB tracks or in wiring will create errors. Large values of capacitance will imply low impedances, which opamps may not be able to drive without excessive distortion or clipping.

(2) Is the least useful, since the range of capacitor values is only half that of 1% resistors. Really strange values can be assured, which will require parallel combinations of smaller caps - messy.

(3) Is useful to check that the components selected will give you the frequency that you first thought of, or something reasonably close after standard component values have been substituted for the theoetical values you will get with the calculation.

The calculator program is far easier and more fun, too. (Of course I like it - I wrote it!)

Capacitor values need to be accurate - the standard offering is +/-10%, which is not really good enough. If you have (or can get access to) a capacitance meter, simply buy more than you need (they are inexpensive), and select the values to be within 2% or better if possible. My experience is that the tolerance of most MKT and MKP caps is actually better than that quoted, but you do need to check!

The easiest way to get the "2C" value is to use two capacitors in parallel, each of value "C".

Resistor values also need to be accurate, and 1% metal film resistors are perfectly acceptable. These are generally available in the E24 series (24 values per decade), allowing a much wider choice of values. Both the E12 and E24 series values are available in the Component Calculator (Help-Preferred Values) for reference. In some shops (oh, really?) you might even be able to get resistors in the E48 or E96 range - these offer an almost limitless range of possibilities (48 or 96 values per decade - awsome!), just don't count on it.

General Notes ...

- As noted above, some opamps create a transient signal upon application or removal of power. Because this will create a loud sound, many builders may want to incorporate a delayed action switch, to ensure that the outputs of the circuit are not connected to the load until the operating conditions have stabilised. One simple solution is described above, and will work perfectly. Alternatively, the circuit described in Project 33 (Loudspeaker Protection and Muting Circuit) can be adapted quite readily to accomplish this.

  The transients are unlikely to cause damage to any amplifier or loudspeaker, but do not sound very nice. For a system that you build yourself, there is a great satisfaction in having it perform flawlessly, so it is probably worth the effort. If you have a separate power amplifier switch, there is no need to use muting. Simply turn on the power amp a few seconds after the preamp, and turn it off a few seconds before you turn off the preamp / crossover. Alternatively, the crossover can be powered permanently - current drain is very low, and it will draw less power than a typical TV set or VCR on standby.

- The crossover as described is phase coherent, in that the phase of each signal applied to each

loudspeaker driver is essentially in phase with all other signals that have passed through the crossover. Because filters are used, the crossover is ***not*** phase neutral - there are wide variations in absolute phase as the frequency changes. This is the case with all crossover networks, from the simplest to the most complex, active or passive.

I mention this because of possible interactions between the main (Left and Right) speakers, and the centre and rear speakers in a surround sound environment. The possibility exists that in some circumstances, the phase interactions between this crossover and other crossovers in a home theatre system may be incompatible with some material. These interactions will always (***always!***) be present unless all speakers in the system have identical crossover networks - not just the same crossover frequencies, but identical networks, drivers and cabinet layouts.This is rarely (if ever) the case in reality.

- If you examine the output waveform, be aware that if your audio generator has more than 0.1% distortion, the high pass output will appear very distorted when you select a frequency more than one octave below the crossover frequency. This is *not* a fault of the crossover. Because the fundamental is attenuated the most, the harmonics are effectively increased by 24dB (for the second harmonic) and about 36dB for the third. This makes the output waveform look very distorted, yet your input signal will appear to be clean on an oscilloscope. It is difficult to see any distortion below 1% on an oscilloscope, but this amount of distortion will make the output look very nasty indeed. Do not despair - all is well.

ESP Linkwitz-Riley Component Calculator

The completed Linkwitz-Riley component calculator is available for you to download. It includes the circuit diagrams for both the high-pass and low-pass sections, and has the following features:

- Calculate resistance from a known frequency and capacitance
- Calculate capacitance from a known frequency and resistance
- Calculate frequency from the resistance and capacitance values (good for checking after standard value components have been selected)
- Includes a chart for the E12 and E24 series. Capacitors generally follow the E12 series, and 1% metal film resistors are always available in the E24 series.
- Calculate the values as a low-pass, then select high-pass. The new values are displayed, along with the circuit.
- Calculates both 12dB/octave and 24dB/octave filters.

This program (ESP-LR12.EXE) is the actual executable file. This is version 1.2 of the program, and is just under 67k, so it is not too big. There are no setup programs or such like, so you simply have to decide where to put it, and create your own shortcut. Feel free to distribute the program to friends, since I have released it as freeware - just don't change the program in any way is all I ask.

The program requires the Microsoft run-time library file VB40032.DLL which can supposedly be obtained from Microsoft's web site (a link to the exact location can be found in the Downloads page) if it is not installed on your machine. Note that the program is 32-bit, so it won't run on pre WIN95 operating systems. The following is a guide as to where the DLL should be installed ...

Windows95/98 - c:\windows\system
WindowsNT - c:\winnt\system32
XP - c:\windows\system32

In all cases, the above assumes that the C: drive is the installation drive. This will usually be the case, but some installations may differ.

Note that although all care has been taken to ensure the file is virus free, ESP cannot absolutely guarantee that this is the case - I don't appear to have any viruses on my machine, but one cannot be too careful. As with all executable downloads, use your own virus scanner to check it before execution.

Download Now (66,560 bytes)

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 1999. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Updated 06 Sep 05 - Added block diagram, info reformat./ 13 Aug 01- update, and linked new calculator./ 25 Mar 2000 - Added Buffer Amp./ 26 Oct 1999 - Modified circuit to reduce high frequency phase errors./ 20 Nov 1999 - added resistors to filter outputs on all drawings.

Elliott Sound Products **Glossary of Electronic Terms**

20 Nov 2000

Introduction

Many of the electronic terms you hear mean something to you already, and others will have you wondering. This list is far from complete, but should cover most of the more common terminology, and hopefully in a meaningful way.

Glossary of Terms

[A] [B] [C] [D] [E] [F] [G] [H] [I] [J] [K] [L] [M] [N] [O] [P] [Q] [R] [S] [T] [U] [V] [W] [X] [Y] [Z]

Algorithm: a set of mathematical "rules" applied to an input. Generally used to describe a section of computer code which performs a specific function

Alternating Current (AC): A current whose polarity alternates from positive to negative over time. The rate of such "alternations" is measured in cycles per second - more commonly known as Hertz (Hz)

Amp / Ampere: The basic unit of current flow

Ampere Hour (Amp hour, Ah): a measurement of the capacity of a storage medium (a single cell or a battery). A cell which can supply 1 Amp for 1 hour before it is discharged to a specified minimum level is said to have a capacity of 1 Amp hour

Amplification: a method for increasing the amplitude (or loudness) of electrical signals

Amplifier: An electronic device which generates a high power signal based on the

information supplied by a lower powered signal. A perfect amplifier would add or subtract nothing from the original except additional power - these have not been invented yet

Amplitude: the loudness of sound waves and electrical signals. Amplitude is measured in decibels (dB) or volts

Analogue to Digital Converter (ADC): A device that converts the infinite range of an analogue signal into discrete "steps". Normally, a good audio ADC will use sufficient "steps" to resolve the smallest musical detail. For CD, this is a 16 bit converter, having 65,536 discrete levels covering the most negative signal level to the most positive

Attenuation: the decrease of a signal's amplitude level over any distance during transmission or through purpose designed attenuators. Attenuation measures signal loss in decibels (dB)

Bandwidth: the measure of a range of frequencies containing an upper and lower limit

Battery: a bank of individual cells connected together to provide the required voltage

Binary: the basic counting system used in computer logic. Two values are available - 0 and 1. A zero is normally represented by a 0 Volt signal, and a one by a voltage of approximately 5 Volts - these levels are dependent upon the type of logic used

Binary Code: a coding scheme that communicates information by using a series of "1s" and "Os" that are represented, respectively, by the digital "ON" and "OFF" states

Bit Stream: the bit rate, or flow of information, between a sender and receiver in digital communication. Also called Digital Bit Stream

Bit: a unit of the binary code that consists of either a single "1" or "O."

Bus: a pathway that connects devices, enabling them to communicate. May be digital or analogue, including power and earth (ground)

Bypass: the practice of using (typically) low value capacitors to conduct high frequency signals either to earth or around a device with limited frequency range

Byte: a unit of the binary code that consists of eight bits. One byte is required to code an alphabetic or numeric character, using an eight-bit character set code

---

Cable: a type of linear transmission medium. Some of the common types of cables include: hook up wire, coaxial (shielded) cables, lamp amd mains cable, figure-8 (zip) cable and fibre optics

Capacitor: A pair of parallel "plates" separated by an insulator (the dielectric). Stores an electric charge, and tends to pass higher frequencies more readily than low frequencies. Does not pass direct current, and acts as an insulator. Electrically it is the opposite to an inductor. Basic unit of measurement is the Farad, but is typically measured in micro-farads (uF = 1 x 10^{-6}F) or nano-farads (nF - 1 x 10^{-9}F)

Cell: one section of a battery. The common carbon or "alkaline" cells used in battery operated equipment is an example

CMOS: (Complementary Metal Oxide Semiconductor) - one "family" of digital logic devices. Some CMOS devices can operate with power supplies from 3 Volts to 15 Volts - others are limited to the traditional logic 5 Volt power supply

Coaxial Cable: a metallic cable constructed in such a way that the inner conductor is shielded from EMR (electromagnetic radiation) interference by the outer conductor. Coaxial cable is less susceptible to more transmission impairments than twisted pair cable, and it has a much greater bandwidth; thus coaxial cable is used by most analogue and digital systems for the transmission of low level signals

CODEC: COder / DECoder - the component of any digital ssubsystem which performs analogue to digital and digital to analogue conversions

Colour Code: used to identify resistors and some capacitors, as well as wires in telephony. For telephone cables, the basic colour code for the first group of pairs is Blue, Orange, Green, Brown, Slate (grey), with white "Mates". The Mate is the most positive lead, and is the Tip connection

Compression (1): the component that joins together with a rarefaction to make a sound wave

Compression (2): the act of compressing (making smaller) a digital data stream - e.g. converting from 16 bit signals to 8bit signals. Most compression schemes are "lossy",

which is to say that some of the original data is discarded and cannot be reconstructed

Compression (3): a circuit used to restrict the amplitude variations of a signal (often combined with a limiter to set an absolute limit). Unlike digital compression, analogue compression can be "undone" to restore the original signal with little degradation

Crossover: A filter network which separates frequencies into "bands" which match the capabilities of the loudspeaker drivers within an enclosure

Crosstalk: a noise impairment when a signal from one pair of wires affects adjacent wires or one channel affects the adjacent channel

Cutoff Frequency: Normally defined as the frequency where the output from a filter has fallen by 3dB from the maximum level obtainable through the filter

---

dB - Decibel - (0.1 Bel): defined (more or less) as the smallest variation of volume detectable by ear. This is measured on a logarithmic scale, so a change of 3dB from 1 Watt is equivalent to 0.5 Watt or 2 Watts. A change of 10dB from 1 Watt is equivalent to 100mW or 10 Watts. In electronics, 0dBm is a reference value corresponding to 1mW at 600 Ohms - this equates to approximately 775mV. The threshold of sound is 0dB, and typical sounds can reach 140dB or more. Any prolonged sound above 90dB may cause hearing damage

Digital/Analogue Conversion: a method used to recreate an analogue signal that has been coded into binary data and transmitted as a digital signal.

Digital/Analogue Converter (DAC): a device used to generate a replica of the original analogue signal that has been coded into binary data and transmitted as a digital signal

Direct Current (DC): A current flow which is steady with time, and flows in one direction only

Distortion (1): Any modification to a signal which results in the generation of frequencies which were not present in the original

Distortion (2): Of phase, any modification of the phase relationship between two or more signals which causes the observed waveform to differ from the original

DSP: Digital Signal Processor - a dedicated computer circuit which performs complex changes or analysis on a digital signal, generally encoded from an analogue source

---

Electronic: The use of active electronic components (integrated circuits, transistors, valves etc) which require a power supply to function. Such "active" components will always be used in conjunction with passive components

Earth (1): also known as ground - commonly used to describe the chassis and other materials that provide a return path for power supplies and signals within any electronic device

Earth (2): also known as ground - a protective connection from wall outlet to equipment chassis to conduct fault currents away from human contact

Electromagnetic Interference (EMI): an unwanted (possibly interfering) signal emitted by any electronic apparatus. The emission of EMI is heavily regulated in most countries.

Electromagnetic Radiation (EMR): a transmission medium that includes radio waves and light waves.

Farad: the base unit of capacitance - equal to the capacitance of a capacitor having an equal and opposite charge of 1 coulomb on each plate and a potential difference of 1 volt between the plates (Abbreviation - F). The Farad is a very large value, and is more commonly referred to as the pico-Farad (pF, 1×10^{-12} Farad), nano-Farad (nF, 1×10^{-9} Farad), micro-Farad (uF, 1×10^{-6} Farad), and (less common) milli-Farad (mF, 1×10^{-3} Farad)

Filter: a circuit which is frequency dependent. The "pass band" is the range of frequencies allowed through, and the "stop band" is that range of frequencies which are blocked

Filtering: a process used to remove or accentuate specific frequencies or frequency ranges of a signal

Frequency: The rate at which an alternating current changes in a cyclic manner from positive to negative and back again (one cycle). The basic unit of measurement is the Hertz (Hz), which equates to one cycle per second

Frequency Modulation (FM): a modulation technique that records changes in an information signal by modifying the frequency of the carrier signal according to changes in the amplitude of the information signal.

Henry: The basic unit of inductance in which an induced electromotive force of one volt is produced when the current is varied at the rate of one ampere per second (Abbreviation - H)

High-pass: A filter which allows high frequencies to pass while blocking low frequencies

Hertz (Hz): the measurement of frequency. Hertz represents the number of cycles of an electrical signal measured in one second

---

Impedance: A load applied to an amplifier (or other source) which is not a pure resistance. This is to say that its loading characteristics are frequency dependent. Impedance consists of some value of resistance in conjunction with capacitance and/or inductance. The equivalent circuits can vary from two components to hundreds.

In-Phase: a condition of two waveforms when they cross the reference line at the same time and in the same direction.

Inductor: A coil of wire which exhibits a resistance to any change of amplitude or direction of current flow through itself. Inductance is inherent in any conductor, but is "concentrated" by winding into a coil. An inductor tends to pass low frequencies more readily than high frequencies. Electrically it is the opposite of a capacitor. Basic unit of measurement is the Henry (H), in crossover networks it will typically be measured in milli-henrys (mH = 1 x 10^{-3}H) and for RF micro-henrys (uH) are common

Insulator: A material that prevents the passage of electricity, heat or sound. The plastic coating on wires is an insulator, preventing the wires from coming into electrical contact with each other. Insulators are extensively used in electronics. Most good electrical insulators are also good thermal insulators

Integrated Circuit (IC): A collection of active and passive devices (e.g. transistors and resistors) mounted on a single slice of silicon and packaged as a single component. Examples include operational amplifiers, Central Processing Units (CPUs), random access memory (RAM), etc.

Intermodulation Distortion (IMD): the intermixing of two frequencies. It is often caused by non-linear distortion within an amplifier or loudspeaker system

---

Laser: Light Amplification by Stimulated Emission of Radiation. Originally, lasers were either gas or precious stone (e.g. ruby), but are now made using semiconductors. Laser light is coherent, meaning that the emitted light waves are in phase, which gives the light a strange appearance since our eyes were never designed to observe coherent light

Low-pass: A filter which allows low frequencies to pass while blocking high frequencies

---

Octave: Musical terminology, meaning the doubling (or halving) of frequency. For example, one octave above "Concert pitch" A440 Hz is 880Hz, and one below is 220Hz. Musically, each of these frequencies is "A"

Oscilloscope: An electronic measurement tool which allows one to view a waveform. The vertical axis shows amplitude and the horizontal axis shows time

---

Passive: Containing no devices which require a power supply. Passive devices include resistors, capacitors and inductors

Phase: Hmmm. Tricky..... Ah-ha! Think of a bunch of soldiers all marching happily (?) to the sergeant's cries of "Hep, rah, hep-rah-hep" - except for Pt. Johnny who is blissfully "Rah, hep, rah-hep-rah"-ing. He is 180 degrees out-of-phase with the rest (or vice-versa). So it is with musical signals, where some signals have a "phase angle" (phase is measured in degrees of rotation) which is different from other signals

Power Amp: An amplifier that is designed to drive loudspeakers or other relatively low impedance loads. Usually combines voltage and current amplification. May be integrated with the preamp (see below)

Preamp: Multiple meanings, but in hi-fi generally refers to a separate section of circuitry that includes source switching, volume and balance controls (as well as tone controls in many cases). Used to raise the level from tape decks, turntables, CD players and other music sources to a level suited to the power amplifier

---

Quasi: to some degree or in some manner, resembling. For example, a quasi complementary-symmetry output stage in an amplifier is *not* in fact complementary-symmetry, but appears to be, and acts in a similar manner

Quiescent: being still or at rest, in an inactive state. The *quiescent* current in an amplifier is that current drawn when the amplifier is "at rest" - i.e. not amplifying a signal, but supplied with power

---

Resistor: An electrical device which impedes (resists) current flow regardless of frequency. Basic unit of measurement is the Ohm

Resonance: The natural frequency at which a physical body will oscillate. An example is when you blow gently across the top of a bottle, the enclosed air resonates at a frequency determined by the internal volume. Also refers to the natural resonance of loudspeaker drivers, cabinets and ports, or the frequency where an inductance and capacitance have the same impedance (this causes maximum impedance with a parallel circuit, and minimum impedance for series circuits)

rms: Root Mean Squared. Applies to voltage and current, but is commonly (although incorrectly) applied to power. Defined as an alternating voltage (or current) which has exactly the same energy content (power) as the same value of direct current

---

Thermal Coefficient (1): Of expansion, describes the amount by which a material expands when heated. Commonly expressed as a percentage per degree Celcius so the exact size at various temperatures may be calculated. Knowledge of the expansion characteristics of different materials is important in high power semiconductor manufacture, since differing expansion rates may cause device failure due to temperature cycling fractures

Thermal Coefficient (2): Of resistance, describes the change in resistance at various

temperatures. Most metals have a positive temperature coefficient of resistance, which means that the resistance increases with increasing temperature. Carbon and some alloys have a negative temperature coefficient of resistance, so as temperature is increased, resistance decreases

Thermal Resistance: The resistance of various materials to the passage of heat energy. Most electrical conductors are also thermal conductors, with the higher electrical conductivity materials usually having higher thermal conductivity. Important in the design of high power electronics, heatsinks, semiconductor casings, etc.

Total Harmonic Distortion (THD): the sum of all amplifier distortion components, plus system noise. THD measurements are sometimes quoted as THD+noise. Usually measured at specified frequencies and power levels

---

Velocity: speed of motion or rapidity. In audio and electronics, we are concerned with the speed of a signal in air and a conductor. Speed (velocity) of sound in air is approximately 345 metres per second at sea level, but it varies with temperature and humidity. Speed of an electrical signal in a wire is approximately 3×10^8 metres per second, but may be influenced by ...

Velocity Factor: a situation that occurs in conductors that are close to another conducting material. For example, a coaxial cable has an inner and outer conductor, with insulation between the two. The velocity factor of such cables varies from 0.7 to 0.9 (i.e. the signal travels slower than in free space)

Volt: The basic unit of "electromotive force". One Volt applied to a resistance of one Ohm will force a current of one Ampere to flow (Abbreviation - V)

Watt: The basic unit of power. 1 Volt across 1 Ohm (giving 1 Amp) dissipates 1 Watt (all as heat with a resistive load)

Wavelength: the length of one cycle of an AC signal. Determined by $Wavelength = c / f$ where "c" is velocity and "f" is frequency. The wavelength of a 345Hz audio signal in air is one metre

---

Xenon: A gas commonly used in flash tubes, HID (High Intensity Discharge) automotive headlamps, and having an intense white light output with a colour temperature close to that of daylight

---

This article (including all text, images and diagrams if applicable) was conceived and written by Rod Elliott. Copyright © Rod Elliott 2000-2003, all rights reserved. Reproduction, storage or republication by any means whatsoever whether electronic, mechanical, or any combination thereof is strictly prohibited either in whole or in part without the express written permission of the author, with the sole exception that readers may print a copy of the article for their own personal use.

Main Index

Articles Index

Page created as a separate entity Nov 2000./ Apr 2003-Added definitions

Elliott Sound Products

Speaker Damage

This is the old "how does clipping blow up a woofer" problem. I know that woofers rarely blow up, even with severe overload, and in essence they seem to be immune (although that might overstate things a bit!). This is (or would be) true, but for one small detail. Few speakers (including woofers) can tolerate their full rated power on a continuous basis. The rating is more for "amplifier power" than speaker power.

Imagine that a 100W amp is just on the verge of clipping with a full range or band limited signal. Although the amp will be peaking at 100W, the average power will be closer to 5-10W, or maybe a little more or less depending on program material. If you push the amp harder, the peaks will clip, which you might not even notice if the mid+high is still clean. The average power may be 50W or so now, and the speaker's voice coil will start to get quite hot. Push it harder again, until even the lower level signals are clipping too, and the average power is now well over 100W.

Note that a 100W amp driven with a square wave or severely overdriven sinewave will give close to 200W. This problem is worse with semiconductor amps than valves, because valves have inherent inefficiencies that reduce the maximum output.

If this is kept up for too long, the voice coil will literally burn up from the heat, the adhesive lets go of the voicecoil windings, and the whole assembly starts to disintegrate. Exit one woofer.

Having said all this, it still doesn't happen too often, but generally you are better off with an amp that is rated for more power than the speaker can handle, rather than less (except for guitar, but guitar amps are driven into clipping a lot (most) of the time).

There are - naturally - limits. A 10W woofer won't last long with a 300W amp driving it at anywhere near full power on transients, but a 100W woofer will last forever with a 10W amp in permanent overload. Some common sense must be applied, but a typical 100W woofer will probably last forever with a 200W amp used sensibly.

The comment above is aimed more at the risk of tweeter damage, but it can happen with midrange drivers and woofers too, given the right (or wrong) set of circumstances. One issue with many modern recordings is excessive compression, and this can increase the average power level dramatically, even before amp clipping.

It is worth noting that tweeters are rarely blown up just by the additional harmonics generated when an amp clips. Although they are undeniably present (and sound undeniably terrible), the actual power of the harmonics is not as high as you may have been led to believe. It is the

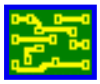
combination of these harmonics and the "compression" effects described above that do the damage, and clipping compression is the greatest offender.

Close

Elliott Sound Products **Project 33**

Loudspeaker Protection and Muting

Rod Elliott (ESP)



Please Note: PCBs available for this project Click PCB image for details

Introduction

Please see Updates 1, 2, 3, 4 and 5 on this project.

Many hi-fi amplifiers and professional power amps (and loudspeaker systems) provide some of protection, either to protect the speakers from an amp fault, and/or vice versa. Some of these are implemented at a very basic level - for example the use of a "poly-switch". The poly-switch is a non-linear resistor, having a low resistance at normal temperatures and a much higher resistance at some designated temperature. Unlike "ordinary" thermistors whose characteristics are more or less linear, the poly switch has a rapid transition once the limit has been reached.

I don't like poly-switches, because I know that the introduction of a non-linear element is going to add some degree of distortion, and because of a finite resistance, will degrade damping. This (i.e. damping) is not an issue IMHO, but to many audiophiles it is of prime importance. (I shall not pursue this argument here, however - see Impedance for more info.)

The basic requirement of a speaker protector requires that any potentially dangerous DC flow to the speakers should be interrupted as quickly as possible. There are a few issues that need to be solved to ensure that this will happen fast enough to stop the loudspeaker drivers from being damaged, and this becomes more critical if a biamped (and even more so with triamped) system is being used.

Naturally, one can simply rely on fuses. Although these also have finite resistance it is small, and use of fast blow fuses can be quite effective. The rating becomes quite critical, and fast blow types are essential. The problem with this approach is that if the fuse is of a suitable value to provide good protection, it will be subjected to considerable thermal stress since it is operating at close to its limits. Metal fatigue will create the problem of nuisance blowing,

where the fuse blows simply because it is "tired" of the constant flexing caused by temperature variations.

This project explains the principles, and shows a suitable detection method that may be applied. The speed of the relay used is another critical factor, and we shall see that the conventional method of preventing the relay's back-emf from destroying the drive transistor also slows down the response to an unacceptable degree.

The circuit also includes a mute function, which leaves the speakers disconnected until the amplifier has settled, and disconnects the speakers as quickly as possible after power is removed to prevent the turn-off noises that some amps generate. These can range from a low level thump 5 to 10 seconds after power is turned off, to whistles, squeaks and other strange noises that I have heard from amps over the years.

Description

It is important to identify the lowest frequency likely to be passed to a speaker, because this determines the delay that must be introduced to prevent low frequencies from triggering the protection circuit (nuisance tripping). For practical purposes, a low frequency limit of 20Hz is satisfactory for a full range system, and this means that a minimum 25ms delay is essential. In reality, due to the combination of low frequencies, and asymmetrical waveforms at higher frequencies, a greater delay will normally be required. Unfortunately, the greater the delay, the greater the risk of drivers being damaged. In a full range system (i.e. using passive crossovers), midrange and tweeters will be offered some protection by the capacitors used in the crossover network, but these are missing in a biamped or triamped system. For this reason, it is important that the circuit can be easily modified to change the initial time delay before the system detects the DC and disconnects the speakers.

The Detector

This is the most important of the functions. It must be capable of detecting a DC offset of either polarity, and be immune to the effects of asymmetrical waveforms and low frequencies. This is a common requirement, and it is most expedient to use a simple (single pole) filter to keep the complexity to a minimum. With this arrangement, a low frequency cut-off of about 0.1Hz is about right. Without boring you with the mathematics behind this, it works out (eventually) that a filter having a time constant of 1.0s will still provide the ability to detect high level DC reasonably quickly, but allow low frequencies through without triggering. With this, the relay could have its supply removed within about 50ms from the time the output voltage reaches the supply rail (this is supply voltage dependent) - due typically to a shorted transistor in the output stage. By changing the time constant of the filter, we can adapt the circuit for operation at other higher frequencies to suit a biamped (or triamped) system.

The detector can be built using an opamp, and will work very well, but this introduces the need

for low voltage supplies within the power amp. This is not always possible (or desirable), so the design uses discrete transistors throughout to allow for the different supply voltages found in typical power amplifiers.

The detector circuit shown in Figure 1 (1) is simple and works well, and as shown will not trigger with a 30V RMS signal at 5Hz, but operates in 60ms with 30V DC applied, and in 50mS with a 45V DC supply. This should be sufficient for most applications, and allows the use of a non-polarised electrolytic capacitor in the filter. These are cheap, small and quite adequate for this purpose.

NOTE: *The power supplies (+ve and -ve) shown in these diagrams will normally be the power amp supply rails. Do not try to substitute different supplies unless you know exactly what you are doing, or the circuit may not work properly. This is especially true of the muting circuit, but incorrect supplies will (may) also affect the DC detection circuit. Like most of my projects, this is intended for experienced constructors.*

Updates 1 & 5: The original circuit could damage Q3 under some conditions due to excess base current. To remedy this, a resistor must be added between the collector of Q1 and the base of Q3, as shown in the amended diagram. This prevents excess base current if the gain of Q1 or Q2 is higher than expected. My thanks to Vince for pointing this out.

21 Mar 2003 - Q3 was shown as an NPN transistor, but was marked as a BC556 - this was wrong. It is a PNP, and the symbol was incorrect. The type shown (BC556) was correct. This is now corrected (thanks David ;-)

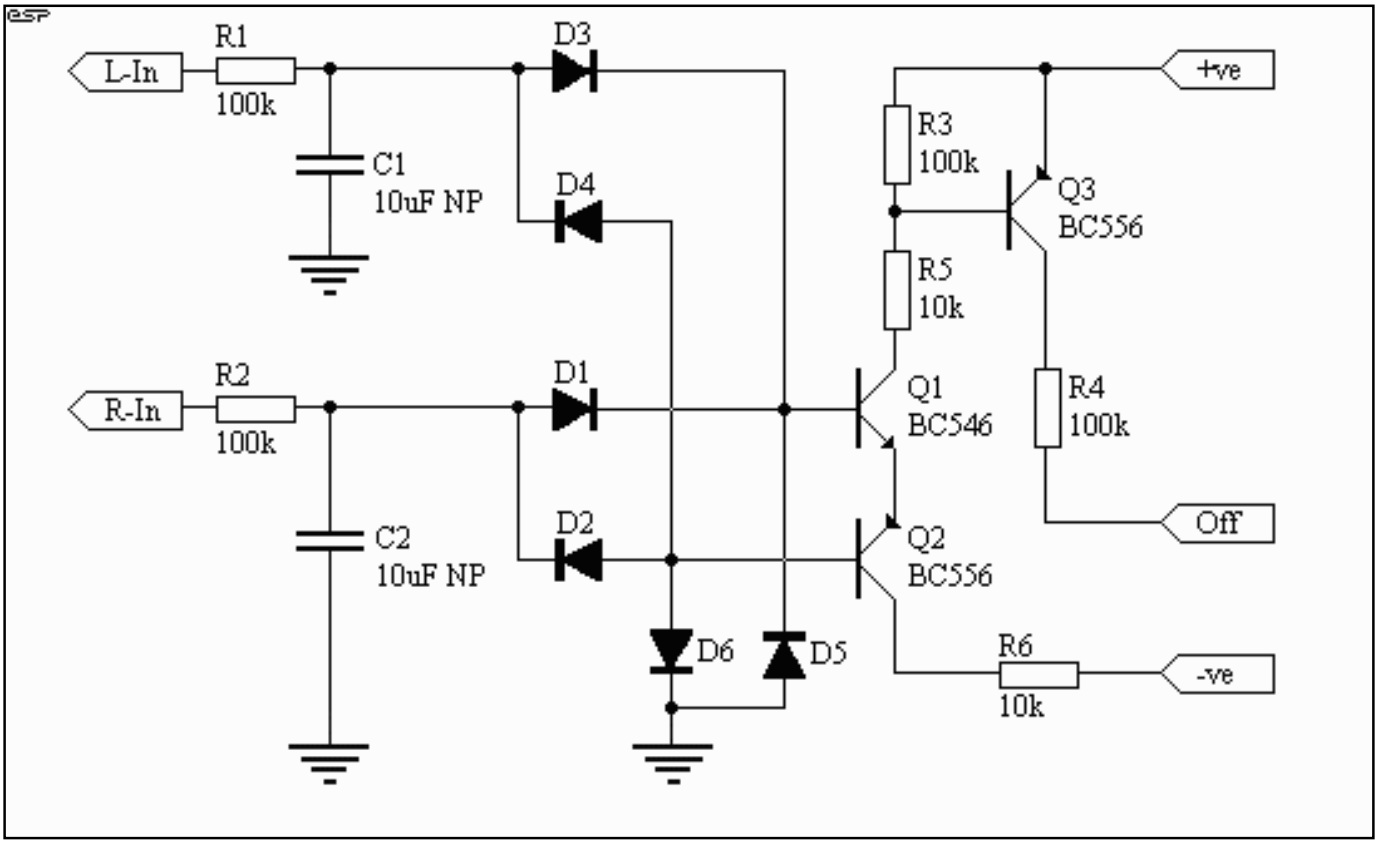


Figure 1 - Basic DC Detector Circuit

Update 4 (18 Dec 2002): It has been found by a couple of people that the detector transistors will self destruct under some conditions. The addition of R6 in the negative supply prevents this. The PCB version uses a single transistor detector and does not have this problem.

The input filter is a simple single pole (6dB/ octave) version, and although it would seem that a "better" filter would be preferable, a two pole (or more) filter will actually degrade the DC detection. This basic circuit is not new (see reference), and has actually existed in one form or another for some time. It is ideally suited for our requirements, as it is symmetrical, and with the input diodes as shown, a single detector can be used with multiple amps and different input time constants for each individual filter. The unit itself can operate on a separate supply if desired, so the complete protection circuit can be in a separate enclosure. Regulated supplies are not needed, and no hum or other "artefacts" are introduced into the speaker lines. (Please see NOTE above.)

The table (below) shows some suggested values for the filter, for use in bi- and tri-amped systems. You will need one filter and two diodes for each amplifier channel connected, and a suitable number of relay contacts to handle them all. In some cases, this will mean multiple relays.

Frequency (Hz)	**C1 Value**
Full Range	10 uF (non-polarised)
100 Hz	1 uF
300 Hz	330 nF
1 kHz	100 nF
3 kHz	33 nF

The resistor should be left at 100k for all frequencies. Do not use a conventional electrolytic capacitor for C1, because any small reverse bias will eventually ruin it. You may discover that with some types of music (especially if at high volume) may cause the circuit to false trigger. If this happens, increase the value of C1, up to a maximum of 47uF. Anything higher than this will slow down the response unacceptably.

Relay Specifications

The relays should be easy enough to obtain. At least one of the Australian component suppliers has relays that are quite suitable, but they are not particularly cheap. The current

rating is very important, and assuming a supply voltage of +/- 40V, this will cause a current of about 6A in an 8 ohm speaker if a transistor shorts. Although 6A may not sound like much, it is at DC, and because there are no periods of 0V as with AC, the arc is longer, fatter, and far more destructive of contacts than the same current using AC.

Do not be tempted to use miniature relays, because if the normal AC speaker signal is too far in excess of the relay contact rating, the contacts may become welded together - this will almost certainly happen if the DC rating is too low. You also need to consider that contact resistance is additional resistance in the speaker lead and may affect damping (albeit very marginally) and will introduce some small power loss, and the miniature types will not be suitable in this regard.

I had a look in the catalogue of one Australian supplier, and they have several relays with a 10A contact rating. I would suggest that anything lower is unwise for long term reliability. Most of the commonly available relays will have a 12V coil, and this will cause problems if the supply voltage is 30V or more. Power relays often draw significant current (typically > 60mA), and it will usually be best to connect the coils in series.

Update 1: A reader (I cannot recall who it was so I can't give the credit) mentioned that in some areas there is significant sulphur content in the air, and this causes heavy tarnishing of silver contacts. If you live in such an area, it would be advisable to obtain hermetically sealed relays if possible, to prevent the contacts from tarnishing.

It is well known that the current required to activate a relay is far greater than that needed to keep the contacts closed, and a common trick is to use an "efficiency" circuit to minimise the relay holding current. I do not feel that the additional complexity is warranted, and have not included this facility. If you really want to do this properly, see reference 1 (below). It has been claimed that an efficiency circuit also speeds up relay drop-out time because of the lower stored magnetic field. I conducted some tests, and the savings are marginal at best, although this could be different with different relays.

Figure 2 shows the relay activation circuit, and includes the connection for the mute and protection signals. No components are critical, but some will need to be modified based on the relays used. I have assumed that a minimum of two relays will be needed (one for each channel), and this increases the total relay coil voltage to 24V. If you are going to use more than two (for example, four single pole relays are needed for a biamped system), then if the supply voltage is 48V or more, all 4 relays can be connected in series. In most cases you will need to work out the value of a suitable dropping resistor from the formula below.

The terminal labelled "Off" is common to all three modules, and these points are simply joined together, as are the +ve and -ve supply connections. A positive current into the Off terminal will de-energise the relays, by turning on Q1. This steals all the base current for Q2, which

then turns off, as does Q3.

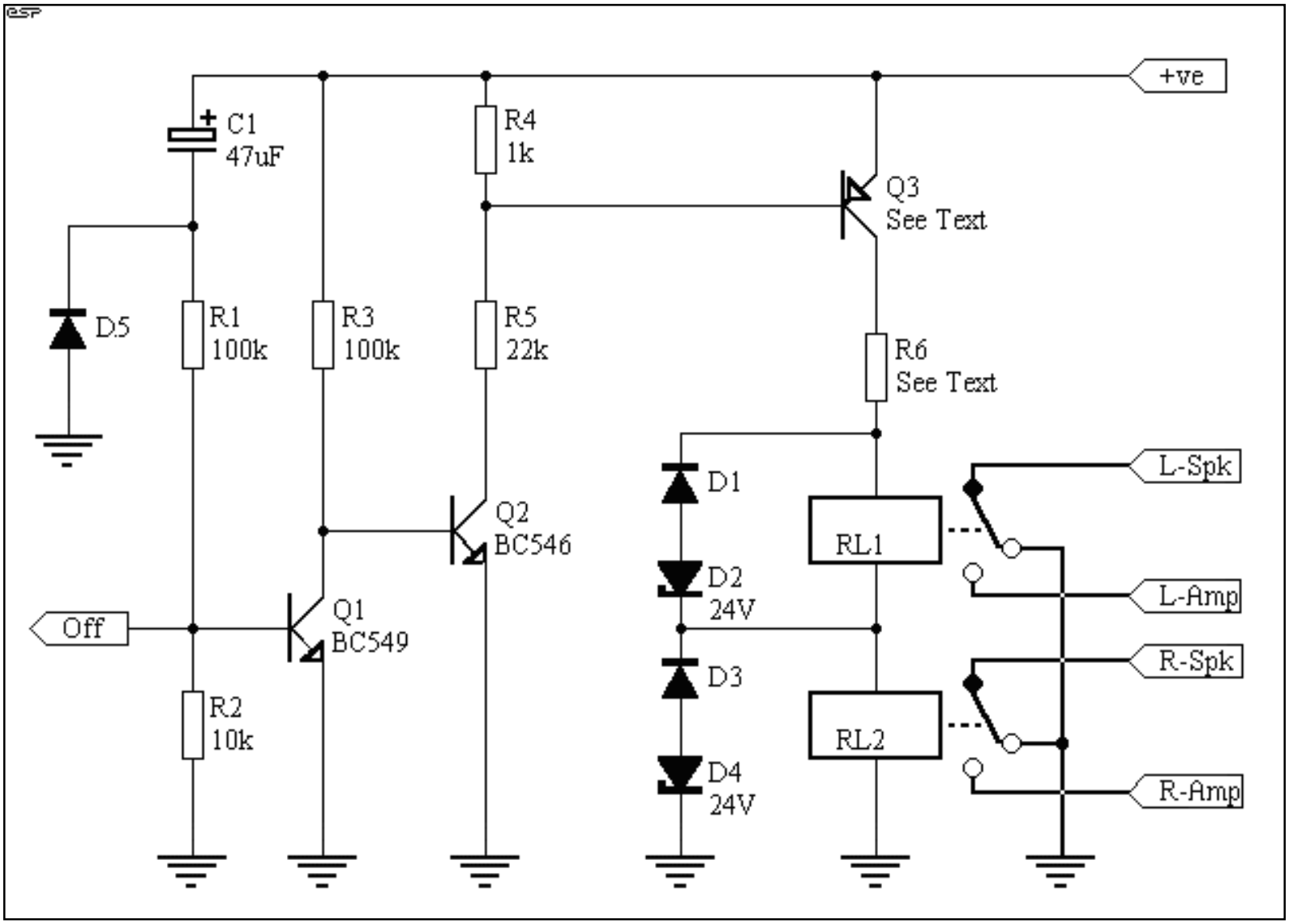


Figure 2 - Relay Activation Circuit

The relay(s) must be turned off in the shortest possible time, so the use of the normal protection diode across the coil should not be used, as it slows the response considerably. Instead, the arrangement shown still protects the driver transistor, but allows the relay magnetic field to collapse without generating a current in the coil (this the what slows the relay's release). I cannot predict the exact delay you will achieve, since the choice of a suitable relay is outside my control. You will have to pester and annoy your local suppliers to find a relay that has suitable characteristics, and be prepared to pay what will seem like an obscene amount of money for a simple electro-mechanical device.

D5 (a late addition) will discharge C1 as the supply collapses. It will not help much in the case where someone switches the power off then straight back on (not that anyone would do that!), but will reset the circuit much faster than would otherwise be the case. My thanks to Vince for the idea.

Update 3 (10 Aug 2000) It has been brought to my attention that the DC arc can (and does) destroy even 10A relays under some circumstances. To provide greater speaker protection, the relay wiring in Figure 2 has been modified to short the speaker to earth in case of a fault. This way, even if the contacts do arc it will be directly to earth. This is much safer (for the speakers), and the arc to earth will blow the fuse a lot faster than if an 8 ohm load is a part of the circuit. It is strongly recommended that this new scheme is used as a matter of course. It is worth noting that any DC protection system that does *not* use this method will almost certainly fail to protect the speakers with a medium to high powered amplifier. (My thanks to Phil Allison for the information.)

Note also that this circuit cannot be used as shown with the 12V relays in series if the supply voltage is less than +/-24V (but you knew that already ;-)

In order to work out the value of R6, subtract the combined relay voltage from the supply voltage (you must know the relay coil current!). To calculate the coil current from its resistance, use the following (I have assumed a 40V supply for the examples):

I = V / R Where V = coil voltage and R = coil resistance

So for a 180 ohm coil (fairly typical) this works out to

I = 12 / 180 = 67mA

The resistor value is worked out with:

R = V / I Where V = the "left over" voltage from the subtraction and I = coil current

You will also need to work out the power rating for the resistor:

$P = V^2 / R$ Where V is the voltage and R is the resistance

Again, for the above example, this works out to

R = (40 - 24) / 67mA = 16 / 0.067 = 239 Ohms (220R should be fine)
P = (16 x 16) / 220 = 1.16W

So for an adequate safety margin, a 2 Watt resistor should be considered the minimum (5W would be better).

To determine the transistor for Q3, add the supply voltage and the zener voltages to give the maximum collector to emitter voltage. In this case it is 40 + 48 = 88 Volts, and I would suggest that a transistor with a breakdown voltage of at least 100V be used to give some

safety margin. The MJ350 (300V rated) will be suitable in nearly (if not) all applications, or you can use a MPSA92 - lower current, but still has a 300V rating.

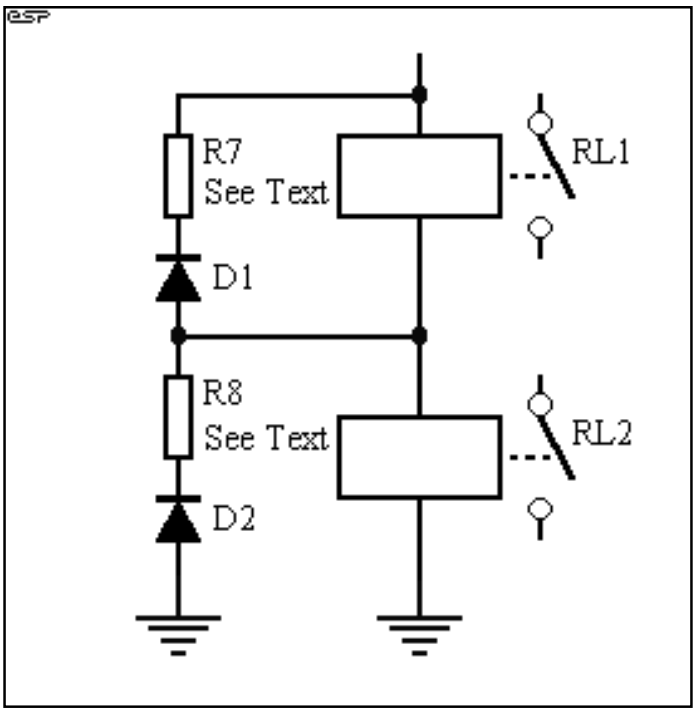


Figure 16a - Alternative Back-EMF protection

Figure 16a shows an alternative method you can use to damp the back-emf from the relay, but to implement it properly, access to an oscilloscope is helpful (if not essential). If the resistors have approximately the same resistance as the relay coils, the back-emf should (!) be limited to about the normal relay voltage, give or take 50% or so. In the tests I carried out (see Tests, below) using a 24V relay, the back-emf was limited to about -30V, which would be fine in most cases.

This method is slightly cheaper than using zeners, but is less predictable. An additional alternative is to use a catch diode to the -ve power supply. A 1N4004 between the top of the relay string and the -ve amp supply will limit the back-emf to the voltage of the -ve supply, so for the example case this would be -40V. I expect that this would be quite acceptable, but have not tried it. Make sure that the diode is connected the right way around - the cathode goes to the top of the relays, and the anode to the negative supply.

Muting

Since we have all this new circuitry, it is most worthwhile to incorporate a muting function, so that when power is removed from the system, the relay will open to stop turn-off transients from being heard. Likewise, we will normally want to mute the system for about 2 seconds after power is applied to stop the turn-on transients as well. C1 and R1 in the circuit of Figure 2 provide the turn-on delay, by supplying current to the "Off" terminal as C1 charges. Once charged, the current falls to zero, and Q1 turns off, allowing Q2 and Q3 to turn on, thus

energising the relays. (Note that this timer will not be reset if the power is turned off and back on again quickly, but since this is a procedure that should be avoided anyway, no provision has been made for it. :-)

To be able to do this effectively, we must have access to the AC from the power amp's transformer, or have the external unit controlled by the main power switch in the system. In some hi-fi installations, there will be a multiplicity of different units to turn on (and off) each time the system is used. I will leave it to the reader to decide which unit to use as the control, but would suggest that where a separate preamp is used, this could be an ideal controller for the entire system. It is unfortunate that hi-fi has not followed the sensible approach of a lot of computers, with a switched IEC connector on the back of the preamp to control power amps and other outboard devices. (I did this on my VP-103 valve preamp, and it is most useful :-)

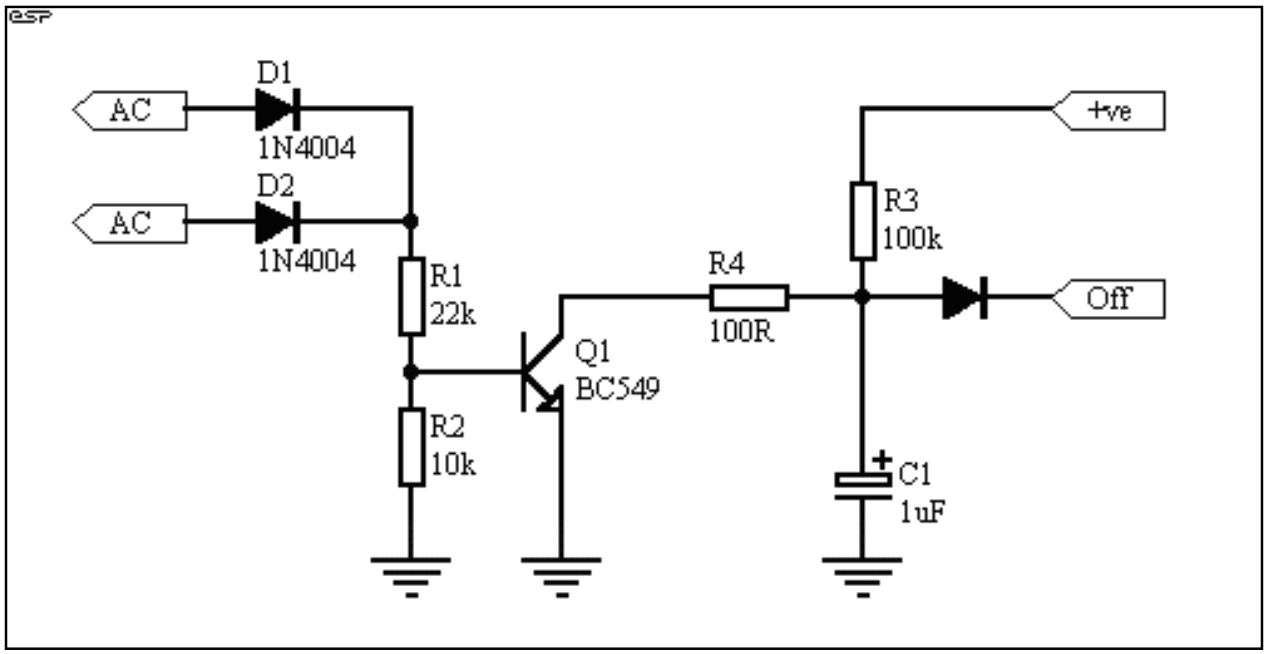


Figure 3 - Loss of AC Detector

The power detector cannot rely on the DC supply, as this may take a considerable time to collapse. The common approach is to use a rectified but unsmoothed output from the transformer secondary. Because it is not smoothed, this disappears instantly when power is removed, and is ideal. Figure 3 shows the basic circuit, and this will remove relay drive within about 50ms of the power being turned off. We could make it faster than this, but there is little point.

The circuit simply uses the current pulses to keep a capacitor discharged via Q1. When the pulses stop, the cap charges until the threshold voltage of the "Off" terminal is reached (0.65V), and the relays are turned off. After power is first applied, the timer circuit will activate the relays after about 4 seconds (typical). This can be increased if desired, by increasing the value of C1 in Figure 2.

Tests

I carried out some tests to see just how quickly the relays could be operated. The results were something of an eye-opener (and I *knew* about the added delay caused by a diode!). The relay I used was a small 24V coil unit, having a 730 Ohm coil and with substantial contacts (at least 10 Amps). With no back-emf protection, the relay opened the contacts in 1.2ms - this is much faster than I expected, but the back-emf went straight off the scale on my oscilloscope, and I would guess that the voltage was in excess of 500V. When a diode was added, the drop-out time dragged out to 7.2ms, which is a considerable increase, and of course there was no back-emf (Ok, there was 0.65V, but we can ignore that). Using the diode / resistor method described above, release time was 3.5ms, and the maximum back-emf was -30V, so this seems to be a suitable compromise.

I was not able to test the zener method prior to publication, since I did not have the 24V zeners needed on hand. I would expect this scheme to be as good or better than the diode / resistor combination. The graphs below show the behaviour of the circuit with and without the resistor and diode. The estimated 500V or more is quite typical of all relays, which is why the diode is always included. This sort of voltage will destroy most transistors instantly. It is exactly the same process used in the standard "Kettering" ignition system used in cars, but without the secondary winding, or the "flyback" transformer used in the horizontal output section of a TV set.

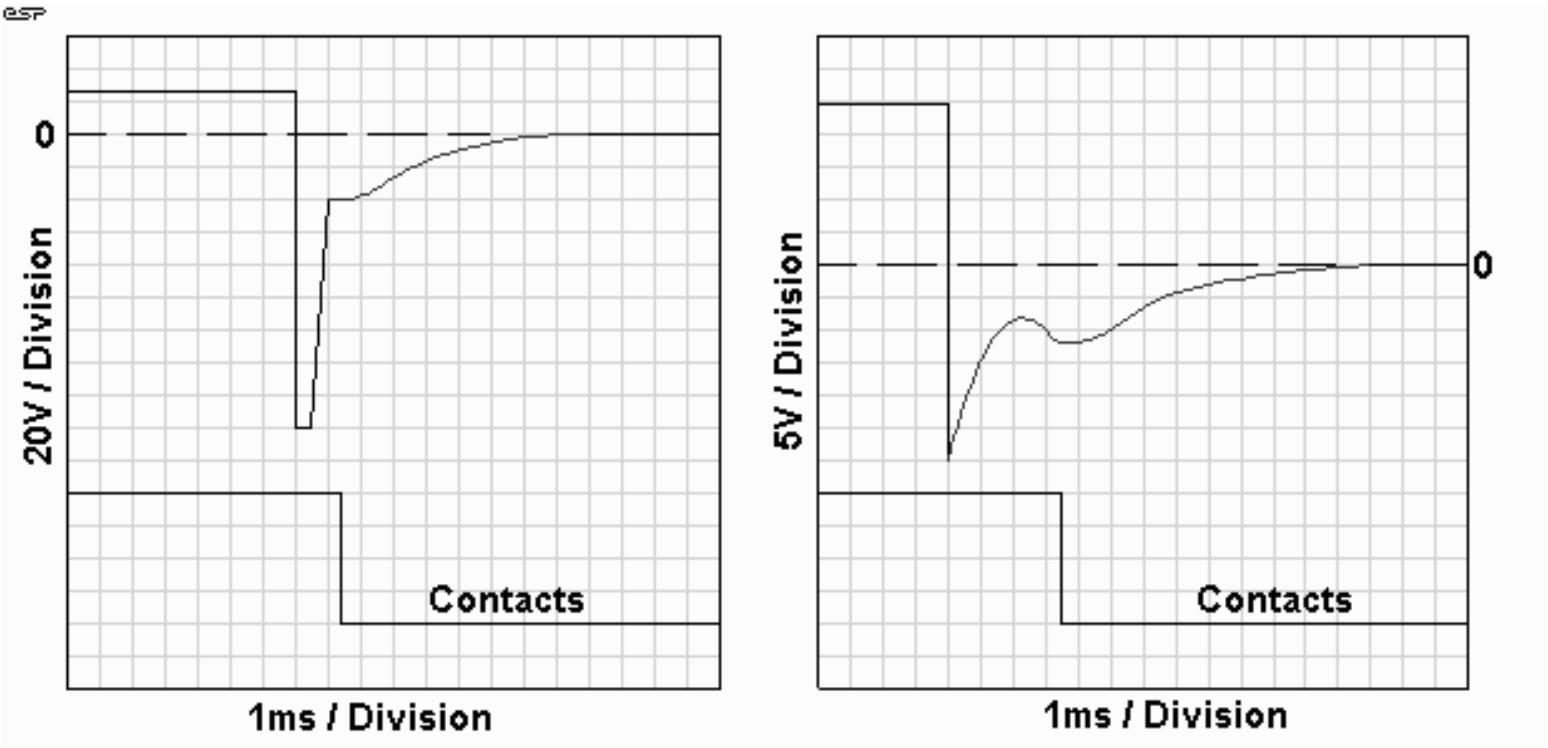


Figure 4 - Relay Voltages

The trace labelled "Contacts" is representative only, and is not to scale. The peak relay voltage (above left) exceeded my oscilloscope's input range (and I was too lazy to set up an

external attenuator), and as shown is cut off at my measurement limit. I estimate that the voltage is greater than 500V.

Note that the kink in the relay voltage curve is caused by the armature (the bit that moves) coming away from the relay pole piece, and reducing the inductance. This causes the stored magnetic "charge" to try to increase the voltage again, but it is absorbed by the resistance and dissipated quickly. The contacts open at the point where the previously closed magnetic field is opened as the armature moves away from the pole piece. As can be seen, this is 3.5ms after the relay supply is disconnected.

These graphs are representative only, as different relays will have different characteristics. As noted above, I cannot predict what sort of relay you will be able to obtain, but the behaviour can be expected to be similar to that shown. All tests were conducted using a 24V relay, having 10A contacts. Upon contact closure, I also measured 2.5ms of contact bounce. Provided your amplifier is stable by the time the contacts close, this will be completely inaudible.

References

1. D. Self - Muting Relays, Electronics World, Jul 1999

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999 - 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Updates: 18Dec02 - Added update 4./ 10Aug00 - added earthed contacts for extra speaker protection./ 26Oct99 - Modified Figure 1 and added R6 to prevent possible damage to Q3./ 19Oct99 - Added some additional info to prevent power supply confusion./ 06Nov99 - Added info on sulphur tarnishing, and comment that mute timer does not reset quickly./ Page Created and Copyright (c) 16Oct99

Elliott Sound Products **More Reader Feedback**

Some of the responses to the article on bi-amping (edited to protect the writers' privacy, etc). Many of these also qualify as FAQs, since many readers have questions, some of which I can (or will) answer, and others I leave open-ended.

Not all e-mails warrant a response in this page, but all correspondence is answered. Volumes have increased to the point where it is not possible to reprint everything, but if it is likely to be of interest to a fairly wide audience, I will include it eventually.

Please note that reference to any brand of product shall not be regarded as an endorsement or criticism of the product. Where included, such references are from a reader, or mentioned in my reply, and are included only for informational purposes.

If an e-mail from you has already been included and you prefer not to be published, please let me know immediately, and your correspondence will be removed.

Main Index

Contact ESP

Readers' Feedback and Responses

C Comment, **R** Response, **Q** Question, **A** Answer

C Thanks for your comments. Bi-amping is something that not very many people understand, no less the average bride or groom looking for a DJ for their wedding reception. You are exactly correct when you said it doesn't tell the whole story. If we went into the detail that your web site has, people would get completely lost. Most don't understand two speakers versus four, let alone bi-amping.

Bottom line, bi-amping creates a better sound, you know it, I know it and we are one of only a few DJ services in the Twin Cities that does it this way.

C I just read your article on bi-amping, and I'm quite impressed. I don't think you left anything out! I am writing to congratulate you an a particularly easy (and entertaining) to read article. I am a devotee of the bi-amping principle myself, and I am currently building my third set of loudspeakers. This time, it's a no-compromise design. It will, naturally, be bi-amped, though I'm beginning to think that it'll have to be tri-amped.

Q What a fascinating web site! Rarely do I come across folks on the web who know how to articulate and present an argument as well as you have done.

This is not to say that I agree with everything you have said, I am still digesting some of your comments about damping factors and remain a little sceptical about the advantages of bi-wiring but your web site was immediately bookmarked under the "good" list!

In your career have you done any lecturing or other form of teaching?

My own 5.1 HT system at home runs 2nd order crossovers somewhere in the region of 2200Hz and the intelligibility is woeful (not helped by poor choice of operating bands for the drivers). I am going to bite the bullet and do a Bi-Amped system. actually, Tri-Amped, since I already run an active 4th order LP crossover into the sub channel.

I am very interested in hearing more about your idea's on minimum phase active crossovers - any further light you can shed on this? I am an Electronics Engineer by Profession by definitely *not* by experience, so my expertise is a little limited. I work in the Telecommunications industry in management now...

A *I don't recall suggesting that you agree, but thanks for your comments.*

I am (at the time of writing - this has since changed) a "technical training engineer" (telecoms speak for teacher), but I was previously invloved in teaching real electronics and audio.

Regarding bi-wiring - I know I bagged it, but it is still better than running the whole shebang down the one set of cable. It does rely on the amp having a good damping factor, but if you look at the specs closely, you will see that damping factor dies as the frequency increases. This is because of internal frequency limiting in the power amp (required to prevent the amp's phase shift exceeding 360 degrees while it still has gain - Murphy's Law - amplifers will oscillate, and oscillators will amplify).

There are some phase coherent designs, but many rely on subtraction - it works, but gives a lopsided crossover response (e.g. 12db/6db) - these were originally described in various magazines over the years, but they can a lumpy response across the xover frequency.

C Just stumbled across your website. Very interesting, especially since I am currently designing a system that conforms to (most of) your beliefs expressed on the page!

I already built an active filter (simple 2th order Sallen-Key, Butterworth) that works but I'm always looking for improvement! The next version might be implemented on a DSP, if that doesn't bite my budget too much.

Keep up the good work.

Q I have been thinking about implementing active cross-overs in my system and came across your web-site. It makes interesting reading and, as I intended starting with only an active low-pass for bass frequencies, is exactly in line with my intentions. I see that you have designed and built your own active crossover; is your circuit publicly available? My speakers (KEF R105) have a completely separate low-pass filter in the low frequency driver enclosure, and I would like to eliminate this completely but am a bit reluctant to because of the nature of the filter. I believe that the filter is shaping the loudspeaker response as well as limiting its supply of higher-frequency signals. Have you taken this into account? It's an interesting article you've written and I look forward to your updates.

A *As regards active crossovers - go for it. Project 09 is the recommended design. Regarding your speakers, all filters shape the response of subsequent stages, it's a matter of minimising the "damage", and this is where active crossovers can make a huge difference. For starters, you get rid of the large bass frequency transitions from the amp, so intermodulation distortion is reduced - this is one of the main contibutors to a "muddy" sound. Problem is that it gets a bit expensive, but that's hi-fi.*

C Thank you for your very thorough and interesting WEB article "Benefits of Bi-Amping". After reading "Benefits of Bi-Amping", I have implemented the following Bi-Amplification HIFI system to drive a pair of JM Megane loudspeakers.

- Integrated Amp: NAD 317 (with Vampire P#YF RCA Adapter on the PRE-OUT connectors to allow multiple power amplifier connection).
- Second Power Amp: NAD 214 (duplicate of the power amplifier in a NAD 317).
- The JM Megane has two drivers, a 1.5 in. inverted dome tweeter and a 7 in, woofer with dual voice coils configured as a 3-way system with crossover frequencies at 3.5 kHz (18 dB/octave) and 1 kHz (6 dB/octave).

This HIFI system began with the NAD 317 integrated amplifier - It was very acceptable beginning ... a good sounding system. I then implemented Bi-Wiring.

Bi-Wiring produced a noticeable improvement in sound stage and detail. I have learned that Bi-Amplifcation should allow for an even more dramatic improvment in sound because this technique is much more efficient in reducing Intermodulation Distortion (undesirable modulation of the high frequency signal by the low

frequency signal) than simple Bi-Wiring.

So far, I have observed that the Bi-Amplifcation has dramatically improved Bass (2nd Amp??) and additional improvement in sound stage over the Bi-Wire system. Bi-Amplifcation also seems to have changed the character of the sound presentation ...

I think I now know what is meant by sound having a liquid like quality.

R *I'm not at all surprised that you experienced a fairly dramatic change in sound quality, this is exactly what I was waffling on about in the article. As for further improvements, I can only suggest that you keep experimenting - there are so many variables in any amplifier/speaker system that you can probably keep yourself occupied for years!*

One thing to be careful about when bi-amping a two-way:- Make absolutely sure that no DC is present at the amplifier speaker terminals (this can arise when the amp is powered on or off, and is usually only fairly brief). Tweeters are not forgiving of even small amounts of DC, and to be safe you should use a capacitor in series (I can't suggest an accurate value, since I don't know the crossover frequency, but I would think that you would need about 20uF or so as an absolute minimum. This will create its own crossover at about 995Hz (for an 8 ohm speaker) and would be (just) ok if you are crossing over at 5kHz or above. If the value is too low, it will create its own phase problems, and if too high, will allow low frequency "surges" to pass through to the tweeter. Only problem is that suitable high value polyester type caps are expensive !!!

C Thank you for writing an excellent article concerning biamping and putting it on to the web for all to see/understand/enjoy. You have confirmed my belief in biamping as an alternative to standard passive xovering. One of the problems I've seen (as a speaker builder) is the complex passive xovers with all the Zobel junk in them. I've seen speaker designs that the xover cost MORE than either one of the drivers (sometimes more than both the drivers). Seems kinda crazy to me. All that loss with the inductors, etc you get with xovers greater than 6 db/ octave.

I'm currently designing a biamped (and eventually a tri amped) system. I'm trying to fully understand the whole section concerning xover freq. I understand the critical range of human hearing part as an area to try to avoid an xover. I'm planning to use one of Vifa's nice 6 1/2" woofers (P17WJ-00-08) xovered to a Morel MDT30 tweeter. The tweeter is good down to just below 2Khz. The Vifa looks good (flat response) to almost 5Khz. Your recommendations make it sound like I should xover at the higher freq, like 4Khz or even 5khz. The pblm I see with most woofers is the not so good response curves for off axis. The Vifa's published specs show the 30 degree response curve start to break away about 3Khz from the on axis curve. By 5Khz, they are approaching 10 db difference. Therefore, before reading your article, I was planning a 3Khz active xover to reduce the off axis response pblms. But after reading it, I'm now thinking closer to 4Khz. Whatta think?

Hey, I'm on a budget, and can't justify spending the big bucks for an "audiophile" system. I have a friend here in town (XXXX, Texas) that has more money in this system than in his house. (Magnapans, All Mark Levisons, including a pair of the monster Levison amps... I think 33s.. I don't remember off hand). I love listening to his system, but little (currently) cheap system is not that much worse than his. But has given me the upgrade fever. I just what to do it at a reasonable price. I guess it is like wine, once you spend a little money on a bottle on wine, you have good wine, the expensive stuff is only marginally better. Almost an exponential curve.

R *Glad you liked the article. In my opinion, there is really no better way to go, since as you said, all the nonsense you have to go through with passive xovers with all their matching networks etc just does not seem worth the trouble.*

I had a similar question from another reader, and suggested that he would need at least 20uF caps. 20uF will create its own crossover at about 995Hz, and will introduce phase shifts above and below (as always). Only problem is that suitable polyester caps are expensive, but I do believe that using electros is a bad idea in high level signal paths.

I agree with your comments on the Vifa (or any other driver with a cone diameter greater than about 100mm (4")) having to provide decent dispersion at higher frequencies. This has been a major hurdle for me in my quest for the Holy Grail of speakers, and so far I have not solved it either. A vertical array of two 100mm drivers has some appeal, but then you start having impedance problems (4 ohms is not an amplifier friendly impedance. Decisions, decisions.

If you can get to 4kHz xover frequency, this is preferable, but at the expense of HF dispersion and lobing.

With your power supply, consider adding a few extra caps (10,000uF 50V) to ensure minimum hum, and this also provides real current capability in the short term. There are many amps about now which use massive amounts of capacitance for just this purpose, and all the reviews and articles I have read indicate that there is a definite improvement in the sound.

Q I am about to purchase some music equipment (I'm not sure if it qualifies as hi-fi) I bought the loudspeakers first. Contrary to my normal buying habits I didn't do a lot of market research. I listened to a range of speakers in one of the few stores (2 actually, at least that I know of) equipped with hi-fi speakers and ended up with buying a pair of Chario Hiper 1000 towers with the following characteristics:

Low frequency load: vented NRS/2pi
Configuration: 3 way vertical array (with the subwoofer mounted in the bottom of the tower)
Drivers: 130 mm doped paper subwoofer
130 mm doped paper midrange

27 mm textile dome tweeter
Sensitivity: 88 dB SPL 2.83 V/m/W
Frequency @ -3dB: 55 Hz
Acoustic crossover freq.: 135-1350 Hz
Acoustic alignment: LKR4
Rated impedance: 4 Ohm
Suggested amps: rated for 50-120 W/8 Ohm
Cabinet: 20 mm thick MDF (no idea what that means, but I guess it is some kind of pressed wood fiber)

They surely sounded the best of all the speakers available in that price category (about 900 EURO), but the choice was very limited (one or two competing models). Any technical assistance from the sales crew was absent, although they remained friendly and were at least willing to switch back and forth between the different speaker systems while I was testing them.

From the instruction manual that came with the speakers I learned that they were suited for bi-wiring or bi-amping. I tried to learn some more about these possibilities via the net and came across the article of Rod Elliott. The possibilities of bi-amping sound attractive. According to the analysis (or what I understand from it) the basic idea is to connect an electronic crossover and 2 amplifiers.

My questions to you (and/or anybody else who is willing to help me out on this) are:

1) given that (for the time being) I don't want to spent more than ± 1000 EURO on the equipment (speakers not included) is there any gain to expect from bi-amping my system, say using two commercial low- end amps (e.g. SONY TA-FB920R) or even using my old Philips amplifier (1984) (I know that to some (most) of you this probably sounds horrible).

N.B.: The other components of the system will be: Sony MiniDisk MDS-JB920, Sony CD player CDP-XB920 (product codes refer to those from the Sony Europe catalogue). Please also feel free to comment on the choice of Sony. Would Kenwood be a better (less worse) option?

2) if the answer to the above question is a yes, I will then need an electronic crossover. Can you get these in specialized shops? do you have to built these yourself? What does it cost?

3) Should I eventually stick with bi-wiring, do I need some sort of signal splitting device (cf. fig 5 at http://sound.au.com/bi-amp.htm).

4) can you give me a rough idea what I would pay for (average quality) speaker wires.

5) are there any introductory books written on the issues in Hi-Fidellity and that pay much attention to the basic concepts of the underlying electronics.

A *Thanks for your response to my web page. First (and foremost, I guess) has to be a comment about Sony. (... this section has been removed - it is not very complimentary and is based on personal experience ...) Otherwise, as I indicated in my article, I am unwilling to make recommendations (or otherwise) about any specific brand or configuration. However - the speakers you mention sound a little suspicious: Sensitivity is 88 dB SPL 2.83V / 1m*

Since the speakers are rated at 4 ohms, 2.83V equates to 2W, so they have a real sensitivity of 83 dB /W/m (i.e. 85 dB SPL at 1 Watt at 1 metre). This is a fairly low sensitivity, even by current standards, so you will need as many watts as they will take to get a reasonable SPL in your listening room. Also with a -3 dB of 55 Hz, you WILL need a subwoofer to fill in the missing bottom octave. As for bi-wiring versus bi-amping. Bi-wiring will provide some benefits to overall imagery and clarity - although the benefits are fairly subtle. Bi-amping will provide a potentially vast improvement in imagery, perceived sound level, and overall clarity.

I will not (apart from my first-hand comments about Sony) pass judgement about your choice of amps (see my disclaimer in the web page). What I will suggest is that you be prepared to experiment - most of the advances made in audio have been from just that. If you stay with bi-wiring, your speakers are already set up for this (based on your description). To bi-amp, you will need an electronic crossover. There are many manufacturers of these, but you should try to get hold of one which provides phase coherency. Have a further read of my article to see why.

As for whether it is hi-fi or not - who cares! If you like it, and it makes you feel good (along with the music you like), then all requirements are satisfied. I know this is an heretical statement, but it really is up to you. If you enjoy what you hear, that's all that really matters - think about it.

MDF stands for "Medium Density Fibreboard", which is about 700 orders of

magnitude better than "chipboard" - the basic ratty stuff that people make into speaker cabinets, dog houses, junk boxes, etc.

C Thanks for your quick and elaborate reply.

First on how I feel about my system. Well ... after having read that the only components I currently own rate as "suspicious" I think I'll ask myself this question again in the morning.

I got suspicious too, though, after I having found no significant information on this brand of speakers on the net. It's an Italian brand which exists for about 20 years. The company made it's first appearance on the international market in 1986. Their most prestigious line is the Academy series. And I must admit that the more I look at them (I haven't listened to them though) the more I come to appreciate their design. Their walnut veneered finishing goes well with the rest of the furniture. But all this is of minor importance when finally I will be disappointed by the way they sound. But what also matters to me is that the sensitivity of my speakers is what one could expect of speakers within the price category mentioned (I paid them about 900 US$). Anyway, I can't turn them back, so I better learn to love them.

On your calculation. I admit honestly that I forgot all about the basics of electricity (and that for someone who even got a course on MOSFET amplifiers. But that has now been 10 years ago). Their most prestigious speaker has a sensitivity of 91 dB SPL (all the rest being the same). So much for their claim "Acknowledged as the world's best speaker by specialized journalists all over the world". The fact that I "WILL" need a subwoofer sound very depressing since they already have one.

As for bi-amping my system. I will buy a new amplifier anyway, so eventually I just might and try to bi-amp the system using my old Philips amplifier. The only thing is the price of the electronic crossover. It should be worth while buying one in the face of using two fairly low-end amplifiers for bi-amplification. So my basic question to you is, whether the mere splitting of the signal to the different drivers already gives considerable improvement, regardless of the quality of the amplifiers used.

R *To give you an idea, to obtain an average SPL (at 1 metre) of 97dB (which is fairly loud, but not ridiculous), you will need an average amplifier power of about 16 Watts.*

Since this is the average, you have to assume an absolute minimum of 10dB headroom (20dB is better). 10dB above 16 Watts is 160 Watts, so if you have an amp rated at about 100W into 8 ohms, this will give the 160 Watts into 4 ohms that you need - provided the amplifier is able to drive 4 ohms! Have a look at the table below.

This is in agreement with the manufacturer's claim that amps rated up to 120 W into 8 ohms are suitable (such an amp should be able to give 200 Watts into 4 ohms, or thereabouts). Also note that the above average 97dB SPL is at one metre. As you move further away, you lose another 6dB each time the distance doubles until you are into the reverberant field. This is entirely dependent on your listening room, but you could hazard a guess that for an "average" room (whatever that is), you would enter the reverberant field at somewhere between 2 and 4 metres.

As for the quality of the amps - bi-amping will not make bad amps sound good. The overall result may be better than conventional operation (internal passive crossover), but if possible, the better of the amps should be used for the mid-high section. The better this amplifier is, the nicer the combination will sound. Some distortion (harmonic and frequency) in the bottom end will not sound as bad as in the mids and highs, because the low frequency distortion is "masked" by the clean output of the top end, and the woofer will not be able to reproduce the really nasty harmonics very well anyway due to its restricted frequency range (sounds good in theory, at least).

As for the manufacturer's claims - don't dismiss them until you have had a listen (I don't know these speakers, so I cannot comment - not that I would anyway).

Based on the sensitivity of your speakers, the following shows roughly what you can expect:

dB SPL (at 1 metre)	Single Amp (Watts - each channel)	Bi-Amped (per amplifier)
85	1	0.25
88	2	0.5
91	4	1
94	8	2
97	16	4
100	32	8
103	64	16
106	128	32
109	256	64
112	512	128

This assumes that the crossover is at the "equal power distribution" frequency of about 280 to 300Hz. You will need to experiment to see if your internal sub-woofer can get that high without sounding horrible if you want to get the maximum benefit of the increased SPL due to frequency splitting with an electronic crossover.

This also assumes that you want (need?) high SPL. If you don't, then you will probably not have an issue with power, so can have more flexibility with crossover frequencies.

Q Your article is excellent- very easy to read yet informative! I graduated from the University of Idaho ten years ago - majored in Electrical Engineering. I am a power engineer and am a little nervous about tweaking with my system. I work with kilo-volts not milli-volts. I have Infinity 8A Kappa speakers, 1 set of Adcom GFA-565 mono amps and an Adcom GFP-565 pre-amp.

I currently have the set bi-wired. I noticed a definite improvement after bi-wiring the system. Saw a bigger sound improvement after replacing my 10-year old Denon CD player with a good transport and D/A converter. I am now kicking around the idea of bi-amplifying my set. If I understand correctly, I would have to remove the passive low-pass crossover in series with the woofer circuit and remove the passive high-pass circuit in series with my mid-base coupler, mid-range and tweeters. Then I would install an electronic crossover to separate the signals to the high and low frequency amplifiers. This is correct?

The speaker manufacturer's literature claims that the woofers passive crossover is tuned to extend the bass response by approx. 1/2 octave to 33Hz. Wouldn't I lose the lower frequencies by removing the passive crossover and using an electronic crossover in its place? I know you don't get something for nothing. What is the downside of having the woofer circuit tuned to extend the low end?

Does it make any sense to install an electronic crossover to split the signals sent to each amp and then keep the passive crossover circuits in the speakers? I expect this would probably cause to much of a roll-off near the crossover frequencies.

I appreciate any response you might have. My wife is particularly nervous about me pulling the speakers apart and snipping inductor coils and capacitors out of the circuits. "Don't you think they knew what they were doing when they designed those speakers???!!!"

I know you don't give advice on equipment brands. If this were your system, however, would you consider upgrading to a bi-amplified system or upgrading one of the other components first?

A *I'm glad you enjoyed my article. You are absolutely right about not getting something for nothing! The bass extension filter almost certainly reduces the effeciency of the woofer at higher frequencies, so it can "give it back" when it is needed in the last 1/2 octave. If you were to biamp the system, this part of the crossover circuit would need to be retained, or a matching bass boost could be done electronically (which is better, but requires more circuitry and effort, and also means that you must know the exact tuning frequency used, and the amount of loss introduced in the higher frequencies).*

All the above is possible, but will take some research and maybe a bit of lateral thinking! A signal generator and oscilloscope or good multimeter would not go astray if you can lay your hands on them - this way you can plot the response of the existing crossover unit - remember to terminate the crossover outputs with a load resistor equal to the loudspeaker impedance, or the results will be meaningless.

If you leave the speakers connected, speaker and cabinet resonances will make meaningful measurements more difficult. If you want to do it this way, make sure that the box sealing is not compromised, and that you drive the system with a power amp to ensure a constant-voltage source. A signal level of 1 or 2 volts is sufficient, otherwise your ears will hate you!

You are also right about having to disable the existing crossover from bass to mid +high. If this is left in place the effect will not be a happy one! Remember that the mid to high crossover must be retained unless you are contemplating tri-amping. Too much too soon I would think.

As for your wife's consternation, she is right, but with a caveat ....

The speakers - like all speakers - were manufactured to suit the majority market in their price range. Few people will jump at the chance to purchase an expensive pair of speakers knowing that they will also have to buy an electronic crossover and another power amplifier to make it work. As a result, the final design is a compromise between simple economics and harsh reality.

Biamping WILL create a dramatic improvement, but it must be done with great care and attention to details, or a disaster will surely be the result. I would not recommend attacking the existing low to mid+high crossover with anything other than a screwdriver (for its removal intact) - this way you can always revert to the original if it doesn't work out, or while you perfect the next stage of your project.

(After all, who wants to be without their sounds?)

If this were my system I would biamp first, then experiment with the result until I was satisfied that it could be improved no more. Only then would I look at the other components (which sound as if you have a reasonable investment in them). I hope this helps.

Q I enjoyed your article very much. I am interested in using a Bryston 3B for the bottom end and a single ended triode amp for the mids and upper. can this setup be made to work?

A *I'm glad you enjoyed the article. I don't know the Bryston 3B, but I'm sure that the combo will work as long as the crossover point is selected to get equal power distribution. You might find that this will require a crossover in the middle of the "intelligence band", but if well done this can still work - thousands of speaker manufacturers do it all the time, so why not you too.*

Q I have acquired a QUAD ELS-57 speakers, that I like very much. Unfortunately, a while ago my old amplifier stopped functioning, and I was advised to buy a new one, instead of paying for a repair. I have searched on Deja-vu News, and found that people suggest certain amplifiers as more appropriate for the QUADs then others. (Bedini 25/25, Spectral DMA-XX were some examples.)

Given the fact, that the input impedance of the QUADs is highly capacitive, and ranging from 2 to 34 ohms, I believe that some amplifiers are more suited than others. What I also noticed was that some, apparently unrelated, people recommended a "fast," "class A," "low power" amplifiers. Do those modifiers have any substance or are they only buzzwords?

Given the price of these amplifiers, is there a DIY amplifier I could attempt to build? Thank you very much.

A *Congratulations on your ownership of the QUAD speakers. They are one of the all time classics, but they are hard to drive. Most amplifiers do not like the load presented by electrostatics or anything else which is highly reactive, so I'm afraid that you really do need something a little different in the amplifier department.*

As for the "recommendations", they are all meaningful in one way or another, but are completely unrelated to your needs - "fast" is not a requirement, nor is Class-A or low power. If I remember correctly, the original QUAD amp was about 50 or 60W (???), but the biggest problem is finding an amp that will drive the capacitive load without a mental breakdown.

I would tend to suggest an amp which is capable of driving a two ohm load (these are not as common as amps which can handle 4 ohms) since such an amp will probably be also capable of driving highly reactive loads such as the ESLs.

The suggestions you mentioned may well be good, as no doubt they have been tried and tested with the ELS series speakers. Having said that, I cannot make a specific recommendation - although P101 has been tested with good results.

Valve (vacuum tube) amps are usually much happier about driving really difficult loads without undue distress, but I have heard of tests conducted on high power PA amps (Australian Monitor), when they were shown to be completely stable driving a 5uF load. Not a recommendation, but it shows that "solid-state" amps are quite capable of driving strange load impedances.

You did not say what sort of amp you were using - and did it die while driving the ESLs? If so, this shows that the load can indeed "blow up" otherwise good amps. If it is a valve amp - get it fixed. Anyone who tells you not to bother has never heard one. My project pages also have many amps, but other than P101 I have never tested them with electrostatics.

Q Hi, enjoyed your web site. i have built jlh 10 watter using 2n3055 &tip35 but as yet have done no subjective comparison. may also try mjl21194 (bit costly in australia). what do you think

A *Did you build the amp from the plans on my site? (I guess you probably did, or you would not be responding to me ...). I know the design is old, but if you find that you like it I'd really like to hear more (I re-published the design, but have not built one yet). A bit of good feedback is always nice, and if you don't mind I will include your comments after a comparison in the Readers' Page.*

*Using the MJL21194 is probably overkill, but these are one of the most linear transistors on the market - as such I would expect the performance to be about as good as it can be. But yes - they are *expensive*.*

Glad you enjoy my site - visit often, there are new things appearing all the time.

Q HI, JUST RECENTLY I PURCHASED A SET OF B&W SPEAKER (DM-603) AND CONNECTED THEM TO MY ONKYO M-508 POWER AMP (2 CHANNEL). I'M A NEWBIE IN HI-FI HOWEVER I'VE LEARNED QUITE A BIT BY READING MAGAZINES AND ARTICLES AND ESPECIALLY YOUR WEBSITE REGARDING BI-AMPING. I'M STILL CONFUSED THOUGH. IS IT POSSIBLE TO BI-AMP MY DM-603 USING ONLY ONE POWER AMP? CAN I USE SPEAKER "A" FOR HI/MIDS, AND SPEAKER "B" FOR LOWS INSTEAD OF BUYING ANOTHER POWER AMP? YOUR ASSISTANCE IS GREATLY APPRECIATED. I BELIEVE IN YOU.

A No, you cannot bi-amp with only one amp. You can bi-wire, but the results are nowhere near as good. I would strongly suggest that you live with your system for a while yet, because bi-amping is NOT trivial - it is possible to damage speakers (or amps) if you are not completely confident, and are experimenting.

Experimenting is good, but please do it with a cheap speaker system and an amplifier you can afford to lose.

Q Was wondering if you had anymore information about designing Capacitance Multiplier power supplies. I am planning on building a pair of Leach amps and am looking at various power supplies. Supply is +/- 57V with average current of around 5 Amps, peak as high as I can reasonably get (The amp itself has current limiting). Any help,information, or recommended reading you could provide would be greatly appreciated.

A *I assume that you read the article in my project pages, and this should have enough info for you to do your own design. Most of it is pretty straightforward, and requires not much more than ohm's law and a bit of calculation of capacitive reactance.*

Have a thorough read of the article, and e-mail me with any specific questions. I don't know of any other sites offering more info than I have.

Q Its been quite awhile since I've done any analogue design, so I'm a but rusty. I assume for increased current the main thing would be to ensure the driver transistor and diode can handle the increased load. Any suggestions for suitable transistors??

The only thing I'm having trouble wrapping my head around is modifying this for the higher voltage. By the looks of it the 200R and caps remain the same and its just a matter of adjusting the 12k resistor. Do I want a 10V drop between the voltage rail and the driver base, or a 10V drop between the base and ground?? Of course I could be completely wrong :o)

A *A re-think on your problem: Since you are running a fairly high voltage and current, you will need some fairly "beefy" devices. However for the Leach amp, I think that anything other than a well filtered supply is not needed - 10,000uF caps are not cheap, but you will need close to that anyway before the filter.*

The Leach amp (like most modern designs) will have quite good power supply rejection, and I would only consider using a capacitance multiplier where hum becomes audible.

I have run some tests on my own power amp, and even with quite high ripple on the DC supply, this does not make itself evident until the amp clips. Below this, my

distortion meter shows zero 100Hz supply ripple at any power level. I doubt that the Leach will be any different. Save your self some money (and energy) and try it out with a simple supply first. If you are not happy, then look at using additional filtering - I bet you will never bother.

If you want to do it anyway, I would suggest MJ15003/4 for the power transistors (250W) and MJE340/350 for the drivers. The rest of the circuit should work as shown, but you might need to use higher wattage resistors. The 10V drop is between the input and output of the supply, and the idea is to ensure that this voltage is higher than any momentary droops (including ripple) as the amp is driven. This is adjusted with the 12k resistor as you guessed. The diodes do not need to carry the current - they are there to stop reverse voltage if the main filter cap manages to discharge too quickly. They don't need to be changed.

C Vulnerability of Tweeter to DC Transient:

There used to an organisation in the UK called the Active Loudspeaker Systems Organisation (ALSO) which had set a standard for active crossovers, active speakers and amplifiers manufactured for bi-amping. The active speakers are without any passive crossover at all to protect the tweeters from DC transcient. Basically, when complying with ALSO standard, there should be no danger having the tweeters blown out.

I am still using the following antiquated British system which are ALSO compliance:
Active Crossover and power amplifier for one channel:
A&R Cambridge SA60X
Control and power amplier for another channel:
A&R Cambridge A60P
Active speakers:
ARC 050 two way speakers with detachable passive crossovers

A&R Cambridge is now Arcam of UK, and ARC is now defunct. I am not sure if ALSO is still active.

R *I did not know of the ALSO organisation, but my suggestion to use protection was based on the fact that most users will not even be aware that a DC "transient" is possible. Many amps produce small transients at power on and off, and these CAN damage tweeters. I prefer to err on the side of safety.*

Q I'm designing a subwoofer module, including an electronic crossover. Could you tell me how to take a speaker level input and have that signal go through an electronic crossover? Can I use a voltage divider or is there another way to attenuate the high level signal?

A *Generally a simple voltage divider is all you need. This can be set up so it will sum the outputs of left and right amps, and give you mono line level for the sub amp.*

If you have (say) 100W / 8 Ohm amps, the maximum voltage is 28V RMS, so you need a divider of about 28:1 - since a lot of bass info is mono anyway, you may need more (you did not say if your sub amp or crossover has a level contol - you will need one). Design of voltage dividers is covered in an article on the ESP site.

Q Your projects page is just what I've been looking for! Do you have PCB's for projects #08 and #09 ? I logged on to Microsofts web page, trying to locate the VB40032.DLL file so that I can use your Linkwitz-Riley Component Calculator Program. However I couldn't find any information on the file. Please, tell me how I can find it. With your help, " My Rig " can sound much better. Thanks.

A *Many project PCBs are available (although not at the time of writing).*

The VB40032.DLL is available from Microsoft

Q Congratulations on a very good page. undoubtably one of the better presented and more informative on the web. <end gratuitous sycophancy>

I was considering building your project 2 - the preamp and just had a quick question. I wish to eliminate the balance control from the design. As I understnd it, this pot seem to add some negative 'feedforward' which reduces the gain for a given channel. If I simply remove this component and sever the link from the balance pot(s) between the two channels will the circuit stillfunction correctly?

Thanks in advance, and keep up the great work!

A *If you don't want the balance pot, just leave it out of the circuit. You will have a little more gain (about 3 dB), but need do nothing else. By the way, I don't really recommend leaving it out, as it can be useful. If you are sure - no problem. There is no feedforward or other nasty stuff involved - it is simply a passive control.*

Q Finally just what I've been looking for! A simple, but pratical amp for my turntable and well under $2,000! Tubes are nice but I already have a tube amp with mmphono inputs! I seem to recall seeing something about circuit boards and or parts mentioned on one of your pages. I am interested in a parts pcb package, minus the heavy stuff. Thank you for the excellant info! well I see that you are away for a few weeks, looking forward to hearing from you, I think I'll put my poor bloodshot eyes to bed!

A *At the moment, I am not in a position to provide kits in any way shape or form (sorry). I am looking into this, and will do so if there is enough interest (at the moment, there is not, unfortunately).*

Which amp were you looking at? I am currently examining the possibilities of a new amp (based on the 60W design) which should be capable of 100 to 150W into 8 Ohms. It's only on the drawing board at present, but should be quite nice when completed. (This is the P3A amp, which has been available for some time now.)

Q Hi, i read your article on bi-wiring/amping and am very interested in it. i was thinking of getting the paradigm moniter 7's or 9's, which can be bi-wired. i have a yamaha r-v1105 reciever and an old denon avr-1200 reciever. the yamaha r-v1105 has preouts, which i can connect to the denon avr-1200. now, is this and some speaker wire all that i need to biamp????

so now am i right in assuming that the volume dial on one of the recievers would control treble, and the other the bass???

i am really unsure if i know what i am talking about so i would appreciate any information. thankyou very much.

A *Be very careful here. The concepts of bi-amping and bi-wiring are very different. The bi-amp approach requires that you do not use the crossovers in the speaker cabinets at all - you need an electronic crossover. It is possible to bi-wire using separate amps, and there might be some advantages over "simple" bi-wiring, but you will not get the full benefit without the electronic crossover.*

*The whole idea is to make one amp and speaker combination responsible *only* for the bass section of the frequency range, and the other only handles the mid +high component. I suggest that you re-read the article very carefully - there are lots of traps for the unwary! If you do it this way, then what you suggest is correct. This is not really bi-amping though, since all frequencies are still handled by each amp. You may get marginally less amplifier intermodulation distortion, but will not gain the real advantage of true bi-amping.*

C First of all, great site! While I really haven't got much experience with bi-wiring, I do love to bi-amp. I build my own speakers and basically am put off by the cost and technical complexity of low frequency passive crossovers.

All the info you've provided is wonderful. I'm thinking of building that phono pre-amp (if I ever aquire the requisite technical expertise to do so).

Here's a couple of jokes for your humour section (maybe?):
A conversation between two audio research technicians (the field of study, not the company):
Tech 1: Gee, this new horn speaker is efficient. At what db level does it become dangerous?
Tech 2: What?

Okay, that was bad. But the following is a play on words (or a saying) and technically accurate.
Q: What did the chemist say when his assistant fell into a vat of acid?
A: Apparently, the problem is, you are part of the solution.

Well, thanks for a great site. I thoroughly enjoyed the visit.

R *Thanks for the vote of confidence (and the jokes).*

It's not only the complexity of passive crossovers, you are just losing so much other good stuff - power, clarity, phase (etc, etc ....)

Q I've been combing the web, looking for an answer, but just can't find anything. Your page appears to be the best on just about everything else, so I guess I'll check if you know.

Basically, I want to construct a Hafler Dynaquad style surround matrix processor. The catch is, I don't want to do it at speaker level. I'd like to do it at line level. I have enough amplification available, and I figured it wouldn't be a difficult item to construct.

The problem is, I can't find any info to guide me. There was, within the past five or so years, a product called the Phasearound, which operated on this principle. According to a review I found, it used an audio transformer to derive the center and surround channels, but the article didn't elaborate beyond that. If I knew more, that would've been enough. I don't need a center channel, as I intend to use it in an audio only system. Any help you might be able to offer would be appreciated.

A *I think I know exactly what you are looking for. The circuit is easy to build (it was also published in an Australian electronics magazine very recently), and I will look at making this a new project within the next month or so (if you can wait that long).*

The magazine article also describes a digital delay - this apparently enhances the "surround" effect quite a bit, by delaying the rear channels. I will not be including this in my project (or I might, I haven't decided yet). (See projects index for details.)

Q I was searching through altavista and found your site, which is definitely the best among the many I've found. I'm not a analog engineer, so there are many things I don't understand, but I find your site very informative. I just have one question. It might have been answered in your site already, but after reading it over and over, I couldn't come up with a definite answer.

Would bi-amping without an active crossover be possible? (source - 2 amps - passive crossover built into the speaker which is bi-wireable) and would it help?

It'd be like a "temporary" path between bi-wiring and true bi-amping. I'm not sure if the HI and LO input of the speaker are truly independent or if there are some interactions between them.

BTW, I have Magnepan 1.6s, which are quite obscenely (what a word!) inefficient at 83dB, 1W, 1m and I find my Bryston 3B-ST struggling (or so I feel anyway). Thanks for your time.

A *You are right - what you are looking for is not in there (at the time - the information has been there for some time now).*

*The approach you are asking about has been suggested by a few other readers as well. It is not really bi-amping in the true sense of the term, because the amps still have to reproduce the full frequency range - its just that there is no actual power being produced at the out of band frequencies. There is probably some merit to doing this, but you will *not* get the power advantage, which is important with really inefficient speakers.*

Q Thank you very much for your informative article on Amplifiers. I am an Engineering student in Washington. Over the next few weeks I will have a great deal of time to experiment and would like to built an Amp. I was wondering if you have a complete schematic of your design as it seemed the design was complete but in pieces that were spread across the document.

I am also interested in whether your design could be applied multiple times off of one power supply as I would like to build a 6 channel Amp for Dolby Digital and THX. If you have the information available what is the approximate cost of all of the components for your Amp design. Again thank you for taking the time to write your guide.

A *Thanks for the nice words. Which article were you interested in? There is one that describes the general design goals, but that is fragmented (as you said). If you want complete schematics, you need to look in my projects pages, since this is where all the complete constructional articles live.*

I suspect that this is not where you were looking, since these are complete in all respects.

*Any number of amplifiers can be run from the one supply (with some interactions of course), but as you do need to ensure that the transformer, rectifier and filter caps are sufficient for the total loading. One trick used by a lot of manufacturers is to use a *really* big tranny, and then use separate rectifiers and filter caps for each amp.*

To give you an idea, using 5 x 60W amps needs a transformer of about 500VA for continuous operation at full power (I would not suggest anything less than this). The filter caps for each amp should be at least 4,700uF for +ve and -ve (i.e. 2 of them) for each amp (10,000uF is better, but expensive). Rectifiers should be at least 10A continuous rating - most manufacturers use less that the above, but you get better reliability when you use bigger rectifiers and caps. I hope you have some money, 'cause this will not be cheap! (It will be lots of fun, though.)

Main Index

Contact ESP

Copyright of this article is the property of Mr. Linsley Hood and Electronics World (formerly Wireless World). It is reprinted here as a reader service, and ESP claims no intellectual rights whatsoever except for the editorial comments. It is reproduced using the original text (or as much as I have been able to acquire), and the descriptions are those of the author (excluding editors notes).

It should be noted that the article was originally published sometime in 1969, and that the transistors are now obsolete. Much of the descriptive text is no longer valid for new designs, and the comments on Class-AB amplifiers may not apply today.

Simple Class A Amplifier

A 10-W Design giving subjectively better results than class B transistor amplifiers

by J. L. Linsley Hood, M.I.E.E.

*** Editors Notes by Rod Elliott**

During the past few years a number of excellent designs have been published for domestic audio amplifiers. However, some of these designs are now rendered obsolescent by changes in the availability of components, and others intended to provide levels of power output which are in excess of the requirements of a normal living room. Also, most designs have tended to be rather complex.

In the circumstances it seemed worth while to consider just how simple a design could be made which would give adequate output power together with a standard of performance which was beyond reproach, and this study has resulted in the present design.

Output power and distortion

In view of the enormous popularity of the Mullard "5-10" valve amplifier, it appeared that a 10-watt output would be adequate for normal use; indeed when two such amplifiers are used as a stereo pair, the total sound output at full power can be quite astonishing using reasonably sensitive speakers.

> * For today's speakers and expectations, this is clearly not the case. 10 Watts is likely to be sufficient for tweeters in a tri-amped system however, and this is the reason for publication of this circuit.

The original harmonic distortion standards for audio were laid down by D. T. Williamson in a series of articles published in Wireless World in 1947 and 1949; and the standard, proposed by him, for less than 0.1% total harmonic distortion at full rated power output, has been generally accepted as the target figure for high-quality audio power amplifiers. Since the main problem in the design of valve audio amplifiers lies in the difficulty in obtaining adequate performance from the output transformer, and since modern transistor circuit techniques allow the design of power amplifiers without output transformers, it seemed feasible to aim at a somewhat higher standard, 0.05% total harmonic distortion at full output power over the range 30Hz-20kHz. This also implies that the output power will be constant over this frequency range.

Circuit design

The first amplifier circuit of which the author is aware in which a transformerless transistor design was used to give a standard of performance approaching that of the "Williamson" amplifier, was that published in Wireless World in 1961 by Tobey and Dinsdale. This employed a class B output stage, with a series connected transistors in quasi-complementary symmetry. Subsequent high-quality transistor power amplifiers have largely tended to follow the design principles outlined in this article.

The major advantage of amplifiers of this type is that the normal static power dissipation is very low, and the overall power-conversion efficiency is high. Unfortunately there are also some inherent disadvantages due to the intrinsic dissimilarity in the response of the two halves of the push pull pair (if complementary transistors are used in asymmetrical circuit arrangement) together with some cross-over distortion due to the I c /V b characteristics. Much has been done, particularly by Bailey, to minimise the latter.

An additional characteristic of the class B output stage is that the current demand of the output transistors increases with the output signal, and this may reduce the output voltage and worsen the smoothing of the power supply, unless this is well designed. Also, because of the increase in current drive with output power, it is possible for a transient overload to drive the output transistors into a condition of thermal runaway, particularly with reactive loads, unless suitable protective circuitry is employed. These requirements have combined to increase the complexity of the circuit arrangement, and a well designed low- distortion class B power amplifier is no longer a simple or inexpensive thing to construct.

> * The thermal runaway referred to is now known to be secondary breakdown, where the transistor suffers from a localised heating on the silicon die. This effect is very rapid, and can lead to almost instantaneous destruction of a transistor. This is one reason that MOSFETs are preferred by many amplifiers designers (the editor is not one of these!).

An alternative approach to the design of a transistor power amplifier combining good performance with simple construction is to use the output transistors in a class A configuration. This avoids the problems of asymmetry in quasi- complementary circuitry, thermal runaway on transient overload, crossover distortion and signal-dependent variations in power supply current demand. It is, however less efficient than a class

B circuit, and the output transistors must be mounted on large heat sinks.

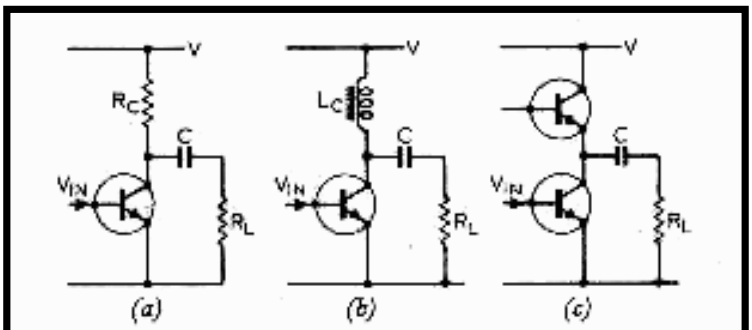


Fig. 1. Basic class A circuits using (a) load resistor R_c giving power conversion efficiency of about 12%, (b) l.f. choke giving better efficiency but being bulky and expensive, and (c) a second transistor as collector load.

The basic class A construction consists of a single transistor with a suitable collector load. the use of a resistor, as in Fig 1(a), would be a practical solution, but the best power-conversion efficiency would be about 12%. An l.F. choke, as shown in Fig1(b), would give much better efficiency, but a properly designed component would be bulky and expensive, and remove many of the advantages of a transformerless design. The use of a second, similar, transistor as a collector load, as shown in Fig 1(c), would be more convenient in terms of size and cost, and would allow the load to be driven effectively in push-pull if the inputs of the two transistors were of suitable magnitude and opposite in phase. This requirement can be achieved if the driver transistor is connected as shown in Fig. 2.

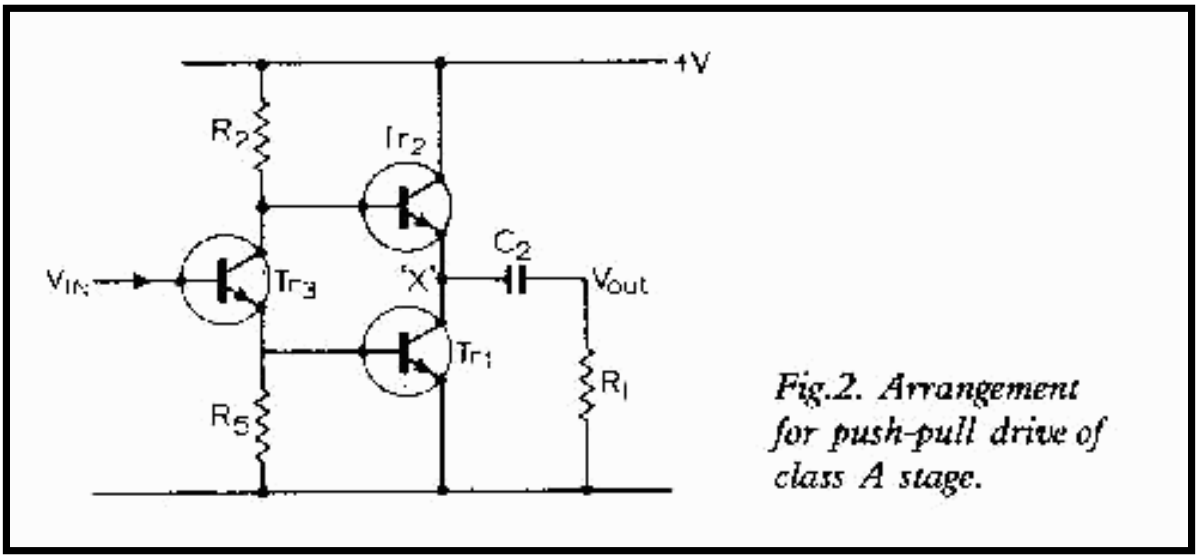


Fig.2. Arrangement for push-pull drive of class A stage.

This method of connection also meets one of the most important requirements of a low distortion amplifier :- that the basic linearity of the amplifier should be good, even in the absence of feedback. Several factors contribute to this. There is the tendency of the Ic / Vb non-linearity of the characteristics of the output transistors to cancel, because during the part of the cycle in which one transistor is approaching cut-off the other is turned full on. There is a measure of internal feedback around the loop Tr1 Tr2 Tr3 because of the effect which the base impedance characteristics of Tr1 have on the output current of Tr3. Also, the driver transistor Tr3, which has to deliver a large voltage swing, is operated under conditions which favour low harmonic distortion :- low output load impedance, high input impedance.

* A potentially worthwhile improvement to this circuit is the addition of a 0.1 ohm resistor in the emitter circuit of Tr1. This applies local feedback to the entire gain stage, providing a significant reduction in distortion. If used, this should be a 5 Watt wirewound type to handle the current.

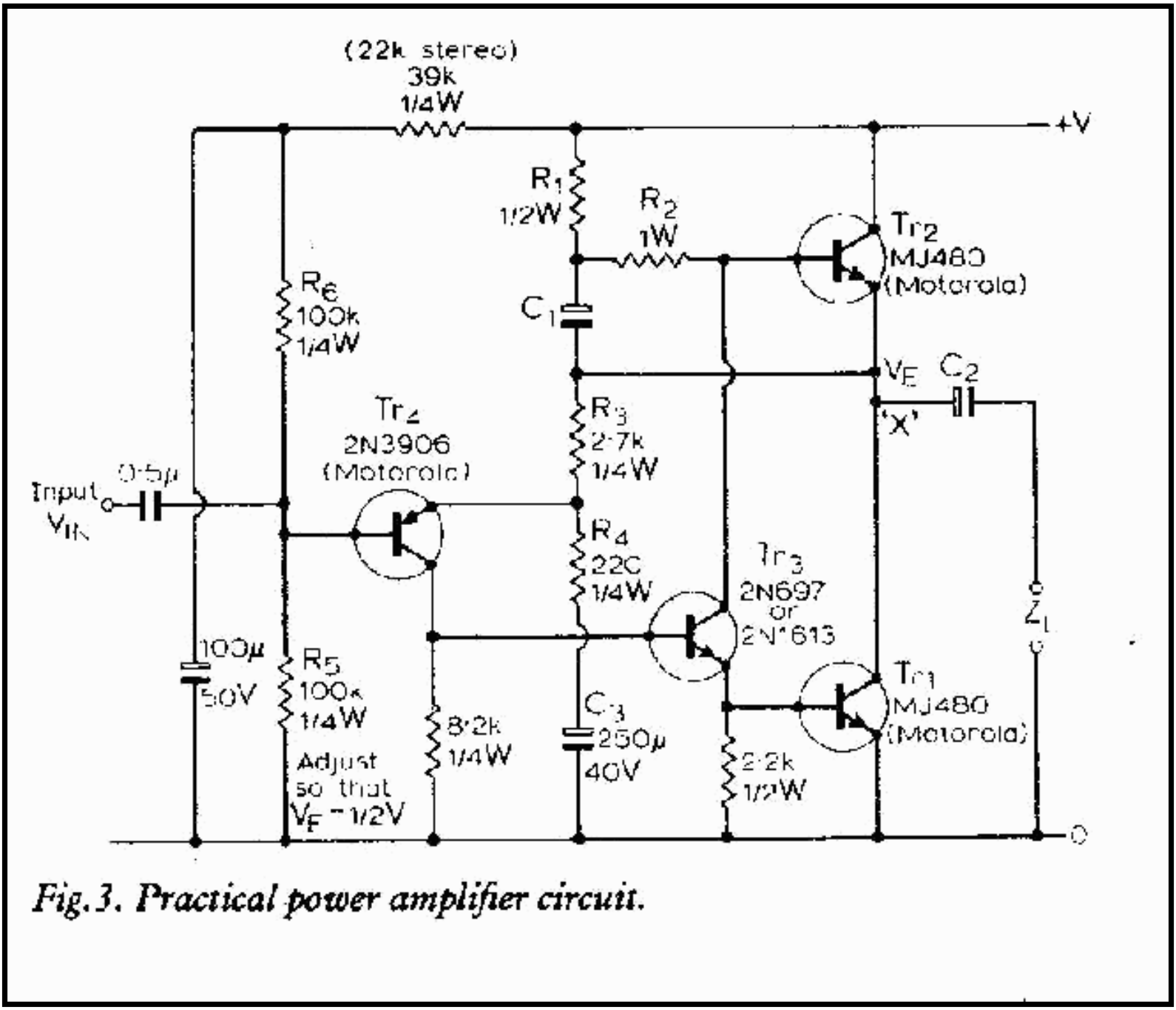


Fig. 3. Practical power amplifier circuit.

A practical power amplifier circuit using this type of output stage is shown in Fig. 3.

* The upper transistor (Tr2) is operating as a current source, whose output current is modulated. This allows the circuit to operate at about half the quiescent current that would be required if no modulation were applied. The values for R1 and R2 must be selected, based on the gain of Tr3. For a 40 Volt supply, if Tr3 were to have a gain of 50 at 1A, then ...

(R1 + R2) = 20V / 20mA (base current) = 1000 ohms.

One problem with this approach is that the current provided by Tr2 will vary with temperature. Readers wishing to experiment with this circuit should ensure that the current is checked at normal operating

temperature (i.e. HOT). There is no mechanism in the circuit to prevent thermal runaway, other than the use of a suitably large heatsink. At some point, the circuit should stabilise the quiescent current. If it does not (and the current continues to increase), then the heatsink is too small. To ensure a useful life for the transistors, they should not operate at greater than 50º C, which in normal conditions should be quite achievable. Since each transistor operates at (or near) 25 Watts, the heatsink for each transistor should have a thermal capacity of about 1º C / Watt. A better (i.e. larger) heatsink will do absolutely no harm, and will ensure freedom from thermal runaway.

There is also a newer version of this amp, but I have no plans to re-publish at any time.

The open loop gain of the circuit is approximately 600 with typical transistors. The closed loop gain is determined, at frequencies high enough for the impedance of C3 to be small in comparison to R4, by the ratio (R3 +R4)/R4. With the values indicated in Fig. 3, this is 13. This gives a feedback factor of about 160 milliohms.

Since the circuit has unity gain at D.C., because of the inclusion of C3 in the feedback loop, the output voltage Ve, is held at the same potential as the base of Tr4 plus the base emitter potential of Tr4 and the potential drop along R3 due to the emitter current of this transistor. Since the output transistor Tr1 will turn on as much current as is necessary to pull Ve down to this value, The resistor R2, which together with R1 controls the collector current of Tr2, can be used to set the static current of the amplifier output stages. It will also be apparent that Ve can be set to any desired value by small adjustments to R5 and R6. The optimum performance will be obtained when this is equal to half the supply voltage. (half a volt or so either way will make only a small difference to the maximum output power obtainable, and to the other characteristics of this amplifier, so there is no need for great precision in setting this.)

* Not mentioned is the purpose of C1 (in conjunction with R1 and R2). This capacitor provides "bootstrapping", which attempts to maintain a constant voltage across R2. If the voltage remains constant across a resistor, it follows that the current through the resistor must also remain constant. The performance of this circuit will be severely impaired if the value of C1 is too small - based upon the lowest frequency of operation, and the parallel value of R1 and R2. For operation down to 20Hz (assuming R1 + R2 = 1000 ohms), the capacitor should be at least 220uF.

Likewise, the reactance of C1 must be low with respect to the speaker impedance (preferably less than 1/2 of the speaker impedance at the lowest frequency of interest - 20Hz is assumed). This works out to be about 2,000uF. A working voltage of not less than 50V is suggested for all electrolytic capacitors, and for optimum h.f. performance, a 1uF polyester may be paralleled with each electro. In my experience this is not needed, but many will disagree, so if you want it, add it.

Silicon planar transistors are used throughout, and this gives good thermal stability and a low noise level. Also, since there is no requirement for complementary symmetry, all the power stages can use n-p-n transistors which offer, in silicon, the best performance and lowest cost. The overall performance at an output level of 10 watts, or at any lower level, more than meets the standards laid down by Williamson. The power output and gain/frequency graphs are shown in Figs. 4 and 5, and the relationship between output power and total harmonic distortion is shown in Fig. 6. Since the amplifier is a straight-forward class A circuit, the distortion decreases linearly with output voltage. (This would not necessarily be the case in a class B system if any significant amount of cross- over distortion was present.) The analysis of distortion components at levels of order of 0.05% is difficult, but it appears that the residual distortion below the level at which clipping begins is predominantly second harmonic.

Stability, power output and load impedance

Silicon planar NPN transistors have in general, excellent high frequency characteristics, and these contribute to the very good stability of the amplifier with reactive loads. The author has not yet found a combination of L and C which makes the system unstable, although the system will readily become oscillatory with an inductive load if R3 is shunted by a small capacitor to cause roll-off at high frequencies.

The circuit shown in Fig. 3 may be used, with very little modification to the component values, to drive load impedances in the range 3-15 ohms. However, the chosen output power is represented by a different current/voltage relationship in each case, and the current through the output transistors and the output voltage swing will therefore be different. The peak-voltage swing and mean output current can be calculated quite simply from the well-known relationship $W=I^2R$ and $V=IR$, where the symbols have their customary significance. (it should be remembered, however, that the calculation of output power is based on RMS values of current and voltage, that these must be multiplied by 1.41 to obtain peak values, and that the voltage swing measured is the peak to peak voltage, which is twice the peak value.)

When these calculations have been made, the peak-to-peak voltage swing for 10 watts power into a 15-Ohm load is found to be 34.8 volts. Since the two output transistors bottom at about 0.6 volts each, the power supply must provide a minimum of 36 volts in order to supply this output. For loads of 8 and 3 ohms, the minimum h.t. line voltage must be 27V and 17V respectively. The necessary minimum currents are 0.9, 1.2 and 2.0 amps. Suggested component values for operation with these load impedances are shown in table 1. C3 and C1 together influence the voltage and power roll-off at low audio frequency performance is desired than that shown in figs. 4 and 5.

> * I suggest a power supply voltage of 40 Volts, which will actually allow close to 20 Watts into an 8 ohm load. Make sure that the heatsinks are sufficient to handle the additional power dissipation caused by the higher supply voltage. Those suggested in the construction notes are probably too small for operation at this supply voltage.

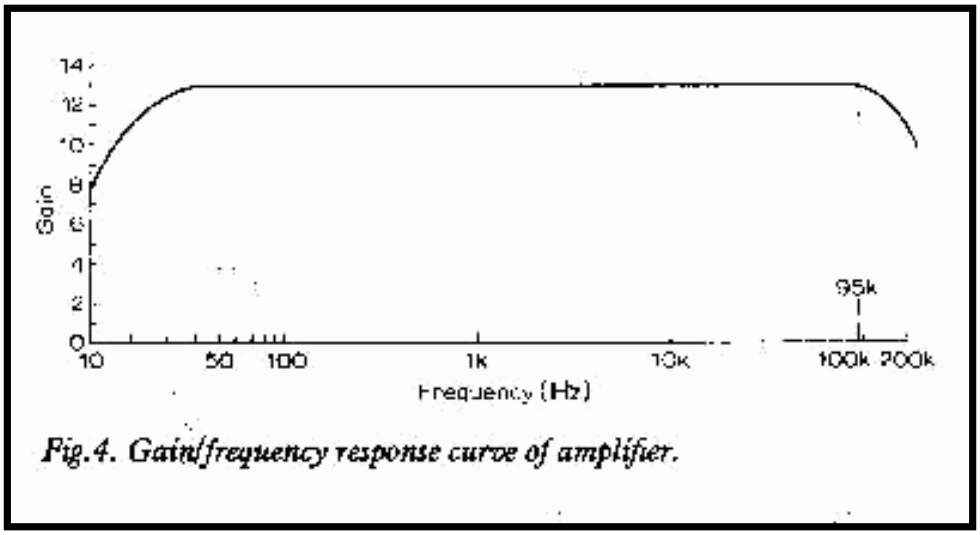


Fig.4. Gain/frequency response curve of amplifier.

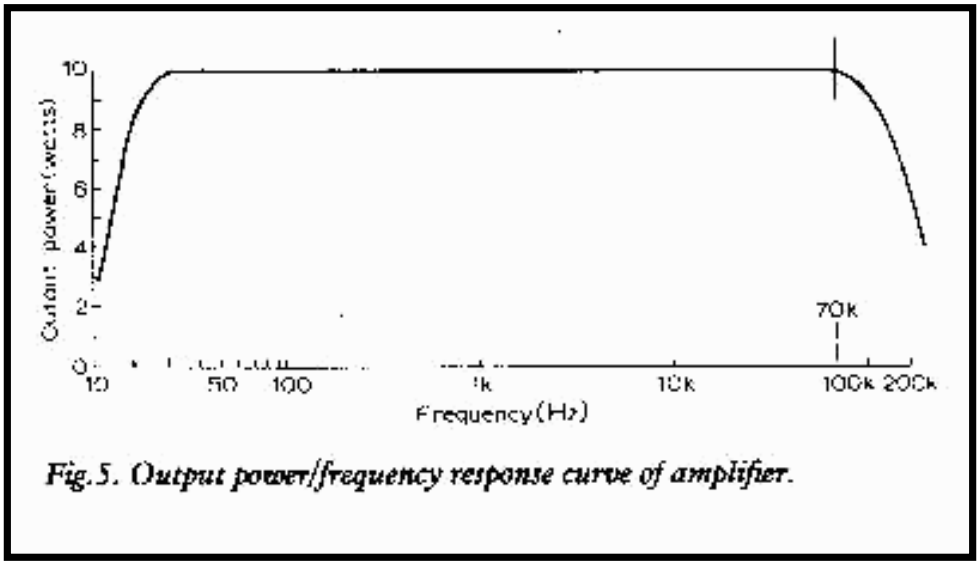


Fig.5. Output power/frequency response curve of amplifier.

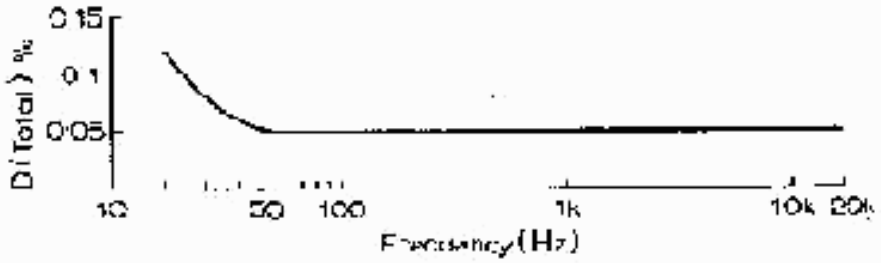


Fig.6. Distortion/frequency curve at 9W.

Since the supply voltages and output currents involved lead to dissipation in the order of 17 watts in each output transistor, and since it is undesirable (for component longevity) to permit high operating temperatures, adequate heat sink area must be provided for each transistor. A pair of separately mounted 125mm by 100mm (5" by 4") finned heatsinks is suggested. This is, unfortunately, the penalty which must be paid for class A operation. For supplies above 30V Tr1 and Tr2 should be Mj481s and Tr3 a 2n1613.

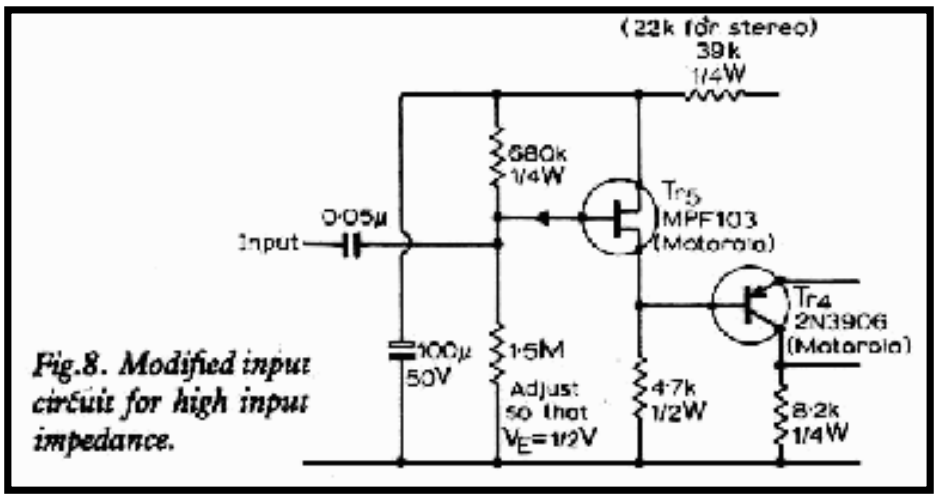


Fig.8. Modified input circuit for high input impedance.

If the output impedance of the preamplifier is more than a few thousand ohms, the input stage of the

amplifier modified to include a simple f.e.t. source follower circuit shown in fig 8. This increases the harmonic distortion to about 0.12%, and is therefore (theoretically) a less attractive solution than a better pre- amplifier.

A high frequency roll-off can be obtained, if necessary by connecting a small capacitor between the gate of the f.e.t and the negative (earthy) line.

Suitable transistors

Some experiments were made to determine the extent to which the circuit performance was influenced by the type and current gain of the transistors used. As expected the best performance was obtained when high-gain transistors were used, and when the output stage used a matched pair. No adequate substitution is known for the 2N697 / 2N1613 type used in the driver stage, but examples of this transistor type from three different manufacturers where used with apparently identical results. Similarly, the use of alternative types of input transistor produced no apparent performance change, and the Texas Instuments 2N4058 is fully interchangeable with the Motorola 2N3906 used in the prototype.

The most noteworthy performance changes were found in the current gain characteristics of the output transistor pair, and for the lowest possible distortion with any pair, the voltage at the point from the loudspeaker is fed should be adjusted so that it is within 0.25 volt of half the supply line potential.

The transistors used in these experiments were Motorola MJ480/481, with one exception, in which Texas 2S034 devices were tried. The main conclusion which can be drawn from this is that the type of transistor used may not be very important, but that if there are differences in the current gains of the output transistors, it is necessary that the device with the higher gain shall be used in the position Tr1.

When the distortion components were found prior to the onset of waveform clipping, these were almost wholly due to the presence of second harmonics.

Constructional notes

Amplifier

The components necessary for a 10 + 10 watt stereo amplifier pair can be conveniently be assembled on a standard 'Lektrokit' 4" X 4.75" s.r.b.p. pin board, with the four power transistors mounted on external heat sinks. Except where noted the values of components do not appear to be particularly critical, and 10% tolerance resistors can certainly be used without ill effect. The lowest noise levels will however be obtained with good quality components, and with carbon-film or metal-oxide resistors.

* Metal film resistors should be used throughout, as these are superior to carbon film types in all respects. These are generally only available as 1% or better tolerance, which will not pose any problems.

Power supply

A suggested form of power supply unit is shown in Fig. 9(a). Since the current demand of the amplifier is substantially constant, a series transistor smoothing circuit can be used in which the power supply output voltage may be adjusted by choice of the base current input provided by the emitter follower Tr2 and the potentiometer VR1. With the values of the reservoir capacitor shown in table 3, the ripple level will be less than 10mV at the rated output current, provided that the current gain of the series transistor is greater than 40. For output currents up to 2.5 amps, the series transistors indicated will be adequate, provided that they are mounted on heat sinks appropriate to their loading.

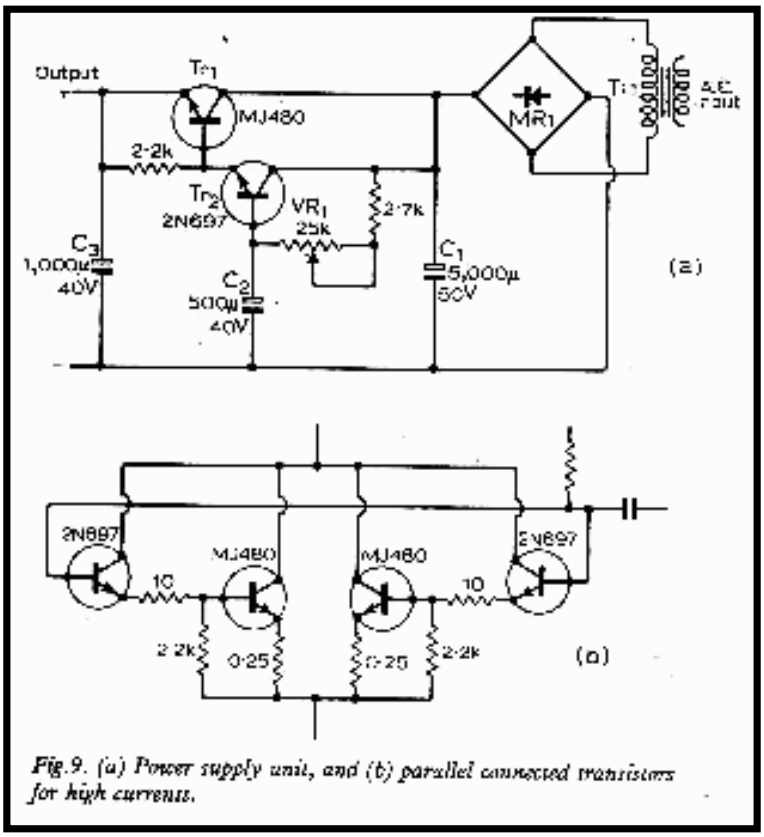


Fig. 9. (a) Power supply unit, and (b) parallel connected transistors for high currents.

However, at the current levels necessary for operation of the 3-ohm version of the amplifier as a stereo pair, a single MJ480 will no longer be adequate, and either a more suitable series transistor must be used, such as the Mullard BDY20, with for example a 2N1711 as Tr2, or with a parallel connected arrangement as shown in Fig. 9(b).

The total resistance in the rectifier "primary" circuit, including the transformer secondary winding, must not be less than 0.25 Ohms. When the power supply, with or without an amplifier, is to be used with an r.f. amplifier-tuner unit, it may be necessary to add a 0.25uF (160V) capacitor across the secondary windings of T1 to prevent transient radiation. The rectifier diodes specified are International Rectifier potted bridge types.

* This supply is not a regulated supply, but is a simple capacitance multiplier. For a more complete description of a better circuit, see Capacitance Multiplier Power Supply Filter in these pages.

Current Limiting

* Although there was no mention of this in the original article (and I managed to "lose" the schematic file for

a time), a current limiter was included. This will ensure that the current through the output devices does not exceed a preset value, although I believe that the concept is flawed, and is of limited value in this overall design.

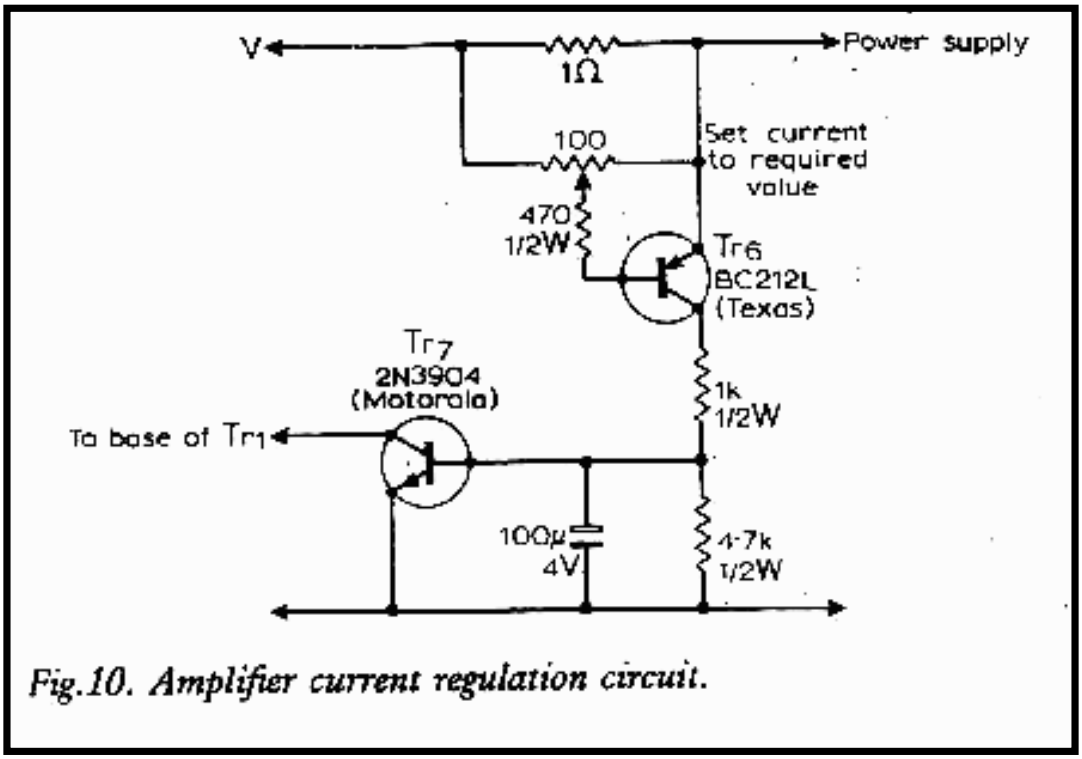


Fig.10. Amplifier current regulation circuit.

The circuit above shows the way the current limiter is connected. It will not stabilise the quiescent (no signal) current, but is only capable of ensuring that the absolute maximum current does not exceed the value determined by the 100 Ohm pot. To be useful, a current stabiliser is needed, which will ensure that the no-signal operating current remains constant regardless of temperature or supply voltage variations. No information is provided to achieve this goal.

Additional Notes

This article (with editorial notes) is reprinted as a service to readers, who are reminded of copyright laws, which may restrict the rights of readers for reproduction, commercial production (etc). The information presented is not intended as a guide for construction, but is primarily for its interest value, and to serve as a starting point for other designers.

The original article is now many years old, and some of the transistor types referred to are now superceded by vastly better designs. I will leave it to readers to experiment with device types. While much of the design is still quite relevant to a new design, I think that this amplifier may be found lacking compared to more recent design trends. In particular, the biasing system is not stable with temperature, and DC drift will be evident. In addition, the open-loop gain is very low, so feedback is far less than might be desirable (although many will feel that this is a good thing!). As mentioned above, additional local feedback (0.1 ohm resistor in the emitter of Tr1) may reduce open-loop distortion, but further reduces the gain. I suggest experimentation (I have only done some computer simulations so far) and would appreciate feedback from anyone who tries out this circuit.

I would also suggest that a single supply power amplifier is not really a proposition for new designs (although the DoZ uses the same principle), and a bi-polar (+/-) power supply may be preferable. DC stabilisation then becomes

a major issue, since small DC offset voltages can prove a disaster to tweeters in particular.

The diagrams are not of high quality, but are the originals from the source WWW page. I do not propose to redraw these, as this design is provided as information only.

The Author

John L Linsley-Hood is a prolific author of amplifier designs, and still presents new ideas and circuits in the UK magazine Electronics World (formerly Wireless World). His influence on the design of quality audio amplifers has been considerable, and continues to this day. This is not to say that I agree with or endorse all his ideas or theories, but at least he has the guts to say what he thinks, and the magazine has the guts to print it, too.

ESP Home Page

Articles Index

Elliott Sound Products **Counterfeit Semiconductors**

Last Update - 04 Feb 2006

Main Index

Counterfeits Index

MJ15003 / MJ15004

Beware of MJ15003 and MJ15004 transistors in aluminium cases (genuine Motorola [now On-Semi] devices use steel cases, and have done since 1982). Don't count on this, though - there are fake Motorola devices in steel cases too. I suggest that you test the markings with a solvent (such as acetone - nail polish remover). Most genuine transistors are marked with non-removable ink, counterfeit devices may be marked with normal screen printing ink that comes off easily.

Test the breakdown voltage with a transistor tester if one is available. Genuine devices are rated at 140V, but will usually be higher than this. Counterfeit transistors will generally have a much lower breakdown voltage. Be warned that the latest batch of fakes will actually pass this test !

Always ask for confirmation from the supplier that the devices are genuine. Feel free to refer them to this page if they claim you are mad :-)

Some time ago, I asked one of the local suppliers (who was inadvertently selling counterfeit devices) to check the authenticity of their stock. I will not name the supplier(s), as it is quite probable that they are innocent, and have been defrauded along with everyone else. Needless to say, I cannot do this checking with any supplier outside Australia, as I do not have ready access to the components they sell or to anyone who might know something.

This Australian supplier had a 'Stop Sale' on their computer for these devices, so it has been noticed by them, at least. In particular, look for a manufacture code of MEX190, with the date code 9H34. Some of the counterfeit devices even have the wrong polarity (an NPN MJ15004 - I don't think so!).

Double Headed Duds!

I have been advised that the frauds - or at least some of them - have two transistor silicon dies internally, wired in parallel in a desperate (but futile) attempt to meet the specifications. These are both quite small for the claimed power rating, and are directly bonded to the steel case. The use of two dies is in itself most unusual, but they are not even bonded to a copper heat spreader as is the normal practice, so

thermal transfer will be much worse than it should be, and thermal expansion coefficients possibly place the silicon at much greater risk of cracking - not from anything the user does, but from normal heating and cooling cycles.

I managed to convince the salesperson at an electronics outlet to sell me one of the 'MJ15003' devices, despite the 'stop sale' warning from the computer. This is fine, since I already explained why I wanted one. Most discouraging was that the salesperson obtained 'advice' from someone else in the store that the one I had (MEX190) was genuine. Well, excuse me. There were some others in the drawer that looked as if they might be real Motorola devices, but not these.

I got it home, and promptly ran some tests before I cut the top off. Gain was (barely) passable at 25 at 0.5A, and the breakdown voltage was above the 140V rating. Then I removed the top, and guess what I found? If you said "Two dies?", you are quite correct. They are exactly as described to me - two small dies, bonded directly to the steel case, and wired in parallel with what I thought were rather flimsy bonding wires. The whole construction was coated with a thin layer of silicone.

Given the sophistication of this fraud, it seems more than likely that these transistors are made in a proper fabrication plant, rather than just being relabelled junk or factory rejects. The construction overall (of my sample at least) was quite neat, and was obviously performed with the proper equipment - if I were to go to that much trouble, it would be worth the effort to use the correct die in the first place! This begs the question of where they come from, and I for one would be very interested to find out. From the latest information to hand, China and India are implicated.

One way to be sure that you have the real thing is to buy ONLY from accredited and authorised Motorola or Toshiba (or whomever) distributors. This may be irksome for home constructors, as these dealers usually have a minimum order value (locally it is AU$100 but will vary in different countries). It is not known at this stage how widespread the rort is, but since I have (over the last few years) received information from the UK, USA, Canada, New Zealand, Sweden and India about similar rackets, we can justifiably assume that no-one is safe. If anyone has further information to add, please e-mail me.

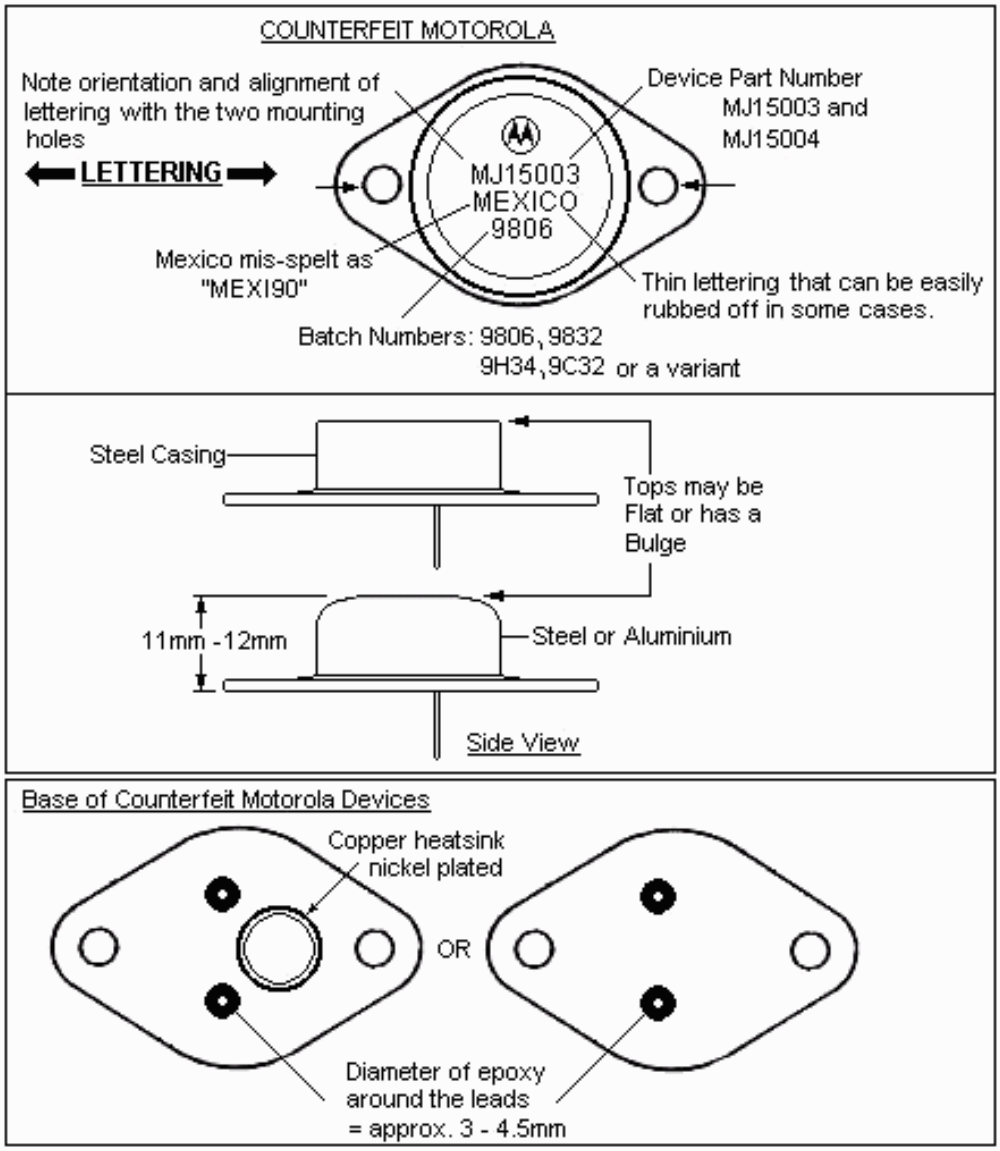
Some info received from a local supplier in response to my e-mail (reproduced verbatim) ...

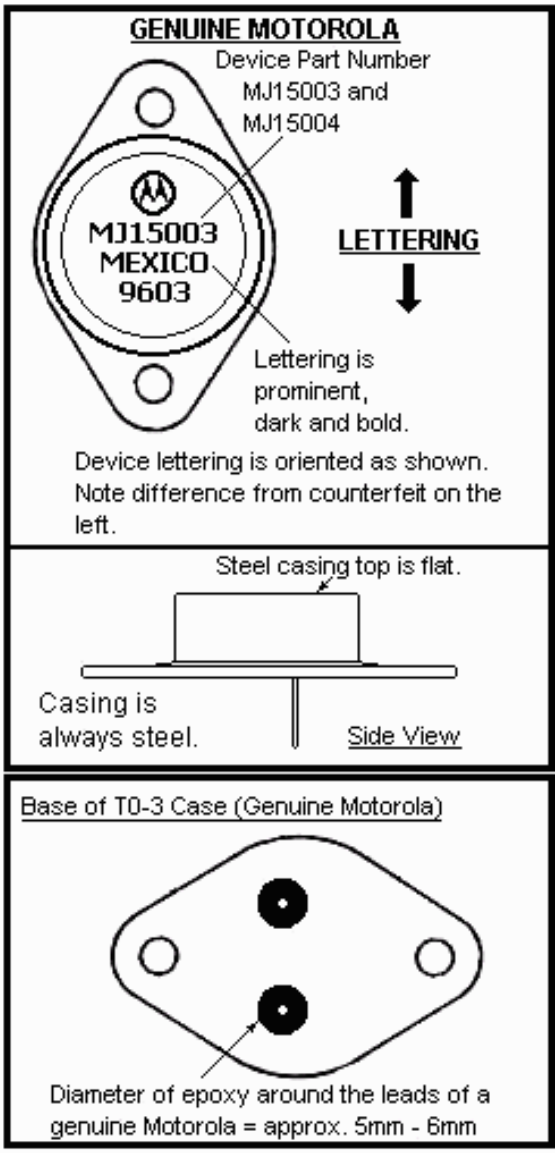
> Dear Mr. Elliott,
> Thank you for your email regarding the above matter. We have already been alerted to the problem about a couple of weeks ago when it was first noticed that some of the MJ15004 were found to be incorrectly polarised - that is, NPN instead of PNP. Our suspicions were raised and we proceeded to cut the devices open, finding their internal construction to be as per your description. Not only that, the chip dies were smaller than the known genuine Motorola types and the internal finishing was abysmal - not the usual high standards that is expected of a Motorola device. The counterfeiters were, fortunately, not too professional and it was possible, on close inspection and comparison to a genuine Motorola device, to tell them apart.
>
> From our knowledge of Motorola manufacturing processes, such a shoddy quality would never have been passed and they are definitely not from Motorola. Upon ascertaining this, we contacted Motorola or rather, ON Semiconductors in the U.S. and notified them of the counterfeits. Together with that, we also provided them with whatever information we have on hand regarding the source trail of our stock which came through a local Australian importer who brought them in through an until-now trusted source in Hong Kong. We have little to doubt the trustworthiness of our supplier as we have been dealing with him for a number of years without any problem encountered. From ON Semiconductor's reply, it would appear that they are already aware of the existence of a counterfeit ring operating out of India and China. We have left any further investigations that ON Semiconductors may want to carry out with them on an international level.

There was some more information regarding store policies that I shall not disclose, since this may identify the supplier to locals, at least. I was suitably impressed with the explanation and the efforts taken to fix this problem, and can only hope that other suppliers are equally responsive.

Oct 2000 - Dick Smith Electronics Issues 'Motorola' Recall Notice

Australian electronics retailer (and wholesaler) Dick Smith Electronics has issued a recall notice on the fake Motorola transistors, and provides a detailed description of how to identify the genuine article from the frauds. This is a good move, and offers some hope to the poor purchaser, however so far no-one else has even acknowledged that this fraud exists, despite that fact that at least one Sydney based firm is still happily selling the counterfeit devices. This is a shameful situation, and one that I would like to see corrected as soon as possible. I am not about to hold my breath, as I expect it will be a long time (if ever) before the others admit their mistake (assuming that it actually *was* a mistake!)





Finally, click either image below to enlarge - they are fairly big photos, and will take a while to load if you are on a dial-up connection. It is worth the effort though, just so you can see some real samples.

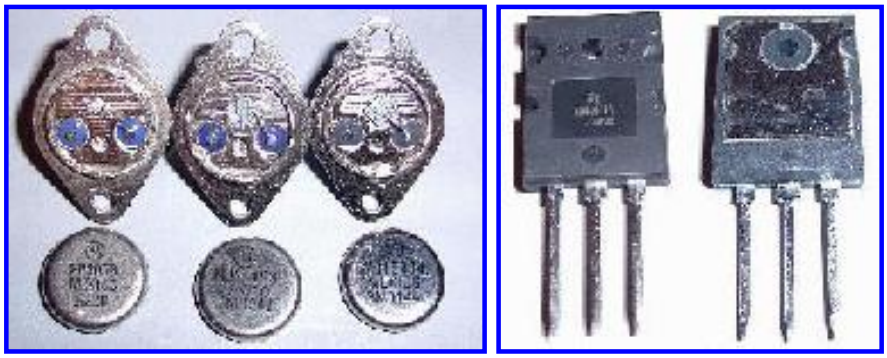
---

06 Jun 2001 - and now, another report (reproduced verbatim)

From South Africa ...

> I live in South Africa and build audio stuff for a hobby (sometimes making the odd amp for a friend and I am presently finishing a friends amp.) I was surfing the web and stumbled onto your site again. Having looked through all the project stuff, I finally went onto the editorials and came across the counterfeit transistor story.
>
> The hairs on the back of my neck started to rise the more I read because the amp that I am finishing is using MJ15003/15004 output devices, but I was at work and had to wait till I got home to check what devices I have installed in the amp. Needless to say they seem counterfeit, see attached jpeg image file, with MEX190/MEX1CO as the place of manufacture and 9H34/9R32 as the date codes (as per Richard Freeman's email to IndustryCommunity.com). I have as yet to open these device but I am sure that they will have two dies internally, when I get the time I will open them and take a photo of the internals. So the counterfeit devices are not only confined Australia, but are probably available throughout the world.

The JPEG image I was sent later confirmed the devices are fakes. They were virtually identical to those shown above.



Some Examples of Genuine ON-Semi / Motorola Transistors

Main Index

Counterfeits Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000-2006. Reproduction or re-publication is allowed due to the importance of ensuring that everyone should be aware of these fraudulent practices, on condition that the name and URL of the original page and author (Rod Elliott) of the information herein is not removed or replaced.

Page created and Copyright (c) 14 June 2000 Rod Elliott./ Updated Apr 2002 - moved section to its own page./ 04 Feb 2006, separated page from main article

Last Update - 04 Feb 2006

Main Index

Counterfeits Index

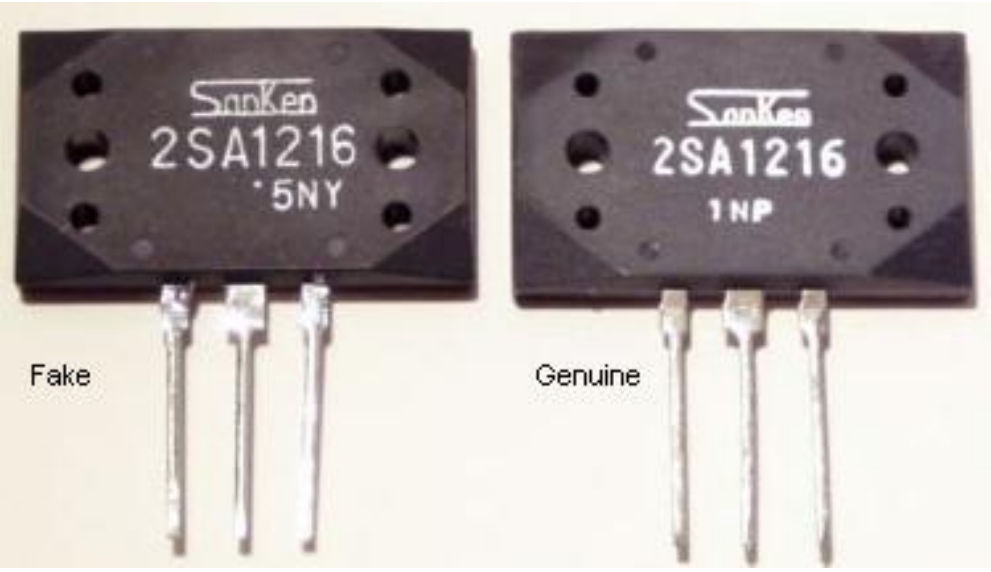
Sanken 2SA1216, 2SC2922

25 May 2003 - From Singapore, Michael Chua provided this information ...

The Sanken on the ***right*** is the original part in each photo, with the fake on the left, as indicated on the photo.

Obvious Features of Original (Front) ...

1) The lettering is thicker and smaller.
2) The production

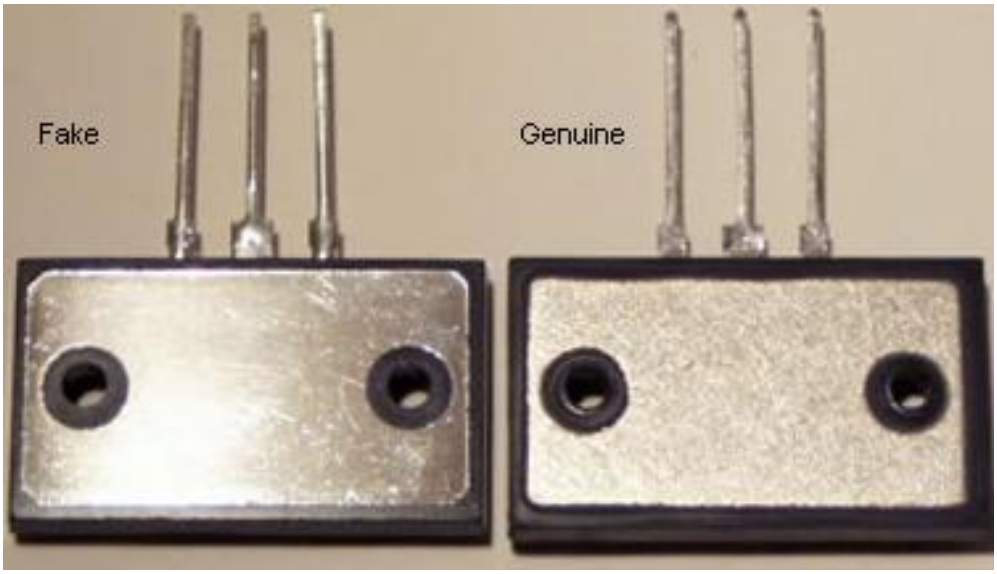


run number is in the middle, below the part number.

Less Obvious Features (Front) ...

3) Overall size - the original is slightly bigger, about 0.5mm on either edge.
4) In the originals, the angles are not as sharp (slightly rounded).

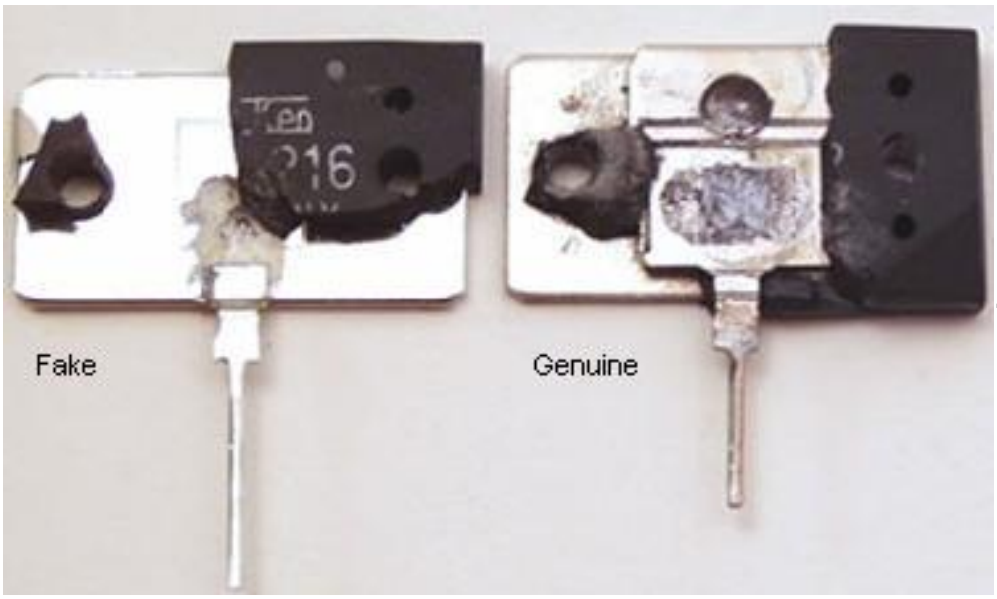
Obvious Features (Rear) ...



1) In the original, the back metal plate is slightly frosted.
2) The fake ones are shiny, almost a mirror finish.

Common

Features: Both weigh approx 10 grams.



... And when cracked open, the difference is very obvious

1) The die on the original is attached to a TO3P type heatsink which is in turn bonded to a larger heatsink contact face.
2) The the die in the original is much larger.

It is difficult to tell exactly, but the die in the fake looks to be around 3mm² which seems to be a common feature of most of the fake devices seen. That it is a great deal smaller than the original, and that it has less ability to spread the heat to the metal heatsink face is quite apparent.

I can only assume that the metal backing plate is slightly thicker in the fake, so that the weight of each unit is reasonably similar. This will not necessarily assist heat removal though, and these fake devices would be lucky to withstand 100W dissipation (based on tests I have done on other fakes with a similar die size).

When I asked if it was ok for me to 'borrow' the info from his site, Mike sent me the following e-mail ...

> I am delighted with your offer. For the sake of the audio community, it is important that these people are exposed. I was told these fakes originated from China. What surprises me is how close they are to the real thing. I will keep you posted if I come across more fakes.

See the original page at AmpsLab, and my thanks to Michael for allowing me to use his photos and info. If these are available in Singapore, you can guarantee that they will be in wide circulation very quickly, since Singapore is a major distribution centre for Asian semiconductors.

Toshiba 2SA1943 / 2SC5200

12 April 2004 - From the UK, Mark W has provided the following ...

I regret to report to you that you may now add 2SA1943 to the list of possibly fake transistors. My story will probably sound familiar. I was powering up an amp of my own design (more or less) for the first time when the fuses blew. I removed the output devices and discovered that one had failed short. I checked the schematic and the pcb layout for design/construction errors. I did this twice. I first decided that the fault was mine since earlier I had a small problem because of forgetting a sil-pad (though not on the failed 2SA1943). But also felt that the circuit and layout were essentially OK - it wasn't THAT original. I try to power up again and when the rails reached ±35V the prior story repeated. I was still inclined to suspect my own ineptness but on a lark, I smashed open one of the 2SA1943's if only to alleviate my frustration.

I won't send the samples. You could mention however that one visual characteristic of the fakes is that the Toshiba making is on a smooth shiny rectangular area of the package. On the fake it is easier to read the markings than the real ones.

What I found looked exactly, I MEAN EXACTLY, like the photo on your site of the inside of a fake 2SA1302. I conclude it's the same bastards just printing a different label on the package, At first I just thought I couldn't find the die or didn't know what I was looking for. Then I saw this 3mm x 3mm square thing pasted on with white glue. There is a clue on the outside. On a real one (I sacrificed one just be sure) the Toshiba label is typically hard to read on the dull surface. On the fake the surface has a glossy area where the label is quite clear.

The 2SC5200s from the particular supplier seem to be real. They don't die when abused and the exterior is dull flat black, somewhat difficult to read unless the light is right.

This was a foregone conclusion - I knew that the criminal bastards would go for the latest Toshiba devices sooner or later, but I must admit I'm a little surprised that it took them so long. This means that it is inviting disaster to use *any* Toshiba power transistors unless you are 100% certain of the source - unlikely in the extreme.

This is a pity, because they have a very good performance (well, the genuine ones do), but the criminals have once again ruined the reputation of a perfectly good transistor, and created a situation where it would be folly to use them in a design.

NTE36, NTE37

23 Jan 2003 - Alleged NTE devices have been found in the US (NTE36/37) that failed instantly, and just to add insult to injury, took the bridge rectifier and transformer with them. An edited quote from the reader who found these latest gems ...

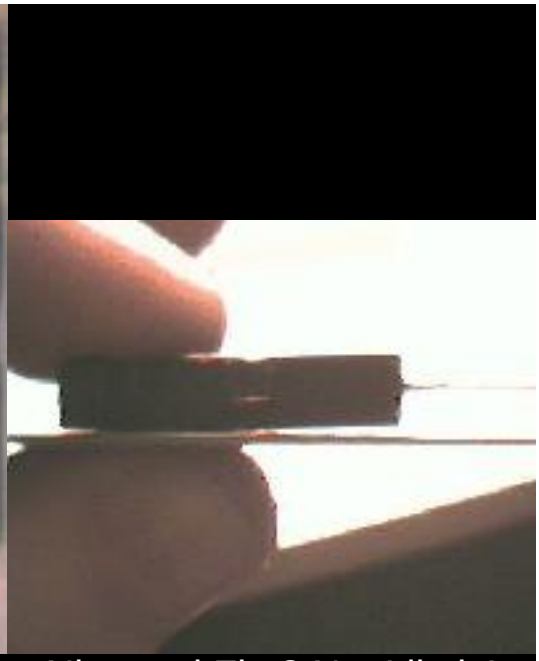
> I just got done reading your article on the counterfeit semiconductors and it all clicked for me. A little while back i was repairing an old amp of mine and purchased 2 pretty expensive matched pairs of BJT's from <name withheld> here in the States. NTE37 and NTE37 are the same as C2581 and A1106 devices. Well these NTE brand semis looked weird to start off with - one transistor of a matched pair was in a green case but the rest were in black ... under the NTE36 and NTE37 I could see a shadow of a device number on each semi that had been taken off. I had no clue about this counterfeiting thing at the time and put them in my amp - they immediately burned out but this time it took out the transformer and rectifiers with them. I thought i was just loosing my touch at fixing stuff at the time but now i see what the real problem is 'cause I cracked open one of the cases a few minutes ago and it definitely didn't look like the original semiconductor die. Tomorrow I'm going to contact the supplier and let them know, because I paid them US$40 for 4 transistors and ended up with about US$150 in damage.

Toshiba 2SA1302, 2SC3281

A reader in Sweden sent me some pictures of fake Toshiba devices, purchased from a local dealer - this shows just how bad these counterfeits can be ...

2SA1302

2SC3281



Nice and Flat? Not Likely!

Notice in particular the rightmost picture - the transistor base (the heatsink surface, not the internal connection ;-) is so convex that it won't sit even *close* to flat. In this case, there is about 0.5 mm convex curvature, which is so completely unacceptable that words fail me!

I have heard other reports of bases that are concave, so the device will never make proper contact with the heatsink, but you can't see it. Either is unacceptable (in the extreme). The printing on the 2SA1302 looks like it was done with a felt-tipped pen (well, maybe a little better than that) - not quite what one expects from a reputable manufacturer, is it? These devices passed the acetone test, so the markings are quite permanent (not all do though - the printing can be removed quite easily from some fakes).

This is an appalling state of affairs. I also recently had an e-mail from a reader in the UK. He built the P68 subwoofer amp, and was only running it into 8 ohms and it failed. After a quick exchange of e-mails, it transpired that he purchased some 'Toshiba' devices from a local supplier for less than the normal price - say no more! His first set came from a reputable dealer (but not the distributor), so almost no-one can be trusted on this score.

From New Zealand ...
A reader sent me a sample 2SA1302 and 2SC3281 to test for him, after his P68 sub-woofer amp blew up during quite gentle testing. The 2SA1302 died well before I could reach my target of 5A at 30V (the limit of the SOA curve for these devices for steady state current at that voltage). The tests were done with the transistor firmly clamped to a heatsink, and were of short duration to prevent the die from overheating - but it still failed!

Someone *must* be joking!

The die is 2.5mm square! This is tiny, and I am actually surprised that the transistor managed even to get to 2A at 30V before it blew. This is well short of the specification, and obviously the reason the amp failed. One can hardly expect a 60W (at best) transistor to provide close to 120W output (the approximate power expected from each pair of devices). Naturally, the devices that I was sent were clearly branded as Toshiba, and are just as clearly fakes.

The 2SC3281 actually managed to survive my SOA (Safe Operating Area) test, but given that the printing was almost identical to the other device, I would be highly unwilling to trust it - it may well be genuine, or simply a "better class" of fake. There were also subtle differences in the case construction of each device, with the 2SA1302 being almost identical in all respects to the Chinese devices that I have (not branded as Toshiba, but also incapable of the rated power during an SOA test). The 2SC3281 was different from any of the other samples I have.

Just in case anyone was wondering, I know it's not my P68 amp design, since I would have had

> a great many complaints by now if the design were flawed, so that only leaves the transistors as being highly suspect. My original amp works fine as well, and has been "punished" many times without failure.

Please take great care when buying semiconductors - especially 'premium' devices. Since these have the highest markup (i.e. they are fairly expensive), they are the ones most likely to be fakes.

From another reader ...

> "Add 2SA1302 and 2SC3281 (Toshiba) to your counterfeit list. I found them (counterfeit ones, that is) here, in dinky Malaysia! The ink printout is totally different from the real thing, being WHITE in colour."

Toshiba plastic transistors are usually marked in white, so this could be misleading. However I do know for a fact that Chinese (unbranded) 2SA and 2SC devices are available, but these make no pretence at being Toshiba. Perhaps (although I think we can be definite on this score) someone has bought the Chinese ones and re-branded them as Toshiba - a worthwhile effort for the criminal element, since the Chinese devices are quite cheap.

Since the Chinese devices are not branded, they cannot be deemed counterfeits, but what sort of quality you could expect is anyone's guess, so one should be wary. It is possible that these transistors are OK, but equally they may be completely useless.

Further Update (06 Jan 2002)

I have checked the Chinese versions, both a load test and visual inspection. The device I checked blew up at well below the normal peak power, and a look at the innards revealed a silicon die about 3mm square - too small for the documented power rating.

On the positive side (if there is one), at least the base plate was copper, and passably (!) flat, unlike the new fakes shown above. These devices would probably be OK in a low power amp, but cannot be used at anywhere near the full capabilities of the real Toshiba transistors.

Are these being purchased by the unscrupulous and re-labelled as Toshiba - *you can count on it* ***!***

One thing that *is* known, is that Toshiba has not made 2SA1302 and 2SC3281 transistors since 2000 (or thereabouts), so the chance of obtaining genuine devices is very low. I

would like to be able to suggest that MJL1320/3281 or the current Toshiba 2SA1943 and 2SC5200 be used, but fakes of the Toshiba devices are already spreading, and even the latest ON-Semi devices can't be far behind).

OP-07 Opamps

Where (or what) next? I received an e-mail from a reader in India, who purchased some premium opamps (at a premium price, naturally). Having paid for OP-07 opamps, one would be disheartened to put it mildly to discover that they were really 741s. I don't know if this has happened anywhere else, but it is fair warning that you could be next.

'Toshiba' 2N2773

Now we have Toshiba branded 2N2773 power transistors. This in itself is interesting, as a search on the Toshiba site reveals that they don't even seem to make this transistor! It would be unusual for a Japanese manufacturer to make a '2N' device at all, but doubly so since this is a very old device now, and seems to be discontinued by just about every other maker.

Again, these have all the earmarks of counterfeits - and naturally enough someone was caught out, and his amp failed with these transistors installed. If you happen across any of these components, be afraid - be **very** afraid!

National LM3915

... And Still They Come ... From a reader in India, and reproduced (almost) verbatim:

> I was reading the article about duplicate/fake transistors. Well, they started faking ICs too!. Don't get me wrong though. I live in India. In my best knowledge, there are no IC/Transistor making factory anywhere in India. So I don't think the fakes are MADE here. But there is a possibility that India is a kind of dropoff-point.
>
> The real purpose of this mail is to add one more IC to the known frauds. (Hundreds more maybe there). I bought this IC, LM3915, supposedly made by National Semiconductor for your LED VU Meter project. It costs about 50Rs (our currency, that's about $US1.00). But it burned out the instant I connected it to the 15-0-15 supply. I bought another one from another store. It looked a bit different. Anyway after reading your article, I got suspicious. I used a simple knife to remove the top cap like thing of the IC (Normal ICs cannot be stripped like that).

> I found another IC. When I scratched the silicon, I saw what I was expecting, LB1405. A vastly inferior (in my experience) and cheap IC. I got ripped by 5 times the cost. I might not have found out this if I didn't power it with 15 volts. I don't know how 'THEY' managed to do this. But it wouldn't have worked anyway. The pin configs are very different. I showed it to the store owner. He discarded it as my mischief. But I couldn't help my poor friend who was making a 10 channel EQ. Poor fellow. He burned out all of the ICs he bought from this store. He doesn't have the budget to replace all the chips. So he's using the EQ without the VUs. Poor chap.

The above was actually received some time ago, and I forgot about it. I have now remembered :-) As you can see, this is widespread, and many store owners are unlikely to admit that they have fraudulent stock.

I think I can say with reasonable certainty that this is the tip of the iceberg. How much reject stock (factory seconds, out of tolerance, incorrectly marked, etc) is gathered up by unscrupulous dealers and sold off as first quality? My guess is - a lot.

Always remember ... Any deal that ***seems*** too good to be true almost certainly ***is*** too good to be true!

2N3773

From Canada ...

> About three weeks ago we received a batch of transistors from Digikey... To be more specific: 2N3773's... (about... 100 of them at $1.25 ea.)
>
> I have worked with the original MOT's and I know the way they are built and labeled. These "new" parts, didn't look like anything I have ever seen, and I have been in this field for almost 20 years working in audio related goodies.
>
> The Manufacturer: MEV (you tell me if you know them)
> Case: Steel or something like that
> The finishing: lead immersed. the whole case looked as if it had been immersed on molten solder to "give" it a "silver coat" look alike. (the pins even looked as if they were used devices and had been cleaned off to strip excess solder material.
>
> The label (markings): looked like cheap paint barely stamped onto the top of the case. some acetone and it rubbed off. (YIKES!!!) And it looks like this:
>
> MEV
> 2N3773
> 94N3
>
> And, if this wasn't scary enough yet, here's the best part of the movie...
>
> I installed one new pair on a switching amp used on a GE servo. Each board makes 1 half of an

H Bridge. So a total of two boards are necessary to form a dual direction servo unit (Each amp uses a total of 12 2N3773's for a total of 90 Amps at 90Vdc, at full load when the trannies are completely either on or off depending on the direction of rotation) all the original devices on the amp were ok, except two that were shorted. After double and triple checking of the board, I installed it on a test bench we have built. (to simulate the working conditions required by GE's service manuals).

The "new" trannies lasted 15 seconds... they started off fine and gradually deteriorated until they went off with a bang!!! The rest is history... replaced them again with two more from the same batch and they worked for an hour...

So I decided to crack the first pair open... (considering i had read your stories on your website...) The Dies are smaller than those of a 2N3055. 25% smaller than the original Motorola devices.

Silicon is Silicon any way you slice it and (normally; did I hear... counterfeit???) current densities are the same from one device or manufacturer to another ... regardless of it's use or purpose. Once you go beyond this set parameter, you're in trouble. Even worse if the TO-3 case (like this aforementioned device) has a coin no bigger than 1 CM wide by 2.5 mm high. (Yes I love Metric system too.)

Footnote: I just remembered another device I ran into that same day... A 2N3055 (supposedly MOTOROLA, as it was labeled. Yet the ink used for the label was the cheap kind.) that looked almost identical to a genuine MOT device, BUT it was made in MEXICO. So far that sounds believable... Right???

Wrong!!! I opened the casing after i had blown one up at only 6 amps, and came to see that the die was slightly bigger than that of a TIP 41C. Like I might have said before, I know very well the dies in these devices. I cracked many of them open to see their guts after they blow. Weird, eh? No coin, or any internal heat spreader at all. The chip looked like it was glued to the case. No traces of the usual solder material that's normally used.

My dear friend...
As much as I love my hobby and my profession (which happens to be the same. I HATE THIS CRAP!!!! (Excuse my french!). I mean... seriously. What's next?????

Let's hope that someone sees this and takes some action!!!. I am, and WILL do my part! I hope this matter (someday) might be resolved, or at least tamed. (Yeah, right.)

I apologize for my rather dry sense of humour. Yet this is no laughing matter. Please feel free to modify this email at your will and post it on your site (if you feel it's worth the effort.)

Main Index

Counterfeits Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000-2004. Reproduction or re-publication is allowed due to the importance of ensuring that everyone should be aware of these fraudulent practices, on condition that the name and URL of the original page and author (Rod Elliott) of the information herein is not removed or replaced.

Page created and Copyright (c) 14 June 2000 Rod Elliott. Updated Apr 2002 - moved section to its own page.

Last Update - 21 Feb 2006

Main Index

Counterfeits Index

MJL21193 / MJL21194

Update - 21 Feb 2006
I have just been informed by the Australian supplier referred to below that these transistors are also branded as ON-Semi. The only way to be certain is to check the package. Genuine ON devices have square-edged steps on the pins, and (needless to say) have completely flat metal rear surfaces. The fakes have tapered pins and a stepped case as shown in the photo, and non-flat backs (although some may be 'better' than others).

The supplier has issued a recall of the ON branded fakes (having recalled the Motorola branded fakes as soon as they were alerted to the problem), but many constructors may have purchased the fakes before they were identified. If you find *any* transistors in your collection that look like the fakes shown below, return them immediately.

These first turned up in Australia in around October 2005. They were first brought to my attention by a reader early in 2006, and although I didn't get to see any in person, I have the photograph reproduced below. The supplier shall remain un-named, because they acted swiftly and decisively, protecting their customers from potential harm.

The thing that gave the game away with these was the date code. According to the code, they were made by Motorola in 2000, but Motorola had handed *all* discrete semiconductor manufacture to ON-Semi in 1998, so a Motorola branded part allegedly manufactured after that date cannot possibly be genuine.

The same (or similar) fake devices are almost certainly still available in Australia, and can probably be picked simply by the price - I

have heard AU$2.50 mentioned for small quantities. Since you can't buy the devices from ON-Semi for that, any MJL21193/4 offered for such a paltry sum is almost certainly counterfeit.

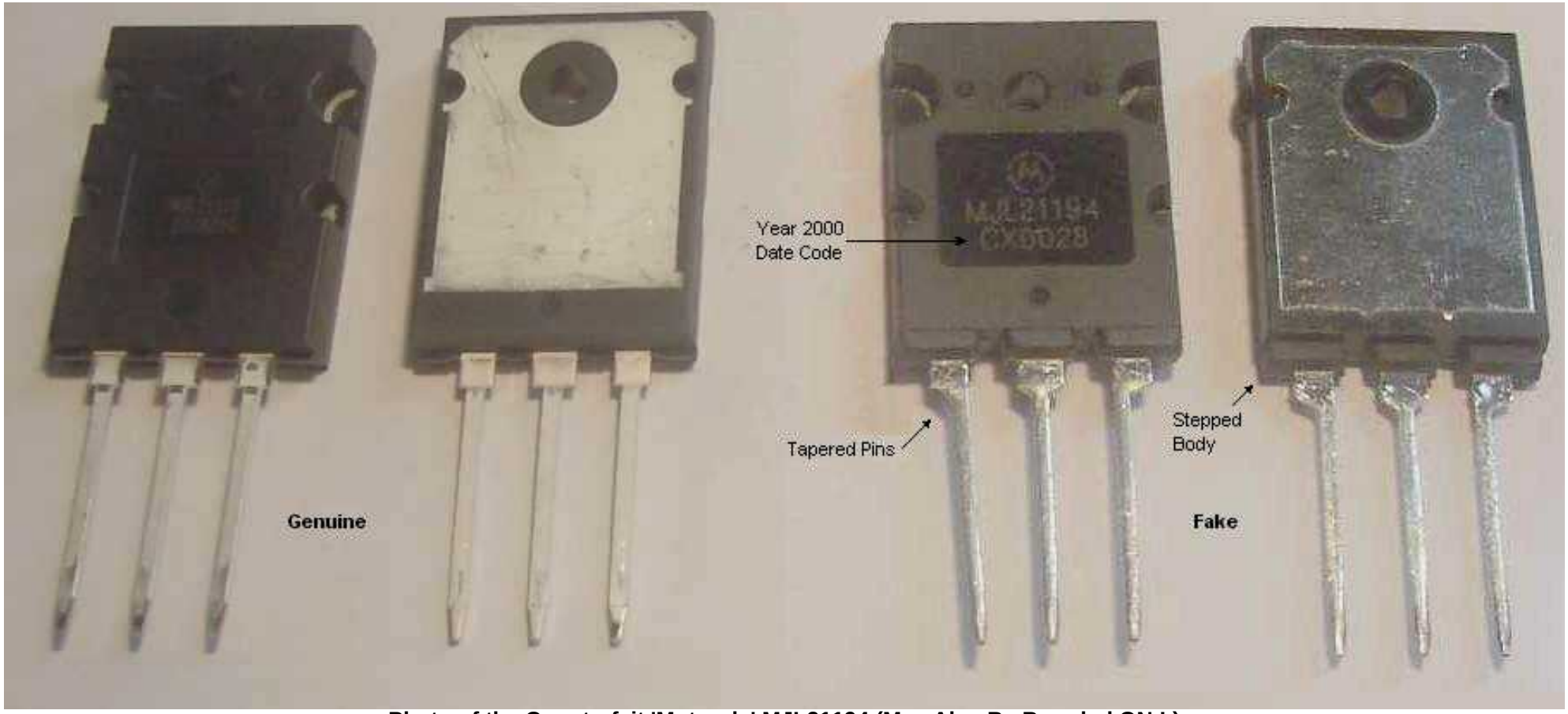


Photo of the Counterfeit 'Motorola' MJL21194 (May Also Be Branded ON !)

The e-mail exchanges are reproduced below, but as you will see, the supplier name has been removed. I did ask, and was told that they would prefer not to be named, and so shall it be. Likewise, the reader's name is not reproduced.

> I have read your article on fake transistors and thought you may be able to shed some light on a possible counterfeit transistor.
>
> I have been working on a project using MJL21193 and MJL21194 transistors and have found what seems to be an anomaly in the packaging. The ON - MJL21193's have come directly from ON-Semi as samples and in all respects conform to the drawings on the data sheet. The Motorola MJL21194's were sourced through (a supplier) and there are some discrepancies between the package and the drawing.
>
> The Motorola data sheet (old version, dated august 31 1995) and the ON data sheet (latest version from on, dated 9 June 2005) both show the same package information, both which match the authentic ON samples.
>
> The suspect devices (see attached image) have steps in the plastic where the legs come out and the outer legs taper towards the center of the

> package rather that the center of each leg. Also, the heatsink surface is not flat but curved (convex). if I hold two together where they bolt on there is a 0.5 mm gap at the other end (0.25 mm each). The ON devices are dead flat (I cannot see light between two devices held together).
>
> (The supplier) was adamant they get their stock from a large Australian supplier and that they had no concerns. Does this sound like cause for concern or is it normal that a manufacturer would use a different style package and is the flatness of the back of the package an issue? Note that the ON devices are lead-free so the metal finish may be different for that reason and is not necessarily another notable difference.

At this stage, the supplier wasn't overly sympathetic, but things changed quickly. The next part to the story comes from ON-Semi, after they were contacted about the devices. The following reply was forwarded to me ...

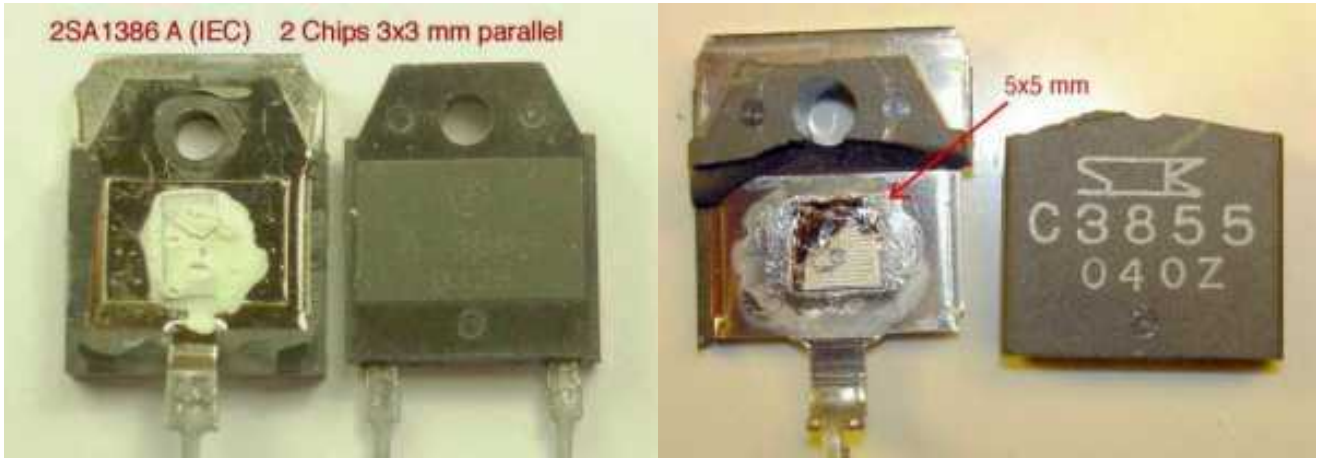
> Response From ON Semiconductor Service Request #66471
> ON Semiconductor Technical Information Center
>
> Thank you for contacting ON Semiconductor
>
> This device does not match our packaging and marking specifications. It is not ON Semi's or Motorola's device, because there is the year of production: 2000. ON Semi had take over the portfolio from Motorola in 1998. Also Motorola is not producing any semiconductors from this date, only whole electrical appliances.
>
> Regards,
> ON Semiconductor
> Technical Information Center

I contacted the supplier and passed on the concerns as soon as I had confirmation that the devices were counterfeit. Later that same day I was able to confirm that all stores were instructed to check existing stock and return any fake devices to the warehouse.

2SA1386 / 2SC3519 (?)

The next batch of fakes are not marked as Sanken, but use the same type number as a Sanken transistor. While this does not make them (technically) counterfeits, they are still fraudulent. A transistor marked as a specific type should perform to a similar standard as the original, and this is very common with a lot of 'second sourced' components. This is where more than one manufacturer makes the same type number. In general, these second sourced devices can be expected to perform as well as the original.

That was definitely not the case with these devices. They are branded as 'IEC', but don't be surprised if this is not a known semiconductor supplier. There was nothing in the information I was given to indicate that the complement was also available as an 'IEC' device, but it is safe to assume that it will be in circulation.



Photos of the IEC 2SA1386 (Left) and for comparison, Sanken 2SC3855 (Right)

The original message I received on this topic came from Germany, but beware of these devices regardless of where you are ...

> Hi Rod,
> there are not many sites dealing with counterfeit transistors. great!
>
> recently I ran into some transistors made by IEC which weren't that wonderful, at least they haven't been relabeled SANKEN.
>
> The original SANKEN 2SA1386 is rated 130Watt with a SOA of 1,5A @ 60V continuous - The IEC survived a 1A 60V pulse of 1 sec.
>
> When I tried it with 1,5A @60V it died after 0.4sec. Case temperature was only 20 (Celsius).
>
> When I opened the case I saw that it has two dies parallel, each about 3 mm square, total area is 18 sqmm.
>
> For comparison, The smaller SANKEN 2SC3855 (100W) has already a die of 5x5 mm = 25sqmm.
>
> If you like you may put the pictures in your gallery, though these types are not used in your projects.

What we see here is a flat-pack version of the 'double-headed duds' described first for the MJ15003/4. This is the first time I have seen this done in a plastic package, and was obviously an attempt to make the transistors perform to *some* standard. Needless to say it failed. The smaller Sanken device (2SC3855) is not the complement to the 2SA1386 (that is the 2SC3519), but was included as a comparison - a smaller transistor with a bigger die area than one supposedly rated for higher power. Need we say more?

It's a very sad situation, but as I have mentioned elsewhere in this section, I consider it unwise to buy *any* Japanese transistors unless you are 100% certain of their pedigree. Whoever is doing the counterfeiting seems to have targeted the Japanese devices hard, and they won't stop now.

Main Index

Counterfeits Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000-2004. Reproduction or re-publication is allowed due to the importance of ensuring that everyone should be aware of these fraudulent practices, on condition that the name and URL of the original page and author (Rod Elliott) of the information herein is not removed or replaced.

Page created and Copyright (c) 14 June 2000 Rod Elliott. Updated Apr 2002 - moved section to its own page.

The Audio Pages

Elliott Sound Products **Audio Project Pages**

This Page Is Updated Regularly
Page Last Updated - 21 April 2006
Current Number of Projects - 124

Introduction

The projects presented here are a mixture of basic (and not so basic) designs. Of these, most are originally designed by me, and have some unique quality which makes them worthy of publication, while others are simple extensions of basic theory. Not all have been built and verified, but with those that have not this is not an issue, since they cannot help but work. Of those which are not tested, they have all been simulated to verify that no stupid mistakes have been made, and will work (if correctly built) on the first attempt.

These projects are not intended for the novice electronics hobbyist, since they all require some basic (or in some cases advanced) knowledge of electronic circuit construction, mounting power transistors, soldering techniques, mains wiring, etc. Make sure you read (and heed) all warnings and the disclaimer before starting construction.

Main Index

Numerical index of projects (with brief description)

Project List	Full list in numeric order
Note Carefully	Important information
PCBs	PCB availability information
Construction	Construction page access

Suggestions	Suggest a new project
Updates	Project modifications and updates

Projects by Classification

Project Type	Number	
All projects in numerical order	124	
Power Amplifiers and Accessories	22	P117 1.5kW Power amp new 21 Apr 06
Headphone Amplifiers/ Adapters	5	
Preamps and Accessories	13	
Crossovers and Effects	16	
Power Supplies	17	P39 Update 06 Jan 06
Musical Instrument	8	
Mixers, Meters, etc.	7	
Digital Audio Applications	1	
Test Equipment	13	P16 updated 21 Feb 06
Microphones and mic preamps	4	
Miscellaneous	11	
Lighting	3	Major update to LX-800
Loudspeakers	4	

Each project description has one or more "flags", that indicate the status of the project itself. These are as follows ...

Flags	Description
new	The design (or update) is less than 2 months (or thereabouts) old.
	Mains wiring is involved, and is potentially dangerous - heed all warnings ! Note that this symbol is displayed on specifically mains powered projects, but other projects may also need a power supply which also requires mains wiring.

PCBs are (or will be) available for any project marked with this symbol. Boards may not be available immediately, but are expected within a month of publication of the project or posting of the PCB symbol.

Date The page was added or updated on the date shown (new updates are shown in red)

Updates

ESP reserves the right to change or update projects without notice, so it is important to be aware that a change may have been made. You should always watch for updates of previously published items. These are shown as a date beside the project (in the "Flags" column). Do not build any of the circuits presented here without checking for updates first. A "new" symbol indicates an update within the last two months or so (but this might vary). Special cases might be listed from time to time - these are likely to be important! Updates less than 3 months old will have the date in **red**.

Note Carefully

Please see the ESP disclaimer for important information

WARNING

Mains wiring should be carried out by suitably qualified persons only. Under no circumstances should any reader construct any mains operated equipment unless *absolutely* sure of his/her abilities in this area. Death or serious injury (to the constructor or others) may be uncommon with home construction of mains powered projects, but the risk is ever-present. The author takes no responsibility for any injury or death resulting from, whether directly or indirectly, the reader's inability to appreciate the hazards of household mains voltages and the correct wiring practices for your country. Please read the disclaimer now if you have not done so already.

Although I am happy to provide assistance to prospective builders, I cannot (and will not) be drawn into prolonged e-mail exchanges if the project does not work as expected. I can say with complete confidence that all projects presented will work ... *if properly constructed according to the published design*. This is not to say that no help will be available - I will help where I can.

It is inevitable that in some cases (due to component tolerances, for example), a project may require a different value resistor, capacitor (or whatever) to correct for an unexpected variation. Since I cannot control or predict the quality of components sourced by readers, nor the standard of workmanship in assembly, it is not possible to allow for every contingency.

Please do not attempt the construction of any project which you do not fully understand, or if you do not feel completely confident that you can build the project without further assistance. Do not expect me to be able to diagnose an obscure fault remotely, and especially if the project has been modified in any way whatsoever.

Main Index

Total Visitors since Jan 2001 - 5322200

Copyright Notice. All projects described herein, including but not limited to all text and diagrams, are the intellectual property of Rod Elliott unless otherwise stated, and are Copyright © 1999-2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author / editor (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use in whole or in part is prohibited without express written authorisation from Rod Elliott and the owner of the copyright in the case of submitted articles.

Elliott Sound Products **Contacting ESP**

Thank you for taking the time to respond. Normally I will respond to e-mails within a few days, but sometimes it may take longer. If it seems that I have forgotten you - probably not, it just means that things are a little hectic. I typically get about 30 or more e-mails each day, and yours is no more or less important than any other.

PLEASE BE CONSIDERATE: I regularly get e-mail from people using anti-spamming services. This is fine, but don't expect a response, since these services just throw my replies away. Please advise me in your e-mail that you are using one of these services, and provide a real e-mail address - *preferably at the beginning of the e-mail rather than at the very end where I might not see it*. It is very frustrating to compose a reply, and later discover that it cannot be delivered. It is a waste of time that I can do without.

As you can tell from the STOP image below and some of the comments, I continue to have a lot of e-mail problems. Apart from stinking spam, one of the biggest issues I have (ongoing) is irritating questions where people are too lazy to look up the information for themselves.

E-Mail Policies - Please *READ!*

Yes. **All** of it!

- ***ESP is not a free information service. Limited support only is available to those who do not purchase PCBs.***
- ***If you send me invitations to join a "friend's network" or similar, it will be deleted without answer. No exceptions.***
- I am happy to answer questions about my own projects and articles, but I will not

be drawn into lengthy exchanges if you build something and cannot make it work (*especially* if you make your own board). In 99.9% of cases where a circuit does not work **you have made a wiring mistake.** It is extremely difficult to diagnose faults remotely, so if you are not sure of your ability to build a project - **don't**. All schematics where a PCB is available work - period! If yours does not, then the mistake is yours, not mine. See the troubleshooting guide before contacting me.

- Before asking questions, read the **FAQ** page, and check that the articles don't have the info you need. In many cases, I am asked questions that are fully explained on the pages. Visit the main index for access to all the information available on this site.

- I ***will not*** attempt to give answers to questions about stuff you find from links or elsewhere. Contact the owner of the site (or the manufacturer for commercial products) for more information.

- Consider posting your question on the ESP Forum rather than sending an e-mail. Others may be able to spend more time to help you with your questions.

- Do ***not*** ask me to endorse (or otherwise) 'Brand X' of anything - cables, amps, speakers etc, because I will not do so. Refer to the disclaimer and other comments relating to the naming, endorsing, or bagging of any product by name or model.

- I reply to most e-mails. If you do not get an answer, the e-mail address you (or your e-mail client) supplied was wrong, or your e-mail service may have problems. There may also be a spam blocker in place - be warned - I will only try to e-mail you **once**, if my attempt is blocked I will not try again.

- In some cases, your question or request was such that I will simply delete it. I regularly receive e-mails that are a blatant attempt to obtain something either free or by deception. If you send me an e-mail like that, I **will** delete it without a second thought.

- If your e-mail looks like spam, my spam filter(s) will delete it. Make sure that you include a sensible subject line, and if at all possible, avoid using free e-mail services (especially Yahoo! and its ilk).

- **Please make sure that the following are included ...**
    - Your name (it is extremely inconsiderate to send mail with no name, other than a "reply to" name that is often incorrect).

- Your e-mail address (one that **works!**).
- Proper English (no 'SMS speak' or other stupid abbreviations). *If u cnt b bothered typing yr req in full, I cnt b bothered trying 2 decode it*. If your primary language is other than English, just do your best. Most apologies from non-English speakers are redundant - their command of our language is better than some who speak nothing else! (true!)

Your computer thinks that the time and date is ...

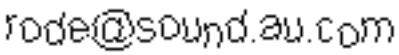
Mon May 29 12:20:37 GMT-0400 (Eastern Standard Time) 2006

If this is not correct, please set at least the date correctly, as any e-mail that appears "old" may not get a reply in a timely manner (if at all).

Cheers,

Rod

If Javascript is *not* enabled, the *only* way to send me an e-mail is to (manually!) copy the address below - it is an image to prevent stinking spammers from harvesting my address! Otherwise, click on the appropriate link.

For general questions ...	Rod Elliott	
For site problems ...	Webmaster	

Main Index

Copyright © 1999 - 2006 Rod Elliott. Page Last Updated - 13 Jan 2006

- SIMetrix SPICE Simulator
- SIMetrix/SIMPLIS Simulator
- What our customers say
- Download

"I have used pretty much all of the simulators out there, Hspice, HPspice, PSpice, Mspice, Accusim, Spectre, SpectreRF and the list goes on... but none are as simple, straightforward to use as SIMetrix."

Click here for more customer quotes

Legal ©2006 Catena Software Ltd.

- catena
- sales
- support
- products
- home

Circuit Si ftware

Catena Software provides reliability, ease of use and accuracy in a range of professionally-packaged circuit simulation tools.

With industry-proven performance and technical support to match, Catena Software offers an affordable alternative to unsupported freeware and high-priced EDA solutions.

Download Latest Release

SIMetrix and SIMPLIS

SIMetrix is a mixed-mode circuit simulation package designed for professional electronics engineers. SIMetrix comprises a substantially enhanced SPICE simulator, schematic editor and waveform viewer in a unified environment.

SIMetrix/SIMPLIS is an advanced software tool that enables efficient design of power electronics circuits. SIMetrix/SIMPLIS combines accuracy and convergence in a unique and powerful design environment, enabling 10-50x faster simulation for power supply designs.

Try SIMetrix and SIMetrix/SIMPLIS for free. Download now.

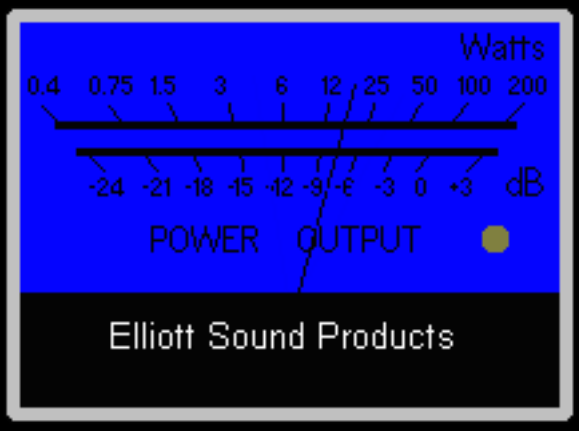
SIMetrix/SIMPLIS product information: Click here

The Audio Pages

Elliott Sound Products - *Professional results for the Do-It-Yourself enthusiast.*

ESP is uniquely positioned to offer you the best of the best in audio, since you decide the quality of the parts used. You can purchase PCBs or other products from ESP with confidence - support is always available.

- **Main Index** Complete indexes of everything on **The Audio Pages**
- **Purchase** Project Printed Circuit Boards and other products
- **Projects** Complete listing of ESP's free project and design articles
- **Articles** Informative articles about audio, building projects, and electronics in general
- **Contact** How to contact ESP



Elliott Sound Products
PO Box 233
Thornleigh NSW 2120
Australia

ESP has no affiliation with any advertiser or product supplier. The Audio Pages are supported entirely by the purchase of printed circuit boards and other products.

You are visitor number 5322201 since Jan 2001

Web page design and all images copyright © 2002 Rod Elliott - All rights reserved.

Elliott Sound Products **Project 57**

The ESP SIM (Sound Impairment Monitor)

Rod Elliott (ESP)

Introduction

The ESP Sound Impairment Monitor is a method of determining just how much your amplifier modifies the original signal. The first version is designed to be built into an amplifier circuit, and although quite simple in concept will indicate if any modification is made to the signal by the amp, for any reason at all.

For example, many amplifiers have overload protection, and this may activate without you even realising it. The SIM will react immediately, since the input and output of the amp no longer match. The meerest hint of clipping - however brief - will turn on the LED, which is designed to stay on for long enough for you to see it.

The internal SIM is the simpler of the two variations, and is actually more accurate than the more complex external version, however it is not as versatile - it is built into the amp after all, so can't be used for anything else.

Description

Nearly all modern amplifiers use a differential input stage, based on the so-called "long-tailed pair". This type of amp will have the input signal applied to one input (the non-inverting input is almost universally used as input), and the feedback to the other. This configuration is identical to that used in an opamp, and just like an opamp, the power amp will attempt to make both inputs have the same voltage at any point in time.

When the amplifier protection circuits operate, or the amp starts to clip, the inputs will have very different voltages on them. Slew rate limiting will also cause the inputs to develop very different voltages, and indeed, any amplifier abberation will change the voltage differential between the inverting and non-inverting inputs of the amplifier.

The internal SIM uses an opamp to measure the signal at each of the amp inputs. This will normally be fairly low - perhaps in the order of about 2 to 5mV under normal operation at just under full power. The SIM has enough gain to allow an indication at voltage differentials down to 1mV between inputs, which means that the finest of amplifiers is easily monitored for any impairment.

Figure 1 shows the differential amp used by the SIM, and it is a completely conventional circuit. This is adjustable by using VR1 to null out any normal variations that the amp might show when there is no distortion or other nastiness in evidence. The second stage is a high gain amplifier, and will amplify the residual signal to a level suitable for the rectifier and indicator circuits.

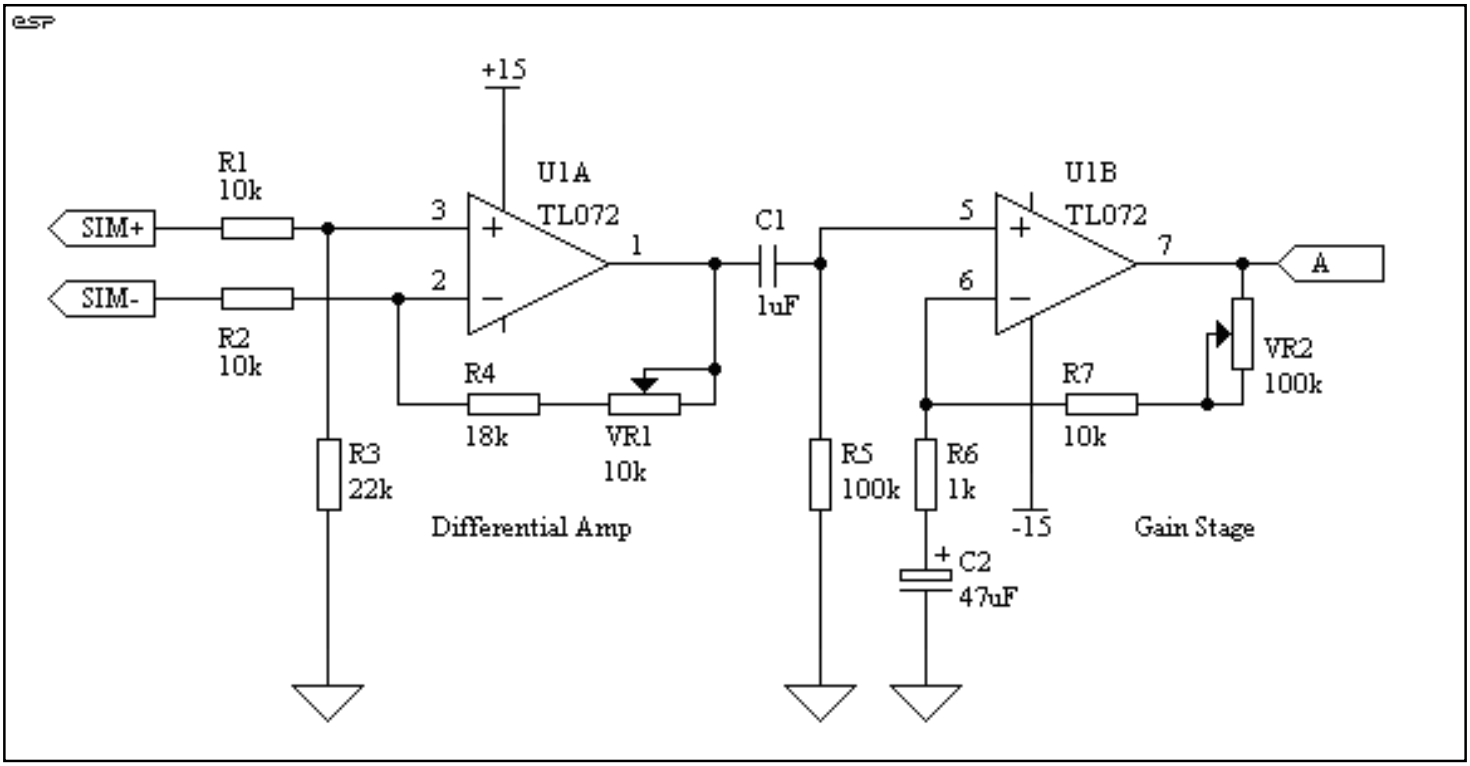


Figure 1 - The SIM Input Stage

The opamp shown is adequate for most applications, but the use of a premium device from Analog Devices or Burr-Brown will improve performance, especially at high frequencies. This is not strictly necessary, because the operation of any amplifier protection or clipping (or any other undesirable effect) will affect all frequencies, not just those that caused the original problem. This is why TIM (transient intermodulation distortion) is so objectionable if it occurs. The SIM will reliably detect any such problem, probably before you are even remotely aware that it is happening.

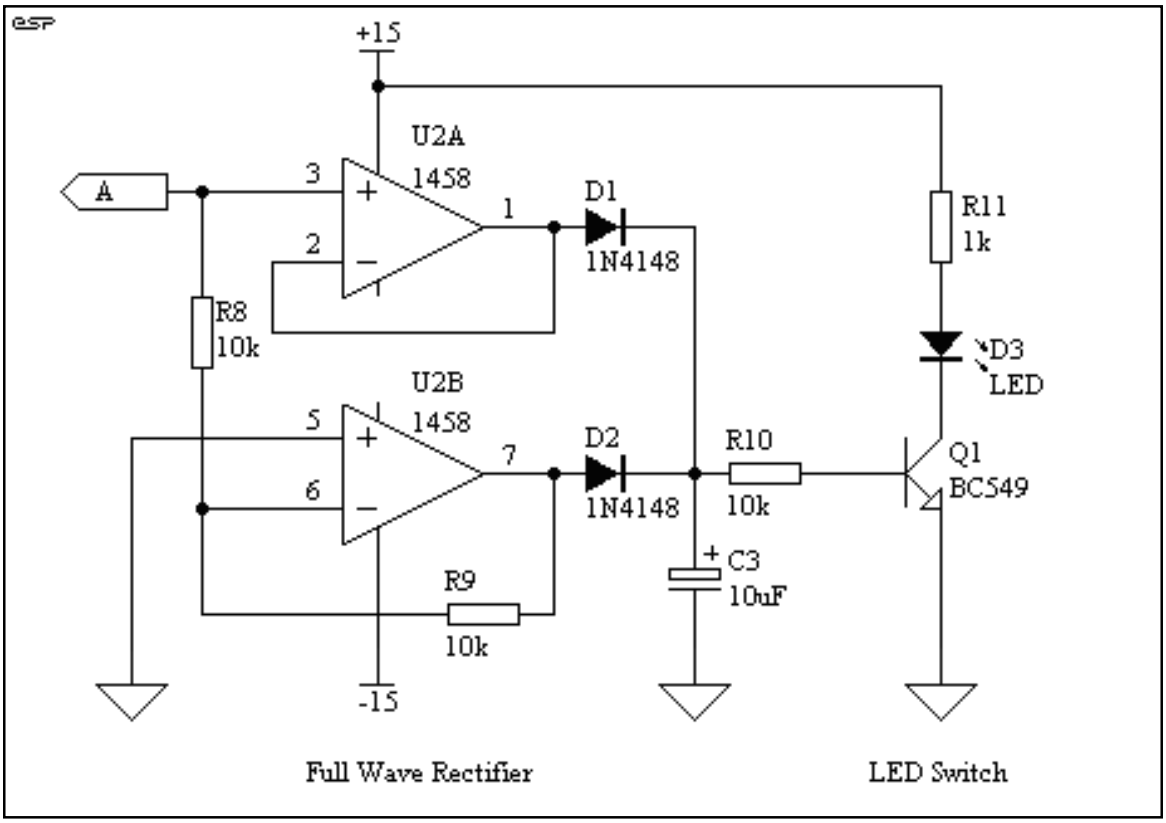


Figure 2 - SIM Detector and LED Switch

Figure 2 shows the detector, which is a simple full wave rectifier. The 1458 opamp specified is quite adequate for this, as the two sections function only as buffers, and accuracy is not important. When the error signal exceeds the detection threshold (about 1.2 Volts), the LED will light, and remain on for long enough to see, even with very short duration signals. The use of a full wave rectifier ensures that an error signal of either polarity will activate the LED.

Connection To An Amplifier

The circuit is designed to use 10k resistors at the amp inputs (as well as those on the SIM circuit itself). Figure 3 shows a typical amplifier (the 60W amp of Project 03) with the SIM connections added. This can be applied to almost any amplifier, with the only effect being that the input impedance will be a little lower than before because of the 42k SIM+ connection input impedance. If desired, the input impedance of the SIM can be increased, and the resistors R3, R4 and VR1 will need to be increased as well to maintain the same ratio.

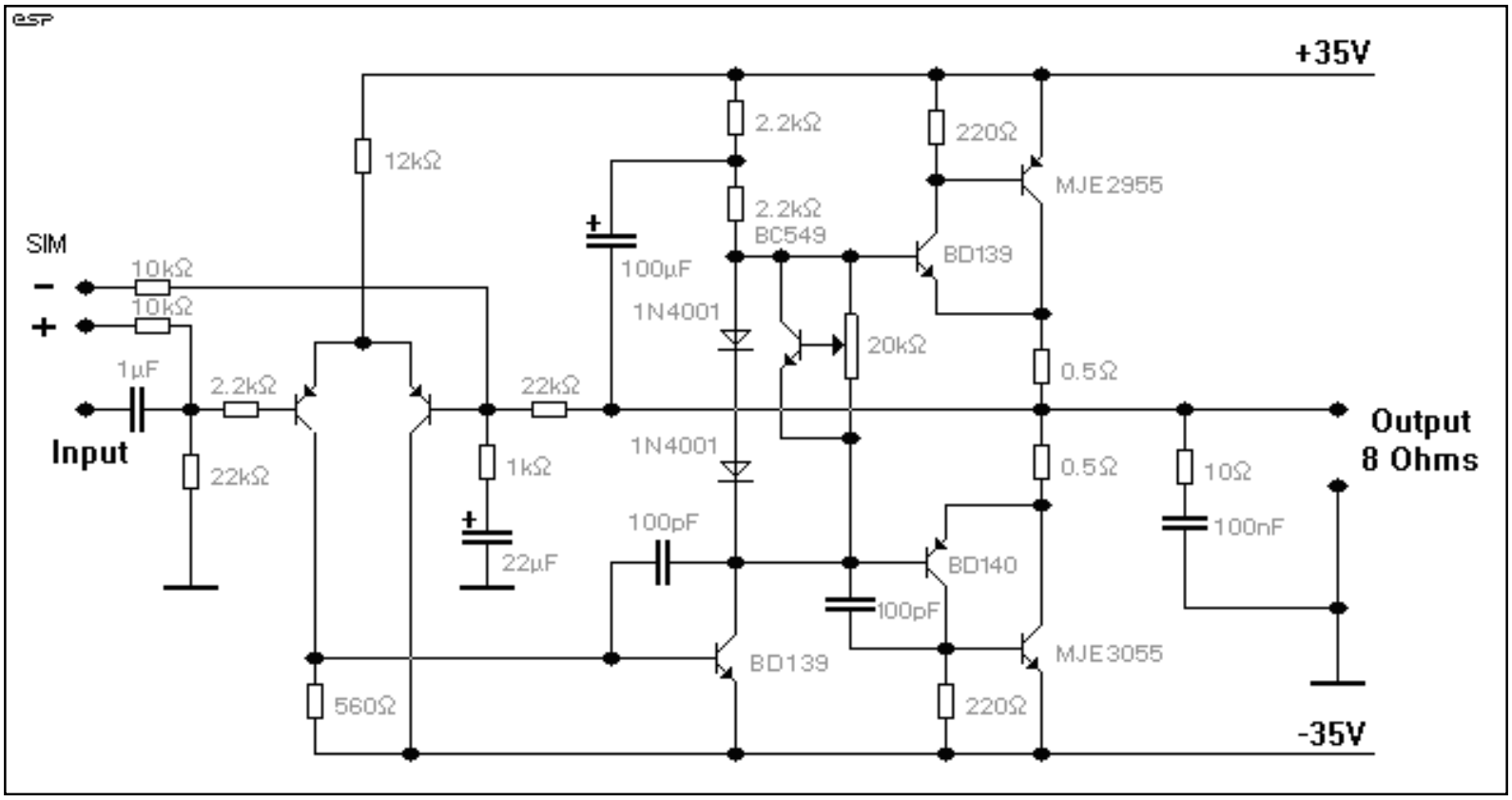


Figure 3 - Connecting The Internal SIM To An Amp

Make sure that all wiring is as short as possible, and the wires should be tightly twisted and preferably shielded as well. This will prevent noise from being picked up by the SIM or the amp. The 10k resistors at the amplifier are to ensure that the stability of the amplifier is not compromised by the capacitance of the cable, and to provide a buffer against noise pickup. These must not be omitted.

Setup And Operation

When the SIM is connected to a power amp, the unit must be calibrated. This will not be as easy for those without an oscilloscope, but a steady signal source is essential. This can be an audio oscillator or a test CD, and the idea is to adjust the circuit so that at all normal settings below clipping, the LED remains off.

First, advance the input level to the amp to a normal to quiet listening level. Disconnect the speaker. With VR2 advanced so that the LED is flickering or barely on, carefully adjust VR1 until the LED is extinguished. Advance VR2 again, and readjust VR1, and repeat this until no further improvement is posible.

If an oscilloscope is available, apply sufficient signal so that the amp is just below clipping. No clipping whatsoever is allowed, so give yourself enough margin to ensure that this is so. Adjust VR2 until the LED is barely visible, then reduce the setting very slightly so the LED is off. If you turn up the signal until the amp is almost clipping, the SIM will almost certainly show the vestiges of clipping well before it is visible on the oscilloscope.

If a 'scope is not available, then a certain amount of guesswork is needed. This will possibly reduce the sensitivity a little, but any audible distortion will still cause the LED to light. You will have to estimate the maximum level that you think is below clipping. If VR2 has to be reduced to near minimum resistance, there is almost certainly too much signal, so reduce it until VR2 can be advanced to close to the centre of rotation. The actual setting can be found by trial and error - at the point when the amp starts to clip, the SIM's LED will go from off to full brilliance (and vice versa) with the tinest change of input signal. Find this point, and reduce the level slightly. Adjust VR2 until the LED is just off.

The SIM is now set up properly, so the speakers may be reconnected and some music played through the system. It is possible that VR2 may have to be wound back a tiny bit if the LED operates at low listening levels, but if the amp is as good as the makers claim, there should be no indication until one of the "defined events" listed occurs - clipping, overload protection or slew rate limiting.

The SIM will indicate if any musical piece causes amplifier distress of any kind. As long as that LED remains off, all is well with your amp.

Projects Index Page

Home Page

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright (c) 22 Apr 2000

Elliott Sound Products **Project 3A**

60-100W Hi-Fi Power Amplifier

Rod Elliott (ESP)

PCBs are available for this project. Click the image for details.

Introduction

Update - 24 Jul 2003. OnSemi has just released a new range of transistors, designed specifically for audio applications. These new transistors have been tested in the P3A, and give excellent results. As a result, all previous recommendations for output transistors are superseded, and the new transistors should be used.

The output devices are MJL4281A (NPN) and MJL4302A (PNP), and feature high bandwidth, excellent SOA (safe operating area), high linearity and high gain. Driver transistors are MJE15034 (NPN) and MJE15035 (PNP). All devices are rated at 350V, with the power transistors having a 230W dissipation and the drivers are 50W.

Having built a P3A using these transistors, I recommend them highly - the amplifier is most certainly at its very best with the high gain and linearity afforded by these devices. Note that there are a few minor changes to the circuit (shown below).

The basis for this amplifier was originally published as Project 03, and although the base design is over 20 years old, as an amplifier it remains "state of the art" - this is an extremely good amplifier. It is simple to build, uses commonly available parts and is stable and reliable. The design featured is a full update on the original project, and although it has many similarities, is really a new design.

This new amp (like the original) is based on an amp I originally designed many years ago, of which hundreds were built. Most were operated as small PA or instrument amps, but many also found their way into home hi-fi systems. The amp is perfectly capable of driving 4 Ohms, provided the supply voltage is maintained at no more than ±35V.

This amplifier, although very simple, is capable of superb performance. This is *not* an amp to be under estimated, as the sonics are very good indeed, and this is due (in part, at least) to the inherent simplicity of the design. The amp is exceptionally quiet, and is reasonably tolerant of difficult loads. It is an ideal amplifier for biamped systems, and may be operated in bridge mode (BTL) if you use the recommended output transistors (which have the necessary power ratings).

The design has had the benefit of many, many years of consistent use, and this version is the best of all - the

refinements ensure minimum "switch on" or "switch off" noise, and the availability of really good output devices has improved on a known and very stable design.

I have heard nothing but praise from those who have built this amplifier - the only feedback I have received has been very positive indeed. The sound quality is up there with the very best. Highly recommended !

A reader has constructed a site based on the assembly of the P3A amp, and he has some useful information that will be helpful for anyone else looking for equivalent transistors, design guides, etc. Tyler has put considerable effort into the site. See the Project 3A Construction Guide.

Description

Note that like the original, there is (still) no output short circuit protection, so if speaker leads are shorted while the amp is working with a signal, there is a very real risk of the transistors being destroyed. The specifications are very similar to those of the original project, but the use of a current sink in the differential pair input stage means that there is virtually no thump at turn on or off.

I have also added the ability to adjust the quiescent current, and with the transistors specified the amp will provide 100W into 8 ohms, at a **maximum** supply voltage of ±42V. This supply is easily obtained from a 30-0-30V transformer.

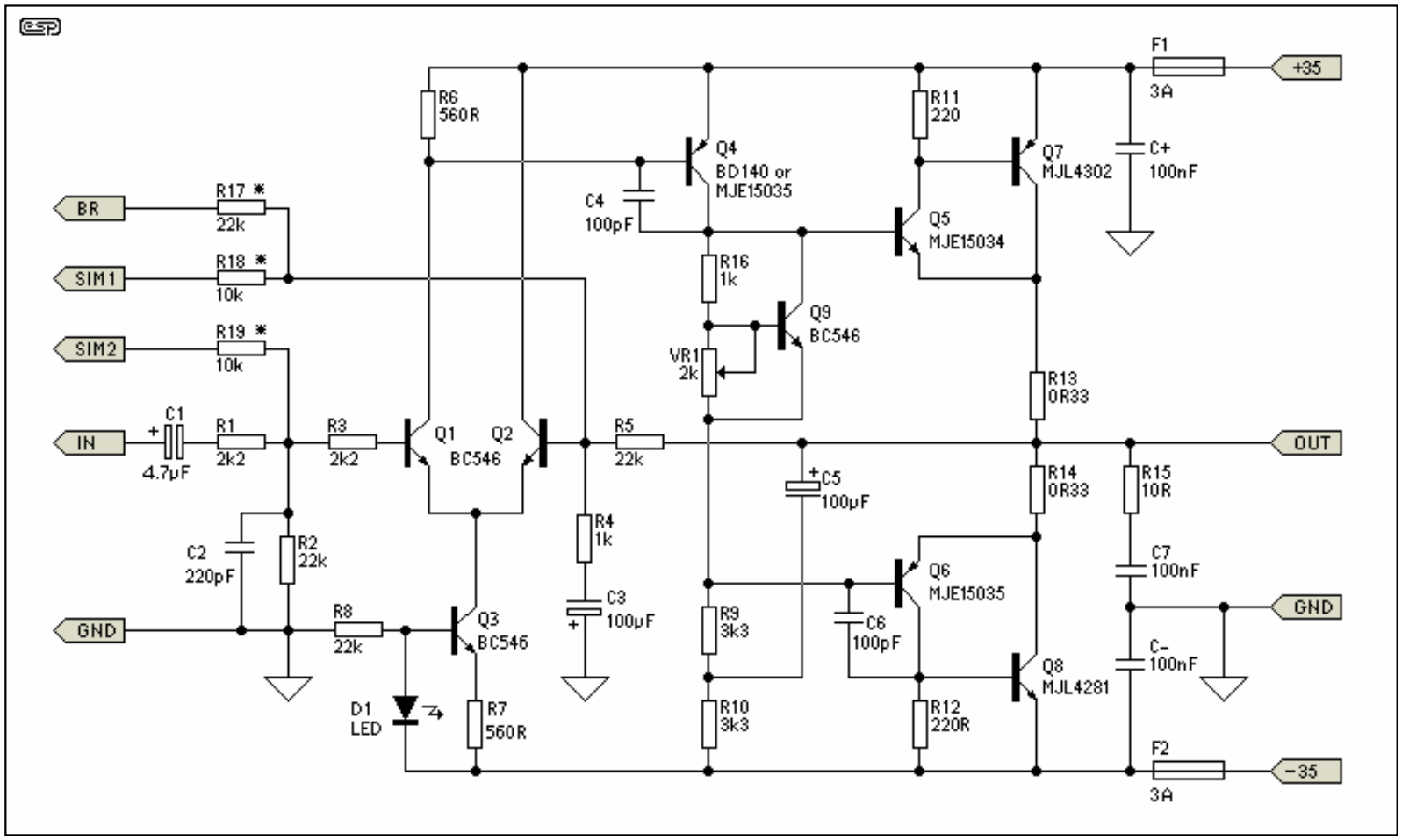


Figure 1 - Amplifier Schematic

As can be seen, it is not a complex amp, but the performance is excellent. Connections are provided for a

SIM (Sound Impairment Monitor), and there is also a resistor (R17) added to allow bridging. This resistor connects to the output of the other amplifier (the master). When used in this way, the input should be grounded - do not omit the capacitor, or DC offset will be too high. When used in bridge mode (also called BTL - Bridge Tied Load), the SIM should be taken from the master amplifier only.

* Components marked thus are optional - if you do not want to use the SIM or bridging, these may be omitted completely.

For use into 4 ohms (including bridging into 8 ohm loads), do **not** exceed ±35V (from a 25-0-25V transformer). Most applications will be satisfied with the lower voltage, and the reliability of the amp is assured with almost any load. In bridge mode, this amp will happily produce 200W into 8 ohms, and will do so reliably even for continuous high power levels. Never attempt to operate the amp in bridge mode into 4 ohms, as this represents an equivalent load to each amp of 2 ohms. The amp was not designed to handle this, and will fail. ±42V is the **absolute maximum** voltage, and should only be used where 4 ohm loads will never be applied.

D1 is a green LED, and should be a standard type. Don't use a high brightness LED, or change the colour. This is not for appearance (although the green LED looks pretty neat on the board), but for the voltage drop - different coloured LEDs have a slightly different voltage drop.

VR1 is used to set the quiescent current, and normally this will be about 50-100mA. The amp will work happily at lower current, but the distortion starts to be noticeable (on a distortion meter monitored by an oscilloscope) at less than around 40mA. The Class-A driver (Q4) has a constant current load by virtue of the bootstrap circuit R9, R10 and C5. Stability is determined by C4, and the value of this cap should not be reduced. With fast output transistors such as those specified, power bandwidth exceeds 30kHz.

With the suggested and recommended 35V supplies, Q4 and the output drivers (Q5 and Q6) will normally not require a heatsink. With 4 ohm loads, you may find that a heatsink for Q5 and Q6 is needed, but my experience is that these transistors should not get hot under most operating conditions.

If using the amp at ±42V, a small heatsink should be used for Q4, as the dissipation will be quite a bit higher and the device will get very warm.

Although I have shown MJL4281A and MJL4302A output transistors, these are very recent and may be hard to get for a time. The recommended alternatives are MJL21193 and MJL21194.

Note: It is no longer possible to recommend any Toshiba devices, since they are the most commonly faked transistors of all. The 2SA1302 and 2SC3281 are now obsolete, and if you do find them, they are almost certainly counterfeit, since Toshiba has not made these devices since around 1999~2000.

Before applying power, make sure that VR1 is set to maximum resistance to get minimum quiescent current. This is very important, as if set to minimum resistance, the quiescent current will be very high indeed (probably enough to blow the output transistors!).

Construction

Since I have boards available for this amp, I obviously suggest that these be used, as it makes

construction much easier, and ensures that the performance specifications will be met. Note that the layout of any power amplifier is quite critical, and great pains were taken to minimise problem areas - if you make your own PCB, it is unlikely that you will be able to match the published specifications.

All resistors should be 1/4W or 1/2W 1% metal film for lowest noise, with the exception of R9, R10 and R15 which should be 1/2W types, and R13, R14 must be 5W wirewound.

The bootstrap capacitor (C5) needs to be rated at at least 35V, but the other electrolytics can be any voltage you have available. The trimpot (VR1) should ideally be a multiturn, but an ordinary single turn pot can be used. Setting the current will be a little more difficult with a single turn pot, and they are not as reliable.

A pair of these amps will be quite happy with a 1°C/W heatsink for normal hi-fi use. Consider using a fan if you are going to push the amp hard. Remember - there is no such thing as a heatsink that is too big.

Basic Specifications

The following shows the basic measurement results ...

Gain	27dB
Input Impedance	24k
Input Sensitivity	1.22V for 100W (8 ohms)
Frequency response [1]	10Hz to 30kHz (-1dB) typical
Distortion (THD)	0.04% typical at 1W to 80W
Power (42V supplies, 8 ohm load) [2]	90W
Power (35V supplies, 8 ohm load) [3]	60W
Power (35V supplies, 4 ohm load)	100W
Hum and Noise	-73 dBV unweighted
DC Offset	< 100mV

Notes

1. The frequency response is dependent on the value for the input and feedback capacitors, and the above is typical of that when the specified values are used. The high frequency response is fixed by C4, and this should not be changed.
2. Operation into 4 ohm loads is ***not recommended*** with the 42V supplies. Peak dissipation will exceed 110W in each output transistor, leaving no safety margin with typical inductive loads. All supply voltages are nominal, at no load - your transformer may not be capable of maintaining regulation, so power may be slightly less than shown.
3. This figure is typical, and is dependent on the regulation of the power supply (as are 1 and 2, above). Worst case power with 8 ohm loads is about 50W, but the supply will be seriously inadequate if the power falls that far.

Four of these amps in a biamped arrangement will give you prodigious SPL, and is similar to the arrangement I am using. Coupled with a Linkwitz-Riley crossover, the amplifiers can be mounted in the back of the speaker box, so only signal and power are needed for a complete system that will leave most commercial offerings for dead.

Powering Up

If you do not have a dual output bench power supply - Before power is first applied, temporarily install 22 Ohm 5 W wirewound "safety" resistors in place of the fuses. Do not connect the load at this time! When power is applied, check that the DC voltage at the output is less than 1V, and measure each supply rail. They may be slightly different, but both should be no less than about 20V. If widely different from the above, check all transistors for heating - if any device is hot, turn off the power immediately, then correct the mistake.

If you do have a suitable bench supply - This is much easier! Slowly advance the voltage until you have about ±20V, watching the supply current. If current suddenly starts to climb rapidly, and voltage stops increasing then something is wrong, otherwise, continue with testing. (Note: as the supply voltage is increased, the output voltage will decrease - down to about 2V, then quickly drop to near 0V. This is normal.)

Once all appears to be well, connect a speaker load and signal source (still with the safety resistors installed), and check that suitable noises (such as music or tone) issue forth - keep the volume low, or the amp will distort badly with the resistors still there if you try to get too much power out of it.

If the amp has passed these tests, remove the safety resistors and re-install the fuses. Disconnect the speaker load, and turn the amp back on. Verify that the DC voltage at the speaker terminal does not exceed 100mV, and perform another "heat test" on all transistors and resistors.

When you are satisfied that all is well, set the bias current. Connect a multimeter between the collectors of Q7 and Q8 - you are measuring the voltage drop across the two 0.33 ohm resistors. The most desirable quiescent current is 75mA, so the voltage you measure across the resistors should be set to 50mV ±5mV. The setting is not overly critical, but at lower currents, there is less dissipation in the output transistors. Current is approximately 1.5mA / mV, so 50mV will represent 75mA quiescent current.

After the current is set, allow the amp to warm up (which it will), and readjust the bias when the temperature stabilises. This may need to be re-checked a couple of times, as the temperature and quiescent current are slightly interdependent. When you are happy with the bias setting, seal the trimpot with a dab of nail polish.

If the temperature continues to increase, the heatsink is too small. This condition will (not might - will) lead to the destruction of the amp. Remove power, and get a bigger heatsink before continuing. Note also that although the power transistors are mounted to the board, never operate the amp without a heatsink - even for testing, even for a short period. The output transistors will overheat and will be damaged.

When all tests are complete, turn off the power, and re-connect speaker and music source.

Power Supply

Before describing a power supply, I must issue this ...

WARNING: Mains wiring must be done using mains rated cable, which should be separated from all DC and signal wiring. All mains connections must be protected using heatshrink tubing to prevent accidental contact. Mains wiring must be performed by a qualified electrician - Do not attempt the power supply unless suitably qualified. Faulty or incorrect mains wiring may result in death or serious injury.

A simple supply using a 25-0-25 transformer will give a peak power of about 75W into 8 ohms, or 60W or so continuous. This is influenced by a great many things, such as the regulation of the transformer, amount of capacitance, etc. For a pair of amps, a 300VA transformer will be enough. Feel free to increase the capacitance, but anything above 10,000uF brings the law of diminishing returns down upon you. The performance gain is simply not worth the extra investment.

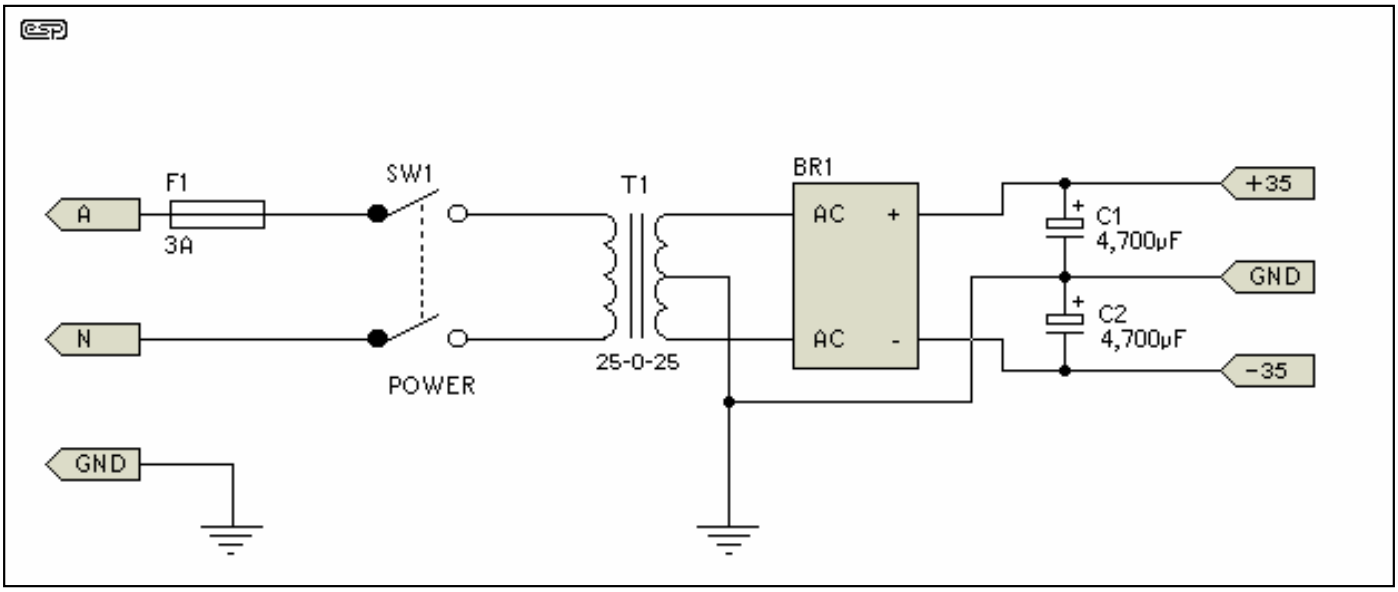


Figure 2 - Recommended Power Supply

For the standard power supply, as noted above I suggest a 300VA transformer. For 115V countries, the fuse should be 6A, and in all cases a slow blow fuse is required because of the inrush current of the transformer.

The supply voltage can be expected to be higher than that quoted at no load, and less at full load. This is entirely normal, and is due to the regulation of the transformer. In some cases, it will not be possible to obtain the rated power if the transformer is not adequately rated.

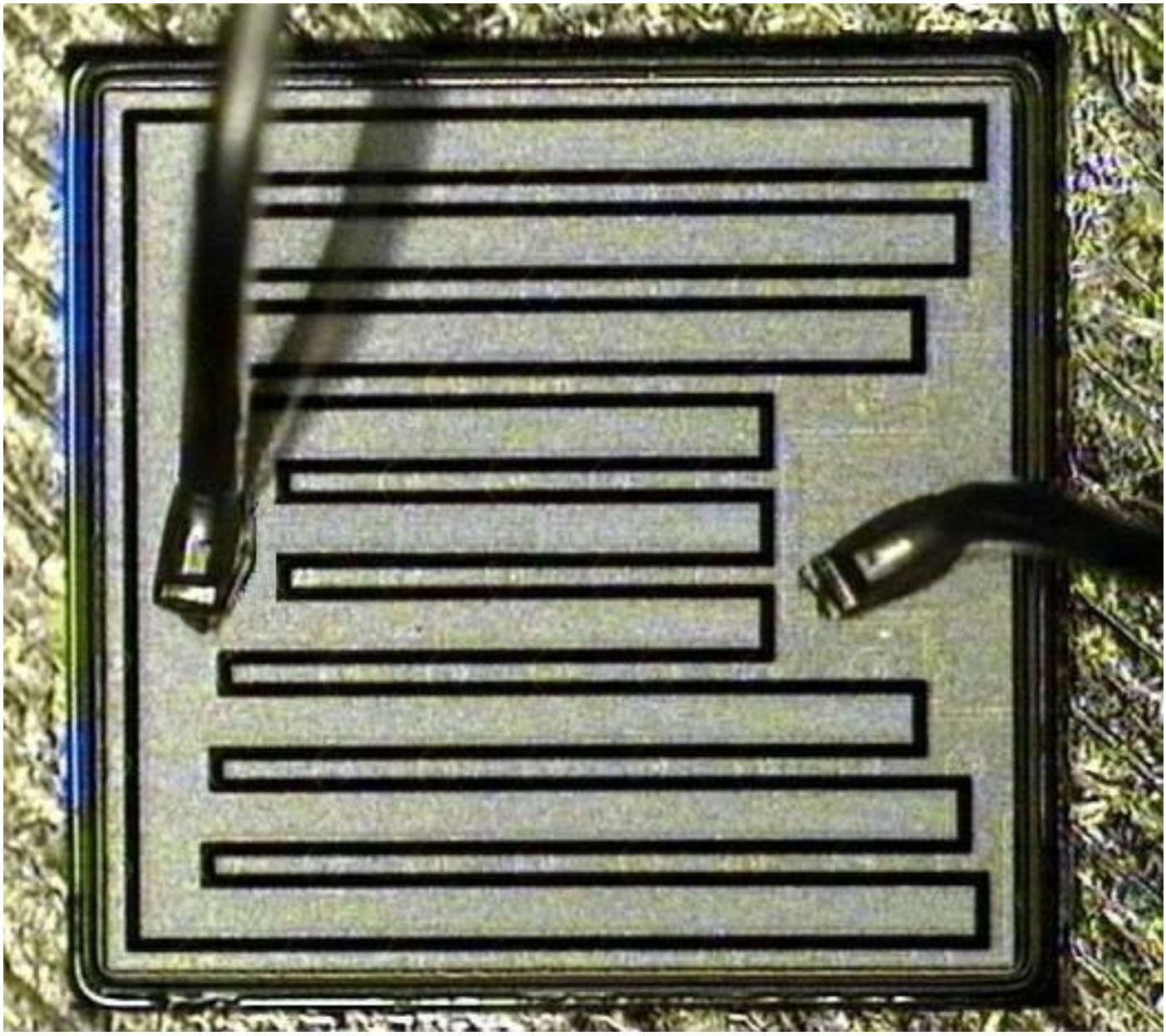
The bridge rectifier should be a 35A type, and filter capacitors must be rated at a minimum of 50V. Wiring needs to be heavy gauge, and the DC must be taken from the capacitors - not from the bridge rectifier.

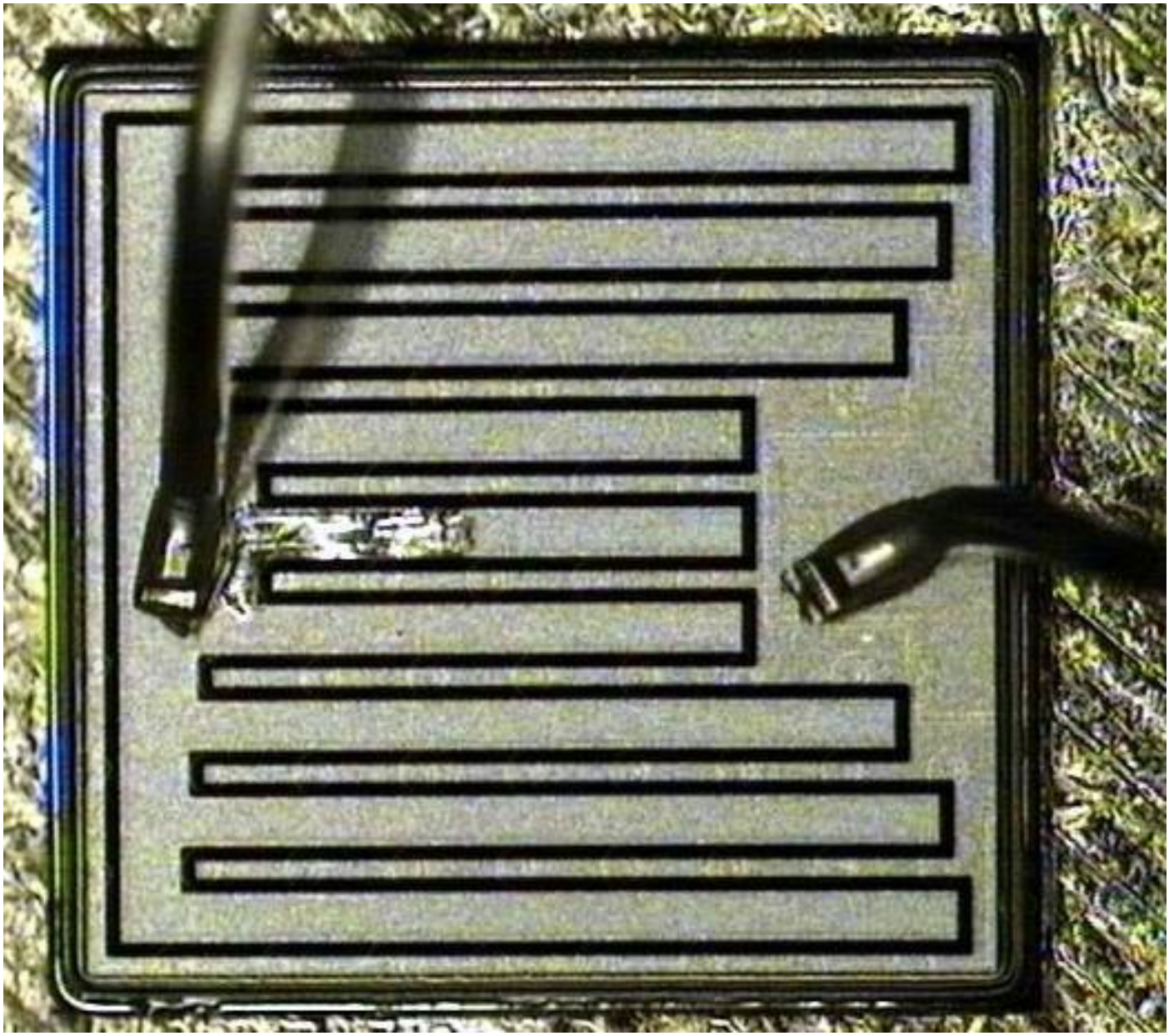
Projects Index

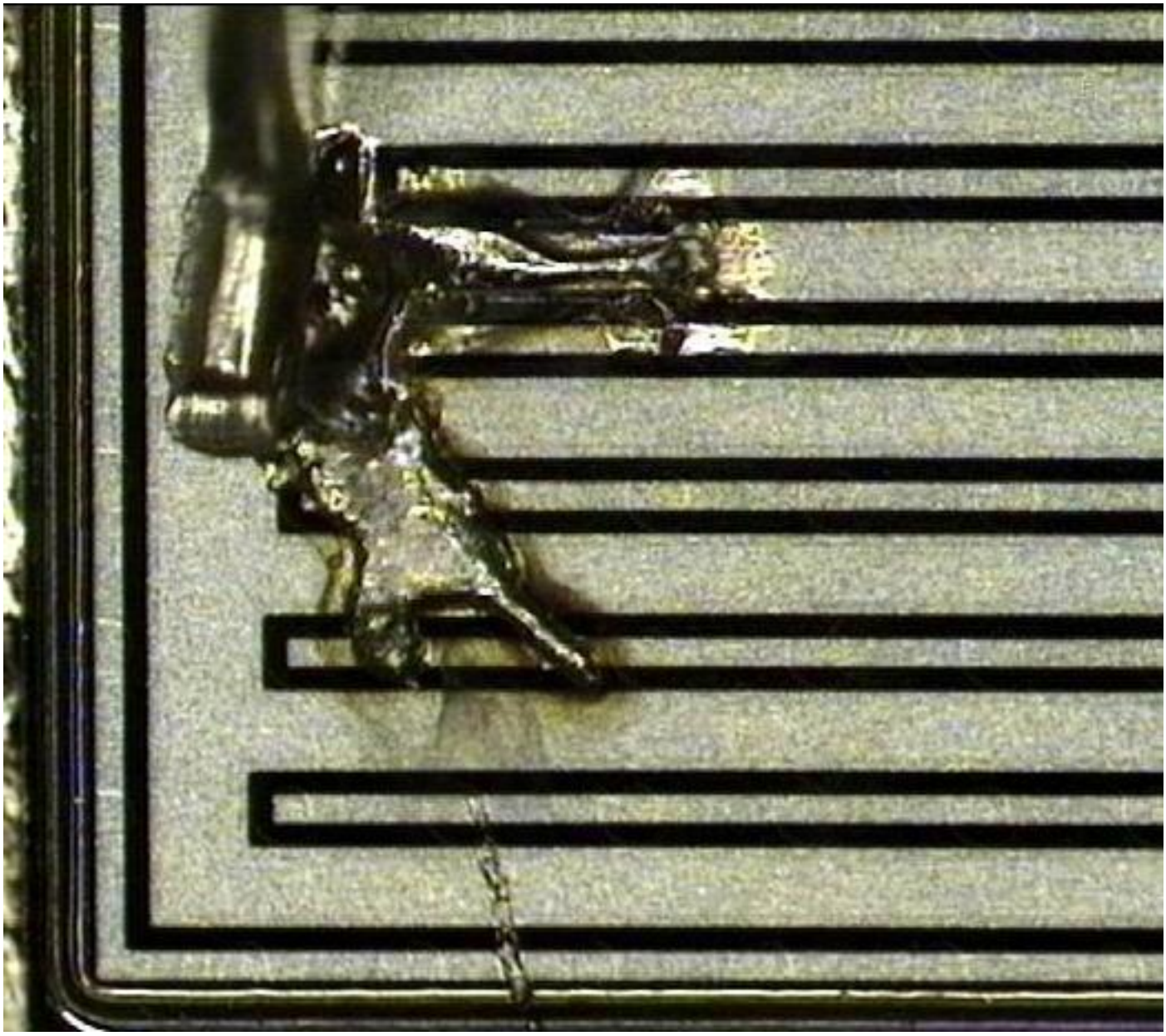
Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000-2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright © Rod Elliott 26 Sep 2000./ Major update 24 Jul 2003







AN1628/D

Understanding Power Transistors Breakdown Parameters

Prepared by: Michaël Bairanzade

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

INTRODUCTION

Among the electrical parameters of a Bipolar power transistor, the breakdown related ones are the most critical to measure. As a matter of fact, as the breakdown voltage can be pretty high, the instantaneous power dissipated during the test must be accurately controlled to avoid a local hot spot on the chip. On the other hand, the breakdown mechanisms are prone to high frequency oscillations and care must be observed to accurately measure this parameter. This application note gives a definition of the breakdown parameters and the associated physics of semiconductor devices. A chapter describes the test techniques and methods recommended to perform this kind of measures.

Since most of the high voltage bipolar transistors are NPN type, all the analysis carried out in this note will be referenced to NPN devices. The PNP mechanisms are essentially the same, with reverse polarities of all biases.

BREAKDOWN MECHANISM

Let us first consider the Collector to Base junction. Like a $p-n$ diode, the avalanche process limits the collector−base voltage the transistor can sustain. When the voltage is large enough, the n−side depletion region will reach the n+ contact diffusion and, if the voltage increases further, the contact will supply electron to the $p-n$ junction. At this point, the Collector−Base junction is essentially shorted and the current is limited by the external resistances.

Basically, three mechanisms control the breakdown: thermal instability, tunneling effect and avalanche multiplication.

THERMAL INSTABILITY

The breakdown for thermal instability is the result of large increase of leakage current when the junction temperature increases. Germanium based devices are more sensitive to this mechanism compare to the Silicon. The process snowball as the leakage current increases, yielding higher junction temperature which can lead to the destruction of the chip if it extends above the melting point of the semiconductor material.

TUNNELING EFFECT

When the electrical field approaches 10^6 V/cm in Silicon, a significant current begins to flow by means of the band to band tunneling process. To get such a high field, the junction must have relatively high doping concentration on both side of the p−n junction. As a matter of fact, the field may be so high that it creates enough force on a covalently bounded electron to free it. This creates two carriers, a hole and an electron to contribute to the current. From an energy−band point of view, in this breakdown process an electron makes a transition from the valence band to the conduction band without the interaction of any other particles.

The breakdown mechanism for silicon device with breakdown voltage lower than 4Eg/q comes from the tunneling effect. For junctions having breakdown voltage in excess of 6Eg/q, the mechanism is generated by avalanche multiplication. At voltages between 4Eg/q and 6Eg/q, the breakdown is a combination of the two effects.

Since the energy band gap in silicon decreases with temperature, the breakdown voltage due to tunneling effect in these semiconductors has a negative temperature coefficient. This temperature effect can be used to differentiate the tunneling effect from the avalanche mechanism which has a positive temperature coefficient.

One must point out that the tunneling effect is dominant for zener voltage in the 5 V to 6 V range. This mechanism is not present for high voltage junctions since the doping concentration is too low to generate the tunneling effect.

AVALANCHE MULTIPLICATION

This mechanism, also named impact ionization, is the most important one in junction breakdown since the avalanche breakdown voltage imposes an upper limit on the reverse bias of the collector voltage. The value at which breakdown occurs depends on the structure of the junction and the dopant concentration used to manufacture the transistor. Both the structure and the epitaxy being controlled, the breakdown voltage of a given semiconductor is reasonably predictable at the design stage.

© Semiconductor Components Industries, LLC, 2003

In this process, free carriers can gain enough energy from the field between collisions to break covalent bonds in the lattice when they collide with it. Consequently, every carrier interacting with the lattice creates two additional carriers and the mechanism snowballs as three carriers can also participate to the collisions. This leads to a sudden multiplication of carriers in the space−charge region when the electrical field becomes large enough to trigger the avalanche.

Let us consider an electron traveling in the space−charge region of a reverse biased pn junction. The electron travels, on average, a distance L (the mean free path) before losing energy by interacting with an atom in the lattice. The energy ΔE gained from the field *E* by the moving electron between collisions is:

$$\Delta E \sum q \int_0^L E.\,dx \qquad (1)$$

Assuming the electron has accumulated enough energy from the field prior colliding with an atom, the bond between the atom core and one of the bound electron can be broken during the collision. This creates three resulting carriers which are free to leave the region of the collision. A simplified schematic representation is given in Figure 1.

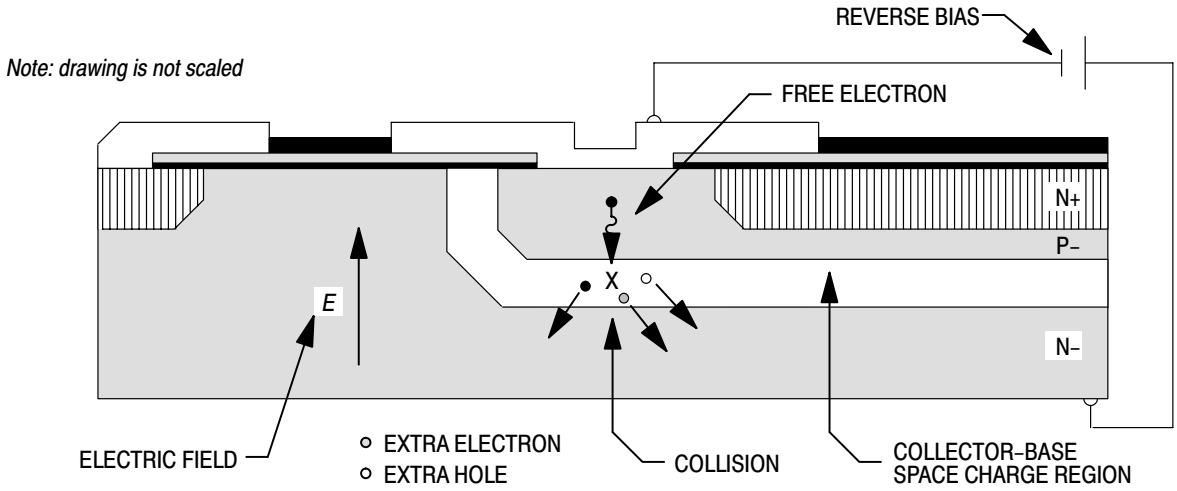
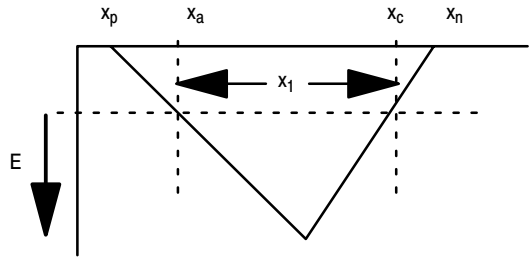
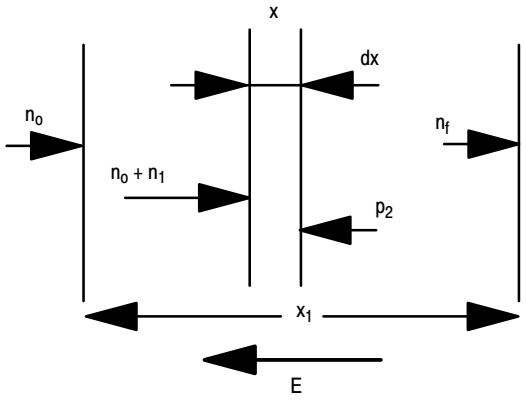


Figure 1. Simplified Representation of the Avalanche Process in a Junction

Near the edge of the space charge region, the electric field is low and practically no carriers can gain enough energy from the field to create a hole−electron pair before they lose their kinetic energy in a collision with the lattice. Consequently, the avalanche is confined to the central area of the space charge region where the field is sizable. This is represented in Figure 2, the central region being identified by x_1.



(a) Ionization occurs in the high field area x_1 of the space charge region



(b) Carriers pair are created in the region dx at x, by electrons flowing from the left and hole flowing from the right

Figure 2. Ionization Process Location for a Non Punch Trough Structure

Let us consider the number of carriers created by avalanche in a small volume of width dx located within x_1 at x (see Figure 2b). Let n_o = density of electrons entering x_1 from the left at x_a.The avalanche phenomenon will increase this density between x_a and x_c in a way that the electrons entering the volume Adx at x from the left have a density $n_o + n_1$. The probability that electrons create electron−hole pairs when travelling through dx is given by the product of the ionization coefficient α**n**, times the length dx. Since the electrons gain energy more rapidly when the field is higher, the ionization coefficient is a function of the electric field, hence of the position in the semiconductor.

In the same manner, the density of holes will increase as a consequence of the avalanche, and the $\alpha_{\mathbf{p}}$ factor represent the ionization coefficient for the holes.

Let n_f be the density of electrons that reaches x_c:

$$n_f \sum n_o + n_1 + n_2 \tag{2}$$

where n_2: density of electrons created between x and x_c.

Since electrons and holes are created in pair, $n_2 = p_2$, we may write:

$$\frac{dn}{dx} \sum (\alpha_n - \alpha_p)(n_o + n_1) + \alpha_p n_f \tag{3}$$

Assuming that $\alpha = \alpha_n = \alpha_p$, we can rewrite equation 3 as follows:

$$n_f - n_o \sum n_f \int_{xa}^{xc} \alpha \, dx \tag{4}$$

The ratio of the density of electrons n_f leaving the space−charge region to the density n_o entering is called the multiplication factor **M**:

$$\mathbf{M} \sum \frac{n_f}{n_o} \sum \frac{1}{1 - \int_{xa}^{xc} \alpha \, dx} \tag{5}$$

As the integral in the denominator of equation 5 approaches unity, the **M** factor increases to infinity. Therefore, the avalanche will occur when:

$$\int_{xa}^{xc} \alpha \, dx \sum 1 \tag{6}$$

The density of ionizing collisions at x is proportional to **n*** which represent the density of excited electrons arriving at x with enough energy to create the electron−hole pair. The density **n***, in turn, is the total electron density **n** times the probability that an electron has not collided in a distance **d** necessary to get the right energy:

$$\mathbf{n^*} \sum \mathbf{n} \exp\left(-\frac{\mathbf{d}}{L}\right) \tag{7}$$

The length **d** may be derived from equation 1 by letting E_1 being the minimum energy necessary to create the ionization process, and *E* being the average field that accelerates the electron:

$$d \sum \frac{E1}{qE} \tag{8}$$

Since the ionization coefficient depends strongly on the electric field, the multiplication factor also increases rapidly with field. Consequently, a small increase in field, as the voltage approaches breakdown, causes a sharp increase in current as observed on a curve tracer. Not only does the ionization coefficient vary with field, hence with the position, in the space charge region, but the width of the space charge region varies also with voltage. Therefore, the evaluation of **M** with equation 5 is difficult and a good approximation can be derived from the simplified equation 9:

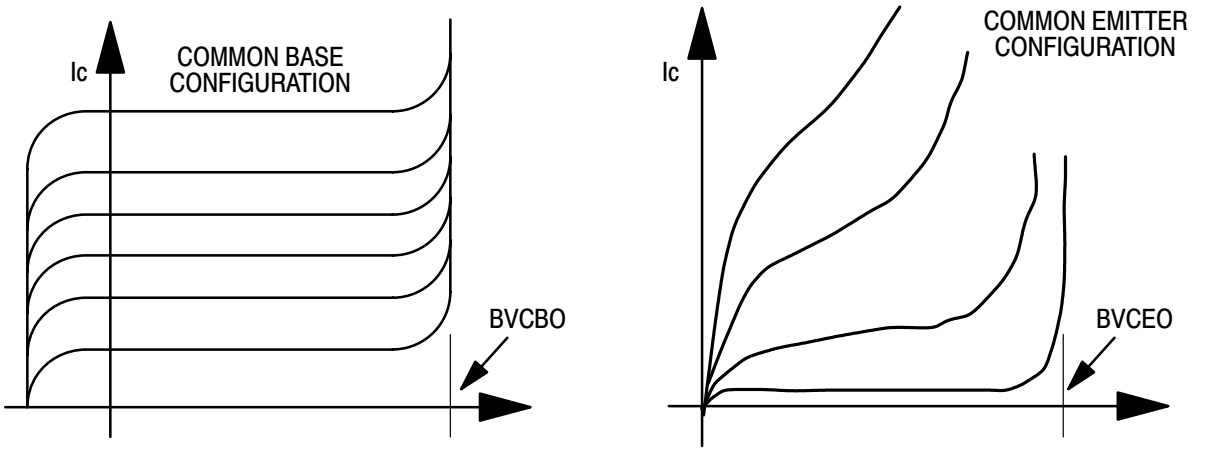
$$\mathbf{M} \sum \frac{1}{1 - \left(\frac{Vr}{Bv}\right)^n} \tag{9}$$

with $2 < n < 6$

Vr: the reverse voltage applied across the junction

Bv: the breakdown voltage.

In the common Base configuration, the breakdown due to the impact ionization (avalanching) gives a well defined voltage as depicted in Figure 3a, the effect of the current coming from the emitter having little effect on the breakdown.



(a) Under breakdown, the emitter current is not affected

(b) At low impact ionization rate, the base current starts to increase, yielding an emitter current increase and, eventually, a collector current runaway

Figure 3. Common Base and Common Emitter Breakdown Typical Curves

For the common Emitter configuration, the breakdown is not as sharply reflected in the transistor output characteristics as depicted in Figure 3b, and occurs at a lower voltage BVCEO. In this configuration, as soon as the impact ionization process starts, the secondary holes are injected into the base, yielding a base current which, in turns, leads to an increase in the emitter current and a current runaway may eventually occurs as the mechanism snowballs.

SECOND BREAKDOWN

The use of power transistor is limited by the second breakdown phenomenon which is an abrupt decrease of the device voltage capability with a simultaneous internal constriction of current. Figure 4 shows a typical Ic/Vc characteristic of a transistor under second breakdown. The avalanche breakdown (also named first breakdown) occurs when the voltage applied across the junction reaches the value specific for this device as discussed above. As the voltage is increased, the second avalanche occurs.

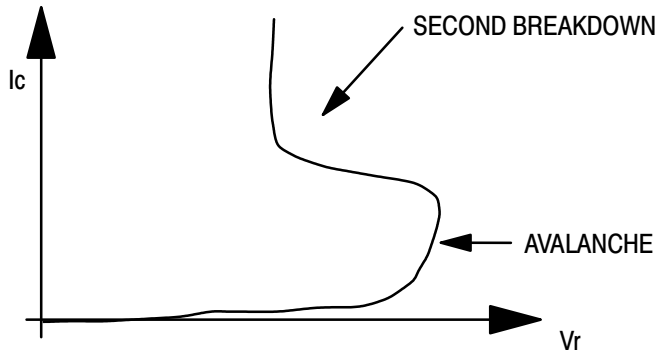


Figure 4. Typical Second Breakdown Characteristic

The initialization of this mechanism is essentially caused by the temperature effect. After a time delay following the power pulse P = Ic*BVCEO applied to the transistor, the device goes into the second avalanche area. During the time elapsed from avalanche to second breakdown, the junction is unstable and may be rapidly downgraded. In particular, the resistance of the breakdown spot becomes extremely low. During the third part of the curve, when the voltage is significantly reduced, the chip is at high temperature and the semiconductor is in the intrinsic mode near the breakdown spot location. If the current keeps increasing, the silicon melts and the transistor is permanently destroyed.

BASIC POWER TRANSISTOR GEOMETRY

The simplified cross section of a power transistor given in Figure 5 illustrates the main areas involved in the

breakdown process. When the voltage rises across the transistor, the electrical field is sustained by the N−epitaxy, the depletion region increasing accordingly. One must point out that, in reality, a modern transistor is not built with a single junction as depicted in Figure 5, but with several emitter stripes diffused into the base P−. Such technology is mandatory to get the expected static and dynamic characteristics requested in today's applications: the photo given in Figure 6 illustrates this concept.

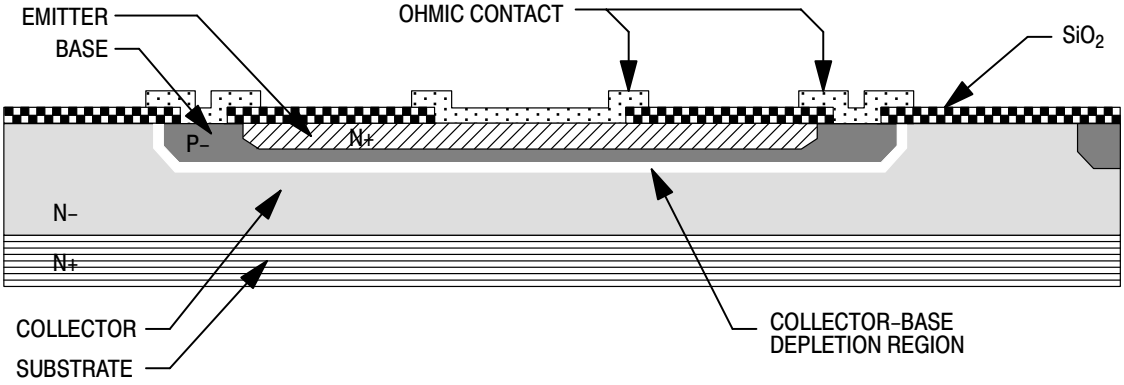


Figure 5. Simplified Cross Section of a Bipolar Power Transistor

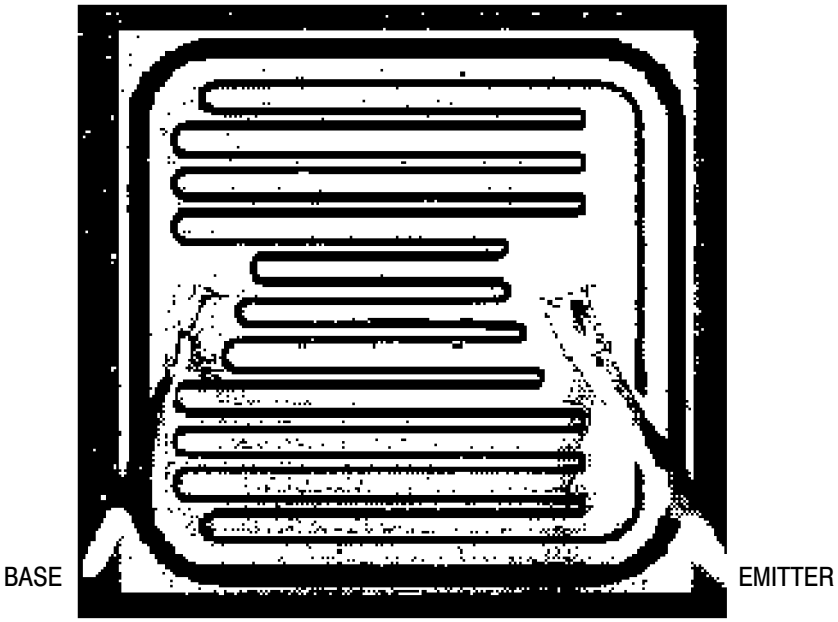


Figure 6. Photo of a Typical Power Transistor Using the PLANAR Termination

On top of that, the electrical field must not collapse at the edge of the chip, where the base diffusion no longer exists. To prevent this mode of failure, the transistors are designed with a junction termination suited for the kind of device under development. A junction termination example is given in Figure 7

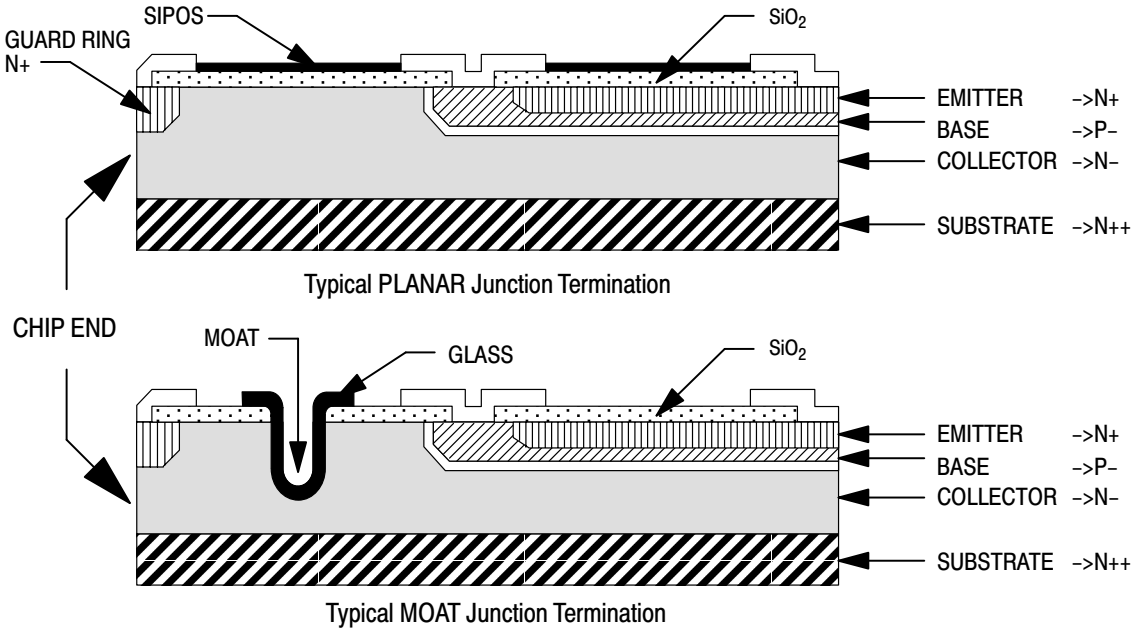


Figure 7. Typical Junction Terminations for a High Voltage Bipolar Transistor

At the time of printing this document, the PLANAR process is the most widely used technology to manufacture power transistors. However, the MOAT termination is still used for very high voltage device with BVCBO of 1500 V and above.

DYNAMIC BREAKDOWN CHARACTERISTICS

Although the static breakdown characteristics are valuable to define the performances of a semiconductor device, they do not provide all the informations needed to safely use a power transistor. Leaving aside the switching parameters, the dynamic behaviour of a power transistor is bounded by three main parameters:

FBSOA: Forward Bias Safe Operating Area

RBSOA: Reverse Bias Safe Operating Area

BVsus: Breakdown Sustaining

The designer of a given application must carefully check that, under no circumstances, the transistor will operate out of the curves provided in the data sheet: severe damage, or destruction, occurs if the operating point moves outside the specified limits.

FBSOA

The FBSOA curves are specified to define the maximum safe operating area for a given device. This parameter depends on the chip geometry and the process used to built the transistor. Particularly, the emitter fingers perimeter is extremely important since it highly influences the current spread across the silicon. As a matter of fact, at turn on, the current starts to flow from the periphery of the emitter finger, as depicted in Figure 8 The rbb' increases as we move toward the center of the emitter finger and the Base−Emitter junction is not fully forward biased during the first microseconds. Consequently, the current density is maximum at the edge of the junction and the emitter becomes prone to local hot spots.

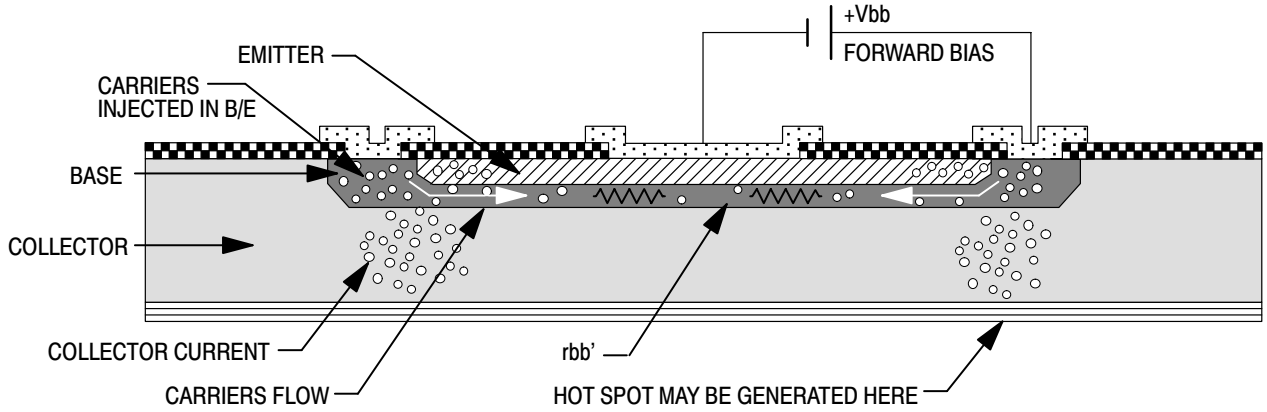


Figure 8. Turn on Mechanism for a Bipolar Power Transistor

The data provided for the FBSOA curve are based on a maximum junction temperature of +150°C and is bounded by four points as shown in Figure 9:

BVCEO: maximum allowable voltage
Icm: maximum allowable collector current
Second breakdown
Pm: maximum power dissipation
Pw: pulse width

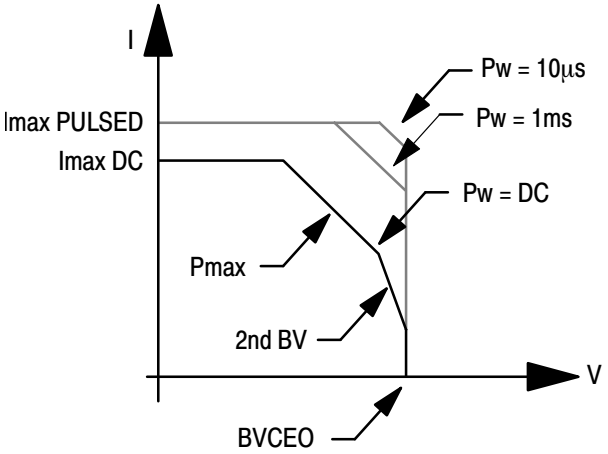


Figure 9. Typical FBSOA Curve

Since this parameter is temperature dependant, for a constant power, the FBSOA depends upon the pulse width. When the pulse is very short, below 1μs, the curves can be extended above the BVCEO limit for transistors specifically designed for switching applications.

The typical FBSOA test circuit is given in Figure 10 Cares must be observed to avoid overheating the silicon during the test. Of particular importance are the short pulses which have a high energy content since such pulses can generate a local hot spot within the silicon chip, even though the case temperature stays at ambient temperature.

Since the FBSOA test can damage the transistor, one must not re−use devices exposed to a characterization procedure. On the other hand, using conventional power supplies to bias the device under test is not recommended since they do not have the capability to accurately control either the pulse width or the current amplitude. It is preferable to use commercial equipment designed for that purpose, or to built a specific test jig, with appropriate dynamic performances, to evaluate a limited number of devices.

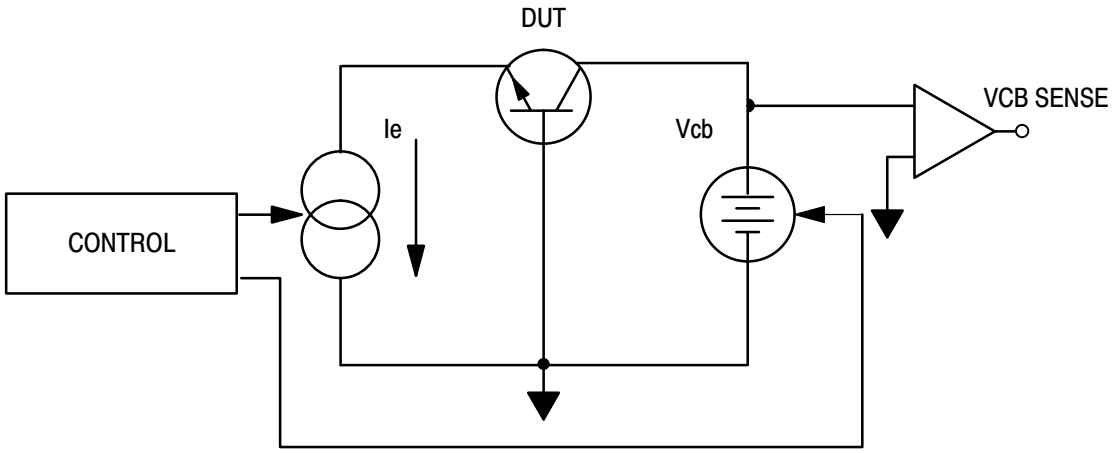


Figure 10. Typical Forward Bias Safe Operating Area Test Circuit

RBSOA

The RBSOA curves are used to define the maximum current/voltage a transistor can sustain under Base−Emitter reverse bias. As depicted by the typical test circuit given in Figure 11, this parameter is essentially an analysis of the breakdown, under dynamic condition, when the device is loaded by an inductor.

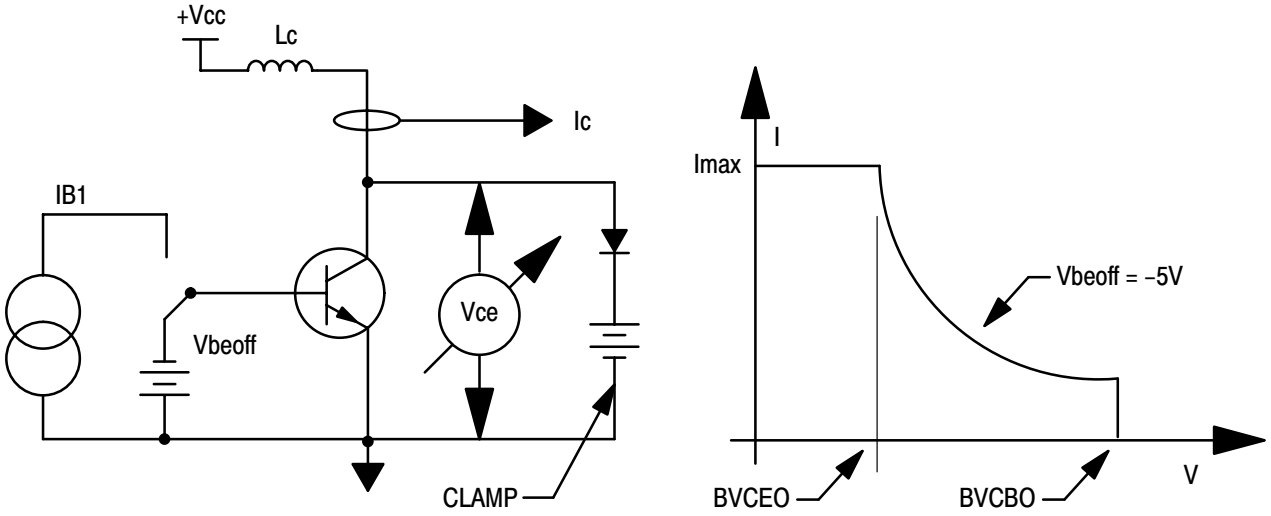


Figure 11. Typical Reverse Bias Safe Operating Area Test Circuit

The test is normally performed with a clamp connected across emitter−collector to limit the voltage within the BVCES/BVCEX of the device under test. In this case, the transistor is not avalanched and most of the energy stored into the inductor is dissipated in the clamp circuit. As a matter of fact, when a voltage clamp is used, the transistor sustains the current−voltage stress only during its collector current fall time.

The same parameter can be evaluated without providing a voltage clamp: in this case, the transistor will be forced in the avalanche region (assuming L*(dI/dt) > BVCES) and all the energy accumulated in the inductor will be dissipated into the device under test. This condition forces much more stress into the silicon as the current will flow during a longer time than the collector current fall time (all of the $1/2L*I^2$ energy must be dissipated) under avalanche condition. Extreme care must be observed when performing such a test since the transistor is very rapidly destroyed if the current is not carefully limited by the external circuit. As a standard rule, the test is carried out by slowly increasing the voltage clamp until the transistor goes into avalanche, the current being limited by the pulse width applied during the charging time of the inductor.

The RBSOA test is sometime named sustaining breakdown (BVCEsus) when it is performed unclamped, or inductive switching when the collector voltage is clamped below BVCES.

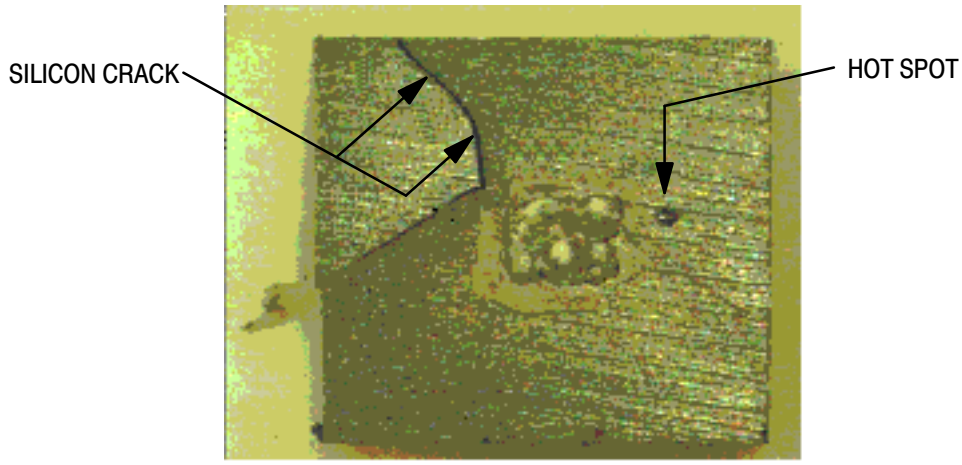


Figure 12. (a) Microphotography of a Power Transistor Damaged by a RBSOA

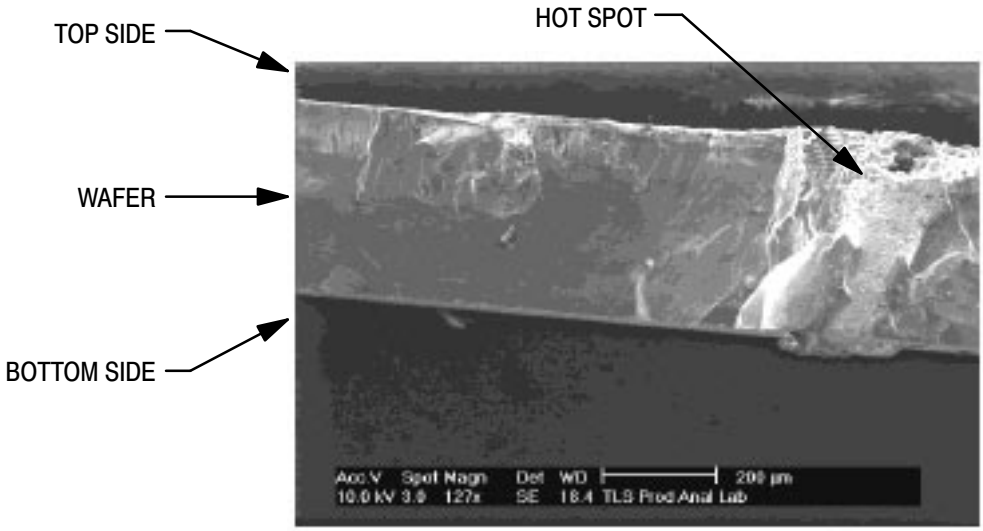


Figure 12.(b) Cross Section of the Damaged Chip (given Fig. 12a) Showing how the Silicon Melts.

POWER DEVICE BEHAVIOUR

The reverse bias characteristics of a power transistor are split into two families: the leakage current and the breakdown.

Generally speaking, the leakage current tests do not generate high stress into the chip, unless the transistor under test has a voltage capability well below the voltage bias. As a matter of fact, since the bias voltage is defined prior being forced across the junction, there is no snap back in the characteristic and the test circuit remains stable. On top of that, leakage currents are in the micro ampere range, for modern silicon power transistors, and the power generated during the test cannot damage the chip. As an example, let us assume an ICEO test performed at 400 V, the maximum current being 100 μA, yielding 40mW in the junction. This value is well within the power handling capability of the transistor and cannot damage the device.

However, although they provide reliable informations, the leakage tests characterize neither the breakdown voltage nor the avalanche mode of a bipolar transistor. These behaviour are analyzed by means of the safe operating area tests as discussed above.

When the transistor is tested in BVCEO, the current/voltage characteristic exhibit a snap back effect when the Collector/Emitter voltage jumps from the breakdown to the test point point defined at a higher current. This phenomenon, illustrated in Figure 13, is prone to heavy uncontrolled oscillation due to the negative impedance which is developed from point A to point B.

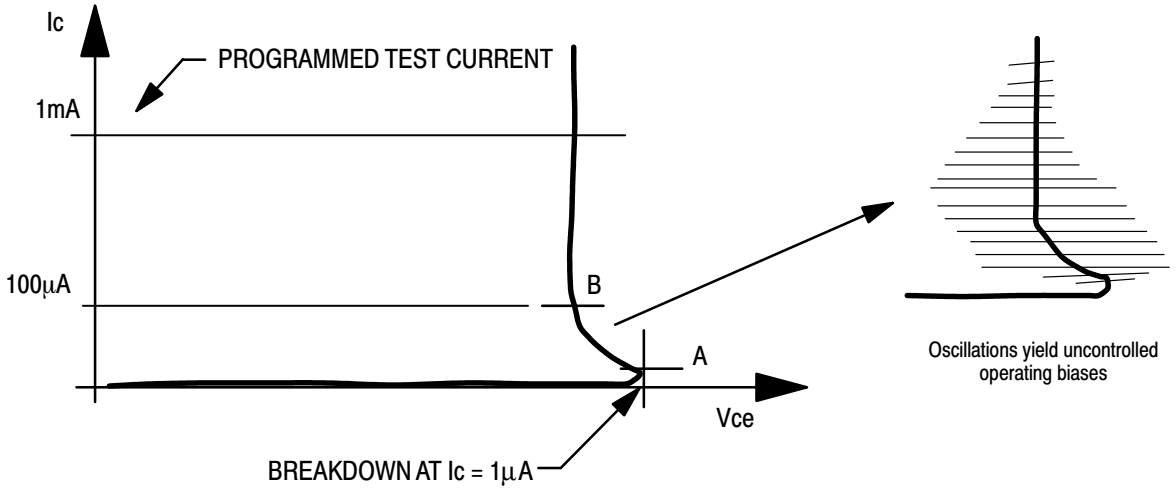


Figure 13. Typical BVCEO Characteristic

Such behaviour can be observed with a curve tracer by limiting the power dissipated into the transistor.

The power transistors are characterized, among others parameters, by a set of breakdown curves as depicted in Figure 14 Under no condition can the transistor be exposed to a voltage−current above the maximum rating provided in the data sheets.

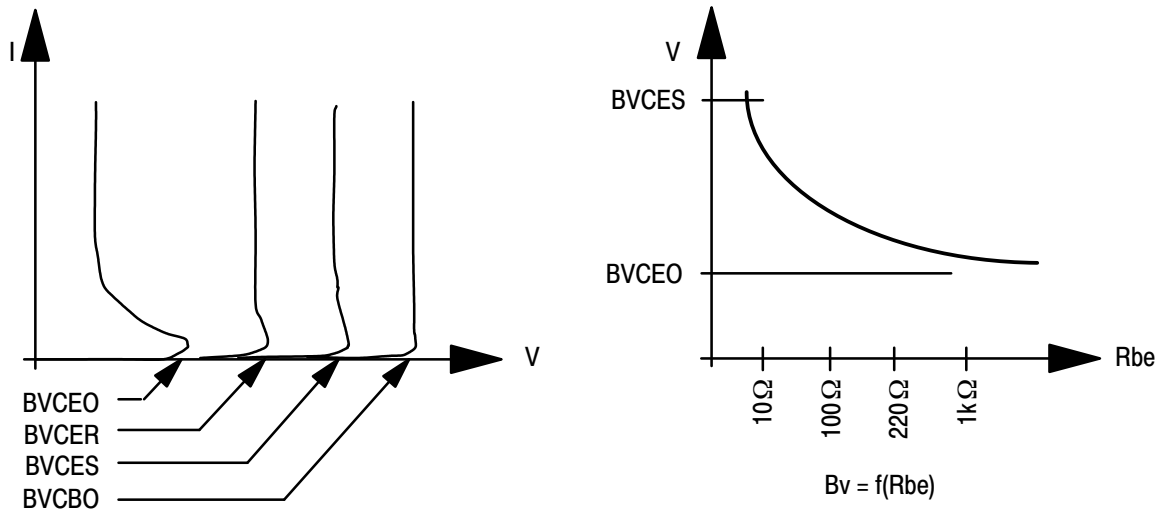
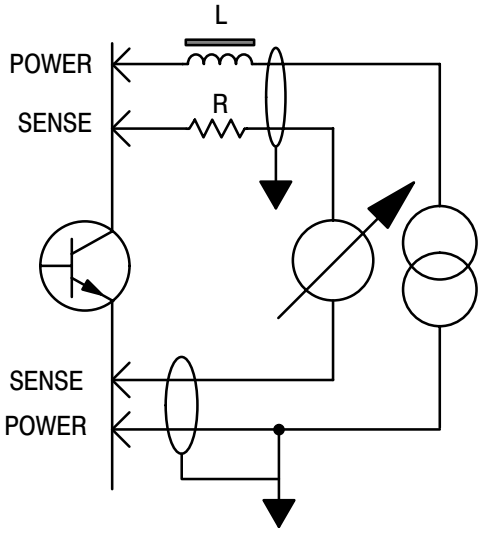
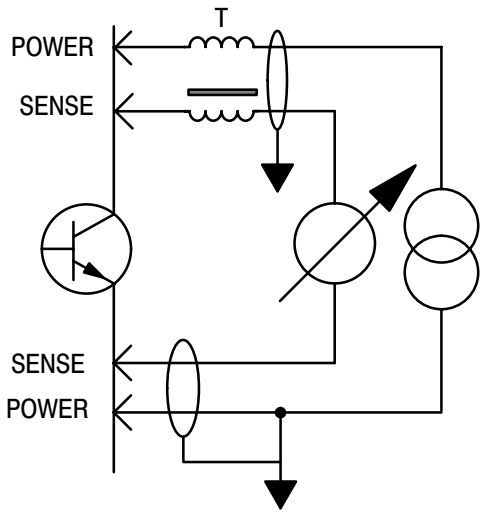


Figure 14. Typical Breakdown Curves (BVEBO not shown)

Because the current source used to perform the breakdown test has a high voltage compliance (in fact, the open load voltage must be higher than the maximum breakdown voltage one expects to characterize), it is not easy to make it oscillation free when the load is very variable. The problem can be cured by using anti oscillation networks right across the transistor electrodes (see Figure 15), but the device can be damaged or destroyed if the current rises above the maximum value allowable for a given chip.



The values of R and L are adjusted for a given transistor and test conditions. Typical values are R=470 Ω and L=100 μH. All the network must be located as close as possible to the transistor under test.
Kelvin contacts are mandatory. Shielding cables are recommended.

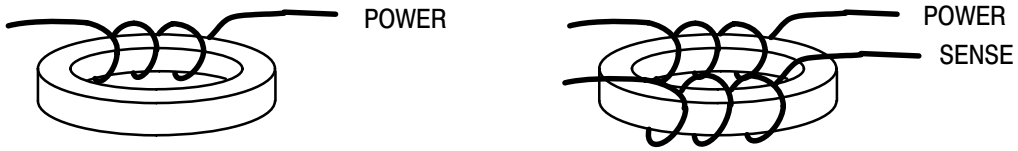


The transformer T1 provides an AC feedback to the current source and is adjusted for a given set of test conditions.
Typically, the primary inductance is 100 μH with Np/Ns=1.

Figure 15. Typical Anti Oscillation Networks

It is also possible to use a capacitor, connected across Collector−Base of the transistor under test, to generate a local AC feedback, although this is not recommended. Keeping in mind that the current source associated with the transistor is a high gain transconductance amplifier, any LC network connected across it generates high frequency oscillations that will make the BVCEO test impossible. Since one cannot define accurately the parasitic inductances and stray capacitances existing in the test circuit, it is preferable not to use a capacitive feedback across the transistor.

The inductor L and the transformer T can be built around a toroid using a high permeability ferrite material. A typical application is given in Figure 16



Toroid: 16x20
Ferrite: N27, B50, A4A
The transformer may require a larger core to accommodate the primary and secondary turns. Using a high permeability material limits the number of turns needed to get the expected inductance.

Figure 16. Typical Inductor Design

One must point out that the design of such an anti−oscillation network very much depends on the test system used to measure the breakdown. Consequently, it's not easy to modelize an accurate circuit and several attempts are usually necessary to tune the network for a given environment.

Another issue with the BVCEO test is the discharge of stray capacitances into the transistor. When the tests are carried out with an automatic instrument, the cables can be several meter in length and the stray capacitance becomes non negligible and are charged up to BVCEO when the current source is connected to the device. Since this source has a finite settling time (i.e., the current does not stabilize to the programmed value in less than a few tens of microseconds!), the current coming from the discharge of the stray capacitance cannot be controlled and may damage the chip within a few micro second. This is particularly true for the BVCEO test as depicted in Figure 17: during the negative going slope, the Collector/Emitter voltage collapses, discharging the

stray capacitance by the delta V from A to B, yielding a peak current above the programmed value.

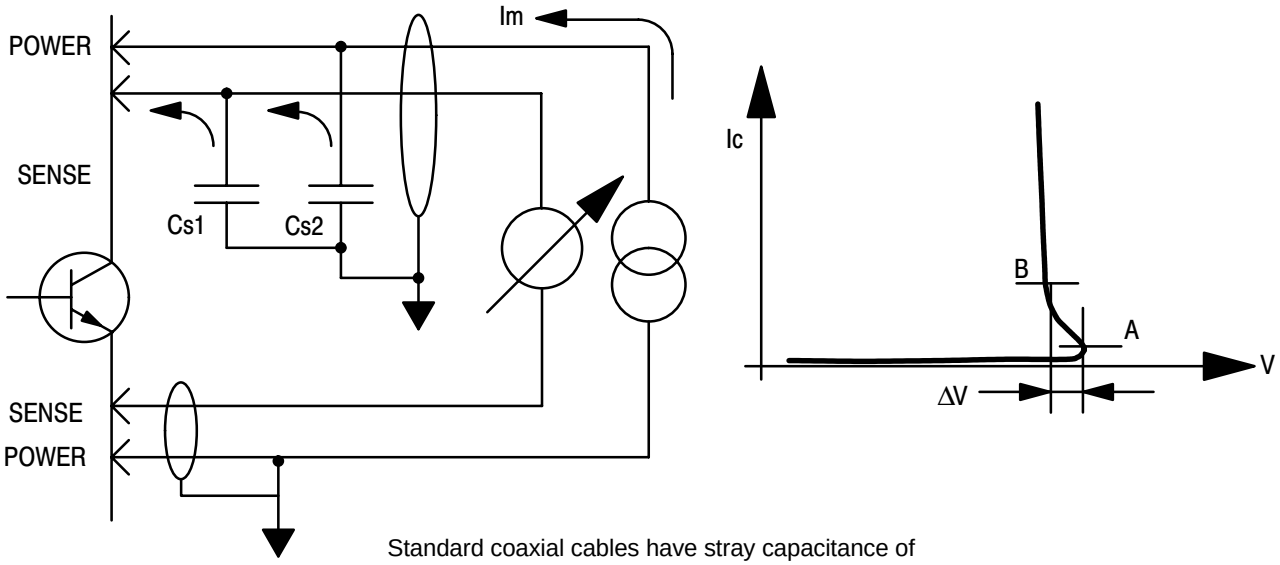


Figure 17. Stray Capacitances Effect

Assuming a $\Delta\beta V$ of 60 V and a total series resistance of 10 Ω, the peak current flowing from the stray capacitances Cs1/Cs2 as depicted in Figure 16, is 6 A. On the other hand, most of the automatic test handler are connected to the tester by coaxial cables up to five meters long. In this case, the total stray capacitance (combining Power and Sense) is 1000 pF, yielding a time constant of:

$Pw = 5*\Sigma Cs*\Sigma R$

$Pw = 5*1000*10^{-12}*10$

$Pw = 50ns$

During this amount of time, the transistor must dissipate the energy accumulated in Cs1 & Cs2:

$Ej = \Sigma Cs*\Delta V^2*0.5$

$Ej = 1000*10^{-12}*60*60*0.5$

$Ej = 1.8\mu J$

This value cannot damage the transistor but, if the circuit oscillates, the $\Delta\beta V$ can be much higher and, for a high voltage power transistor, the energy can rise to a value above the maximum capability of the junction. As an example, let us assume $\Delta\beta V$ = 500 V, then Ej = 125 μJ. Of course, a power transistor can easily handle such amount of energy, but the same transistor can be rapidly damaged when the energy is generated during the transient phase of the breakdown mechanism. As a matter of fact, during uncontrolled oscillation, most of the energy developed in the circuit will be dissipated in a very small area of the die, yielding a very high current density which, in turn, generates a hot spot in the silicon. Ultimately, the silicon may melt and the device is destroyed. This is a consequence of the multiplication factor **M** as discussed above.

Consequently, it's recommended to reserve the breakdown test for engineering and characterization purpose, using the leakage current test to perform, incoming inspection. However, the breakdown parameters are fully tested, under several biases and case temperatures, during the development of a new power transistor. Regular samplings also provide statistics, at production level, to make sure the devices are fully within the guaranteed limits published in the data sheet.

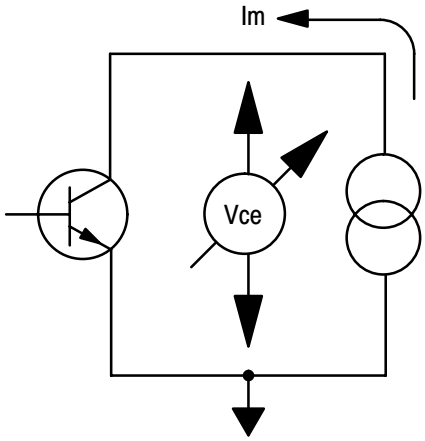
PARAMETERS DEFINITIONS

The off condition parameters are split into two families: the leakage currents and the breakdown voltages. Both are characterized during the development of a transistor and tested during the final test process. The voltage spans from a low 6.00 V for the Bvebo, to a high 1800 V for the Bvcbo, the leakage currents ranging from nano ampere to one milli ampere at room temperature.

BREAKDOWN VOLTAGES

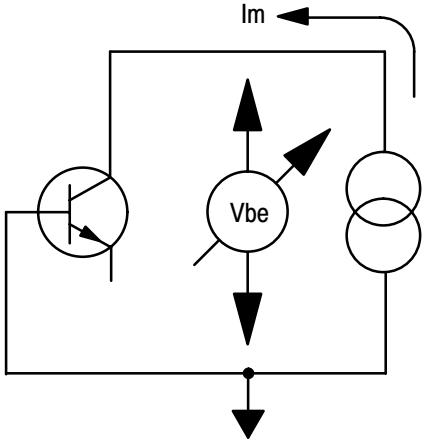
BVCEO: reverse Collector to Emitter voltage, with the Base open, under a given collector current bias.

Im: programmed
Vce: measured



BVCBO: reverse Collector to Base voltage, with the Emitter open, under a given collector current bias.

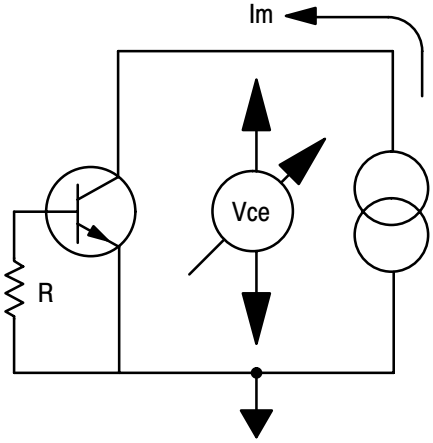
Im: programmed
Vcb: measured



BVCER: reverse Collector to Emitter voltage, the Base connected to the Emitter with a low ohm resistor, under a given Collector current bias.

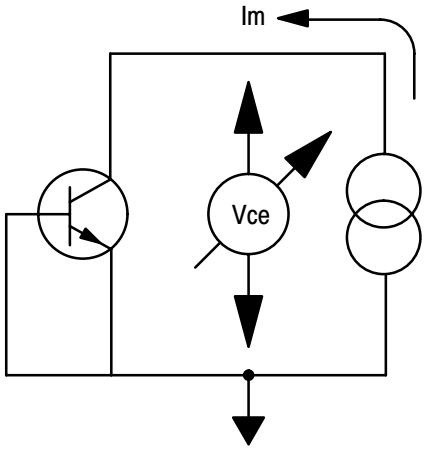
Im: programmed
Vce: measured

Resistor R must be located as close as possible of the device under test.



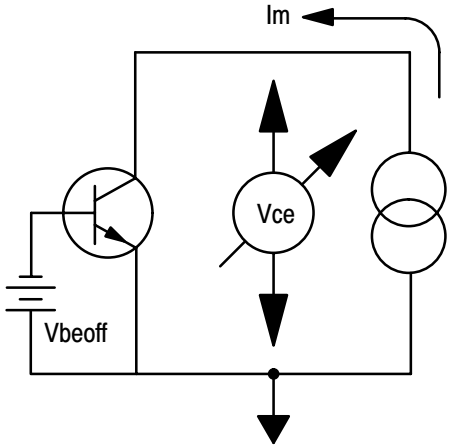
BVCES: reverse Collector to Emitter voltage, the Base shorted to the Emitter, under a given Collector current bias.

Im: programmed
Vce: measured



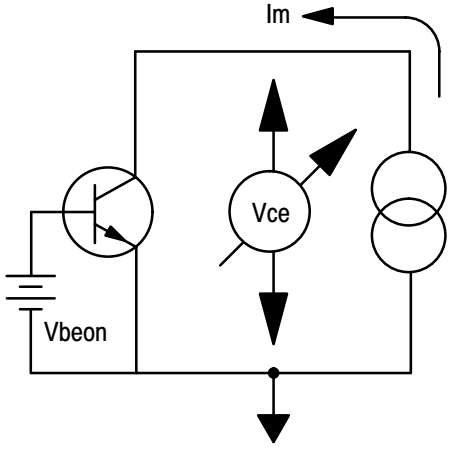
BVCEX: reverse Collector to Emitter voltage, with a reverse Base to Emitter bias, under a given Collector current bias.

Im: programmed
Vce: measured



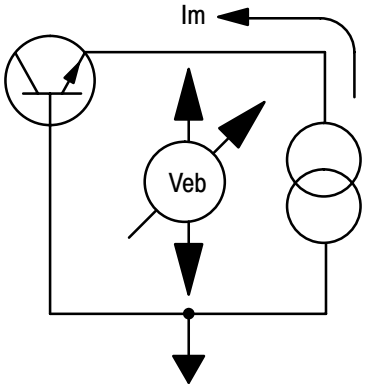
BVCEY: reverse Collector to Emitter voltage, with a forward Base to Emitter bias, under a given Collector current bias.

Im: programmed
Vce: measured



BVEBO: reverse Emitter to Base voltage, with the Collector open, under a given Emitter current bias.

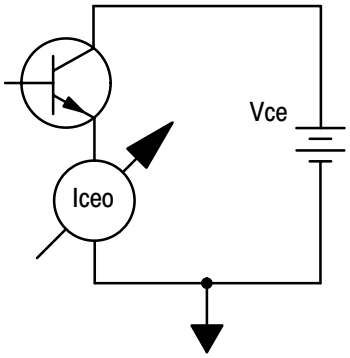
Im: programmed
Vce: measured



LEAKAGE CURRENTS

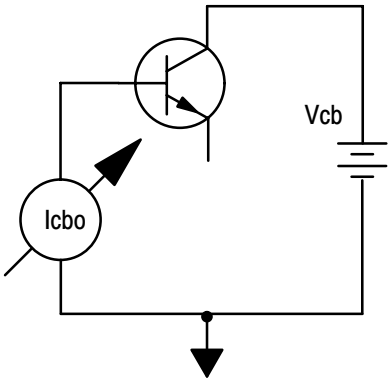
ICEO: Collector to Emitter current under reverse Collector to Emitter voltage, Base open.

Vce: programmed
Iceo: measured



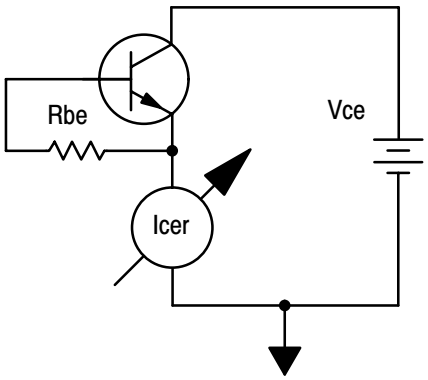
ICBO: Collector to Base current under reverse Collector to Base voltage, Emitter open.

Vcb: programmed
Icbo: measured



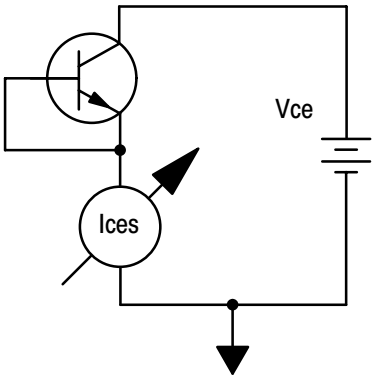
ICER: Collector to Emitter current under reverse Collector to Emitter voltage, Base connected to Emitter by a low ohm resistor.

Vce: programmed
Icer : measured.



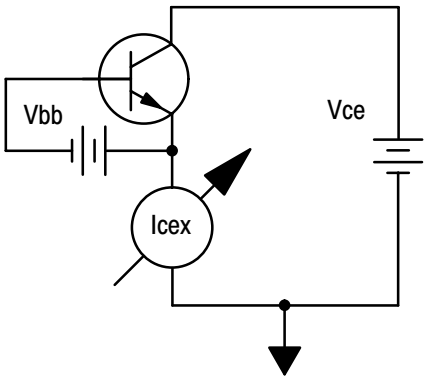
ICES: Collector to Emitter current under reverse Collector to Emitter voltage, Base shorted to Emitter.

Vce: programmed
Ices: measured



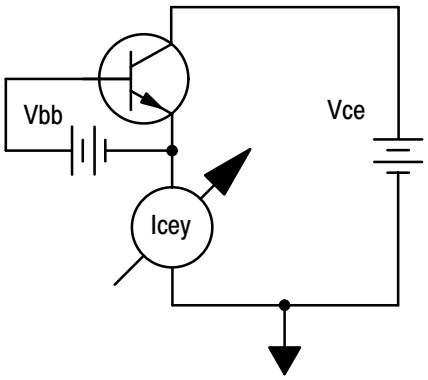
ICEX: Collector to Emitter current under reverse Collector to Emitter voltage, with a reverse Base to Emitter bias.

Vce: programmed
Icex: measured



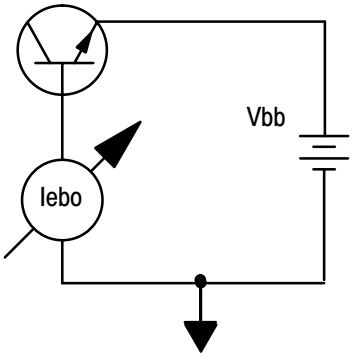
ICEY: Collector to Emitter current under reverse Collector to Emitter voltage, with a forward Base to Emitter bias.

Vce: programmed
Icey: measured



IEBO: Emitter to Base current under reverse Emitter to Base voltage, Collector open.

Vbb: programmed
Iebo: measured



ANNEXES

Symbols

E:	Electrical field
E_1:	minimum energy for ionizing collision
E_g:	Band Gap energy
q:	electron charge
n:	electron density
np:	excess electron density (p region)
L:	mean free path
M:	avalanche multiplication factor
α_n:	ionization coefficient for electrons
α_p:	ionization coefficient for holes

Breakdown Electrical Fields in Semiconductors

Material	Bandgap (eV)	Breakdown Electrical Field (V/cm)
GaAs	1.43	$4*10^5$
Ge	0.664	10^5
InP	1.34	–
Si	1.10	$3*10^5$
$In_{0.53}Ga_{0.47}As$	0.80	$2*10^5$
C	5.50	10^7
SiC	2.9	$2.5*10^6$
SiO_2	9	10^7
Si_3N_4	5	10^7

BIBLIOGRAPHY

Jasprit SINGH: “Semiconductor Devices, an introduction’’, McGRAW−HILL International

Richard S. MULLER & Theodore I. KAMINS: “Device Electronics for Integrated Circuits’’, second edition, Wiley International

S. M. SZE: “Physics of Semiconductor Devices’’, second edition, Wiley International

P. LETURCQ & G. REY: “Physique des composants actifs à semicondcuteurs”, Dunod Université

ON Semiconductor and are registered trademarks of Semiconductor Components Industries, LLC (SCILLC). SCILLC reserves the right to make changes without further notice to any products herein. SCILLC makes no warranty, representation or guarantee regarding the suitability of its products for any particular purpose, nor does SCILLC assume any liability arising out of the application or use of any product or circuit, and specifically disclaims any and all liability, including without limitation special, consequential or incidental damages. “Typical” parameters which may be provided in SCILLC data sheets and/or specifications can and do vary in different applications and actual performance may vary over time. All operating parameters, including “Typicals” must be validated for each customer application by customer’s technical experts. SCILLC does not convey any license under its patent rights nor the rights of others. SCILLC products are not designed, intended, or authorized for use as components in systems intended for surgical implant into the body, or other applications intended to support or sustain life, or for any other application in which the failure of the SCILLC product could create a situation where personal injury or death may occur. Should Buyer purchase or use SCILLC products for any such unintended or unauthorized application, Buyer shall indemnify and hold SCILLC and its officers, employees, subsidiaries, affiliates, and distributors harmless against all claims, costs, damages, and expenses, and reasonable attorney fees arising out of, directly or indirectly, any claim of personal injury or death associated with such unintended or unauthorized use, even if such claim alleges that SCILLC was negligent regarding the design or manufacture of the part. SCILLC is an Equal Opportunity/Affirmative Action Employer.

PUBLICATION ORDERING INFORMATION

Literature Fulfillment:
Literature Distribution Center for ON Semiconductor
P.O. Box 5163, Denver, Colorado 80217 USA
Phone: 303−675−2175 or 800−344−3860 Toll Free USA/Canada
Fax: 303−675−2176 or 800−344−3867 Toll Free USA/Canada
Email: orderlit@onsemi.com

N. American Technical Support: 800−282−9855 Toll Free USA/Canada

JAPAN: ON Semiconductor, Japan Customer Focus Center
2−9−1 Kamimeguro, Meguro−ku, Tokyo, Japan 153−0051
Phone: 81−3−5773−3850

ON Semiconductor Website: http://onsemi.com

For additional information, please contact your local
Sales Representative.

Elliott Sound Products — Cables, Interconnects & Other Stuff - Part 2

Rod Elliott - Copyright (c) 1999/ 2000/ 2001/ 2002

Page Last Updated - 07 April 2002

Main Index

Articles Index

Contents

- **Part 1**
  - Introduction
  - Preamble
- **Part 2**
  - Speaker Leads
- **Part 3**
  - Interconnects
- **Part 4**
  - Power Leads
- **Part 5**
  - Audiophile Capacitors
  - High Current Amplifiers
  - Monoblock Amplifiers
  - Power Supplies
  - "Special" Designs
  - Opamps
  - Valves (Tubes)
  - Speakers
- **Part 6**

- Conclusion
- Further Reading
- References

- **External Sites**
  - The ABX Data Page

Speaker Leads

Speaker leads have been discussed extensively in my article on impedance, but I shall repeat some of this here for the sake of completeness. For the full text, see Impedance [2] and Amp Sound, an article discussing the influences that affect the sound of amplifiers.

This was pointed out to me by a reader, and was originally published in the New York Times (on-line edition) a while ago ...

http://www.nytimes.com/library/tech/99/12/circuits/articles/23down.html

Copyright 1999 The New York Times Company

December 23, 1999

A Spat Among Audiophiles Over High-End Speaker Wire

By ROY FURCHGOTT

In the last year, Lewis Lipnick has tested high-end audio cables from 28 manufacturers. As a professional musician with the National Symphony Orchestra and as an audio consultant, he counts on his exacting ear to tell him if changing cables affects the accuracy of the sound from his $25,000 Krell amplifiers.

His personal choice is a pair of speaker wires that cost $13,000. "Anyone would have to have cloth ears not to tell the difference between cables," he said.

"In my professional opinion that's baloney," said Alan P. Kefauver, a classically trained musician and director of the Recording Arts and Sciences program at the Peabody Institute of Johns Hopkins University. "Has the wire been cryogenically frozen? Is it flat or round? It makes no difference, unless it makes you feel better." His choice for speaker wire? Good-quality 16-gauge zip wire.

The disagreement would be unnotable except for one thing: experts are in agreement that most cables that claim to improve the sound of audio equipment don't. Even cables costing thousands of dollars per foot are often little more than sonic snake oil, experts say. Consumers trying to purchase audio cables often find themselves buying high-end replacements because the only cables in the store are expensive ones.

A purchaser of an entry-level $550 stereo system might be sent home with $55 worth of the least expensive middle-quality audio cables. While experts agree that most cables make exaggerated and unfounded claims about improving sound, they cannot agree on which cables actually do improve sound and which do not.

The scientific record is unclear. So far no research paper contending to prove or disprove the value of fancy wires has been accepted by the leading industry publication, The Journal of the Audio Engineering Society, said Patricia M. MacDonald, its executive editor. She said there were dozens of reasons a research paper might not meet her journal's standards.

"I don't think anyone should infer anything from it," she said.

The manufacturers and sellers of audio goods like to stay above the fray. Cables are a highly lucrative item that may account for a modest percentage of sales but a greater percentage of profit. Even audio manufacturers not directly involved in the cable business like to steer clear of the debate.

[Related Articles: Do It Yourself: A Little Soldering Goes a Long Way; (December 23, 1999)]

Polk Audio, a well respected manufacturer of loudspeakers in Baltimore, no longer makes cables but declined an invitation to set up a listening test in its laboratories. One reason it gave was that the test could affect relationships with audio stores. "We would be hearing from every retailer in the country," said Paul Dicomo, communications director for Polk Audio. Kerry Moyer, staff director for the Consumer Electronics Association, which represents manufacturers, said accessories were usually the highest markup items, wires included. Sales of high-margin accessories have become critical in the current market, where prices of components like receivers, amplifiers and DVD players, have had profit margins squeezed by competition.

"It becomes a question of where are we going to make a little money?" he said. Mr. Moyer, whose $3,000 sound system uses about $300 worth of cables, said the technological superiority of a cable is not the issue -- it is the perceived value to the hobbyist.

"If someone feels good about buying it, whether it works or it doesn't, it makes them feel good," he said. "I don't think we should question."

John Dunlavy, who manufactures audiophile loudspeakers and wire to go with it, does think questioning is valid. A musician and engineer, Mr. Dunlavy said as an academic exercise he used principles of physics relating to transmission line and network theory to produce a high-end cable. "People ask if they will hear a difference, and I tell them no," he said.

Mr. Dunlavy has often gathered audio critics in his Colorado Springs lab for a demonstration.

"What we do is kind of dirty and stinky," he said. "We say we are starting with a 12 AWG zip cord, and we position a technician behind each speaker to change the cables out." The technicians hold up fancy-looking cables before they disappear behind the speakers. The critics debate the sound characteristics of each wire. "They describe huge changes and they say, 'Oh my God, John, tell me you can hear that difference,'" Mr. Dunlavy said. The trick is the technicians never actually change the cables, he said, adding, "It's the placebo effect."

This leads to disagreements based on competing science. Bruce Brisson, who owns Music Interface Technology, an ultrahigh-end wire manufacturer in Rockland, Calif., also wants to see cable charlatans revealed and may use his extensive laboratory to do it.

"I am getting ready to expose this in the year 2000," he said. "People are paying a lot of money and getting nothing for it." But he disagrees with Mr. Dunlavy on the effectiveness of wires, saying that the theory Mr. Dunlavy uses to design his cables is not the right theory and that is why listeners cannot hear a difference. **

Some scientists say it would be difficult to prove one way or another. Changing cables leaves a time lapse that makes comparison difficult. Putting several stereos side by side with the different wires would mean that the speakers would be different distances from the ear, which could have an effect. And while a switch could be made that would send a signal through each of several cables to a speaker from a single sound system, cable makers say the switch itself might spoil the advantages of their wires.

Part of the difficulty is that there are still unexplained acoustic phenomena. William Morris Hartmann, a professor of physics at Michigan State University in East Lansing, works on psycho-acoustic projects, which investigate the way sound is perceived, rather than the way it is produced. There are examples, he said, of sounds that measure beyond the range of human hearing, and yet some people seem to perceive them. That means the market is left open to wild claims and psuedoscience. "It's annoying, but it's hard to disprove," Professor Hartmann said.

Perhaps the closest thing to middle ground is the position taken by Russ Hamm, an electrical engineer whose New York company G Prime Ltd. installs digital processing

equipment for studios. Mr. Hamm said that indeed, wires do make a perceivable difference, but very little, and then only to professionals, like the engineers at BMG Music. He
lent them new high-grade cables for use on roughly $250,000 of equipment. On his system, Mr. Hamm uses a specialty cable manufactured in Vienna that costs $2 a foot.

"We are talking subtle differences, but that is what the high end is all about," he said.

It is a subtlety he describes as a 2 percent difference on a high-end system. "If you had a fine Bordeaux wine, how much does it matter if it's in a nice wineglass or a Riedel crystal glass?"

His advice to audiophiles: "I would say that you want to put the first $10,000 into your equipment."

Copyright 1999 The New York Times Company

The above is reproduced verbatim, and I hope that this information is helpful to your understanding of the topics to follow. Remember that the purpose of this article is *not* to try to sell you something, but to inform and rationalise the many myths that abound regarding the audibility (or otherwise) of different cables.

** It is worth noting that the year 2000 has come and gone, and to my knowledge, neither Bruce Brisson nor anyone else has produced a scientifically sound explanation for the alledged superiority of any one cable over another. There is simply little or no validation available for the vast majority of the outlandish claims made.

We hear so much about damping factor, the effect of speaker leads (and how much better this lead sounds compared to an "ordinary" lead), and how amplifiers should have output impedances of micro-Ohms to prevent "flabby" bass and so on. But what does it all really mean?

Before an informed judgement can be made, we need to look at some of the real factors involved. There are a multitude of impedances involved in a typical amplifier to loudspeaker connection, most of them having a vastly more profound effect than the impedance of the speaker lead alone.

For example, my own (tri-amped) hi-fi uses separate amps for the bass and mid with a

designed output impedance of about 2 Ohms. This provides a useful extension of the bottom end (I'm using sealed enclosures), without excessive peaking at resonance. Much the same effect is found with most valve amplifiers, which typically have an output impedance of 1 to 6 Ohms.

Ignoring the losses in the speaker lead (which are usually very small), the impedance of the cable is very low compared to that of loudspeaker crossover networks and the like. While there is no denying that some speaker leads do sound different, the important thing is "different" rather than "better"

A double-blind test carried out by an Australian electronics magazine many years ago found that most listeners thought that the really thin figure-8 type speaker cable had better bass than all the more expensive ones. Treble response was generally thought better using a heavy duty 3-core mains cable. No-one thought that any of the high priced cables sounded better than anything else.

Other workers in the field, such as Douglas Self [1], have determined much the same, so even in the light of some convincing evidence to the contrary, we have reviewers still extolling the virtues of cables costing more than a decent set of loudspeakers.

Generally, resistance and inductance in the speaker lead can (and does) cause minor variations in level, especially with difficult loads. These deviations are likely to be less than 0.1dB for reasonable cable constructions, with inductance less than 4uH. The resistance of a typical cable (perhaps 0.1 Ohm) causes response variations across the band, following the loudspeaker impedance curve, but these are usually even less at around 0.05 dB. Neither variation is audible.

You will even find references in some cases to the cable's characteristic impedance - a value that is only useful if cables are used for radio frequencies, or are many kilometres in length. These are uncommon in audio listening rooms in my experience. The characteristic impedance of a cable has no effect whatsoever on signal frequencies that are low compared to cable length. At the worst (using coaxial cable) a signal travels at 0.8 of the speed of light (3x10^8 m/s).

Assume that for an adequate safety margin we want to be able to pass up to 100kHz through the speaker cable. The wavelength at this frequency is 3000 metres, or in coax, 3750 metres. A typical listening room may require up to 10 metres of cable, so at the very worst case, the cable is 1/300 to 1/375 times the wavelength of the signal. The effect is utterly insignificant in all respects. The signal will be delayed by an amount that is less than that experienced if the listener were to move his/her head by 1mm towards

or away from the loudspeaker. This is of course a common occurrence, and often by several millimetres, even while asleep.

Difficult Loads

While it is true that reasonable quality twin cables (figure eight or zip cord) are adequate for nominal 8 ohm loads over short distances, there are a number of popular loudspeakers that are anything but nominal at high frequencies.

Two that a reader advised me about are the AR11 and the Quad ESL (old model). Both of these drop below 2 ohms in the treble frequencies. The AR bottoming out at 5kHz and the Quad at 18Khz (although anything from 15kHz to 18kHz is common). The dips are fairly sharp and so the load impedance is highly capacitive on the way down and inductive on the way up. The frequencies are high enough to not worry good amplifiers but what about the response at these dip frequencies?

Twin wire cables all have significant inductance which increases in proportion to length. With 10 amp rated twin flex over only 5 metres the response was down by 2.5 dB into one Quad ESL at 18 Khz, and 3.5 dB into the other speaker which had 8 metres. This was audible and unacceptable.

The only way to reduce cable linear inductance is to make the two wires talk to each other. Running in close parallel is a start, tight twisting is better but only by using multiple wires for each and interweaving can you really get the inductance down. Several cable makers have done this and sell them as low impedance cables, which is exactly what they are. There are several different cables that use this method, and twin coaxial cable is also used to achieve a similar result.

One construction uses two groups of 72 strands of enamelled wire plaited around a solid plastic core. Using these cables with difficult loads, the droop at either 5 or 18 kHz disappeared and the sound was distinctly better. There would be virtually no other way to solve the problem short of mono amplifiers sited next to each loudspeaker.

One (potentially major) drawback ocurs if you own certain amplifiers that are unstable with capacitive loads. Typical multiple twisted pair cable has about 9nF per metre of capacitance with little resistance or inductance, which causes many amplifiers to go into parasitic oscillation. The fix is simple, wind twelve turns of wire around a pen and put it in series with the beginning of the cable. This tiny coil has far less inductance than even one metre of twin flex.

This description of the possible issues with speaker cables is the first I have seen that

makes some sense from a technical perspective. There is sufficient evidence from my own measurements and those of many writers that there are indeed some detectable (and measurable) differences. With this in mind, and wanting to provide all the information I can, I have included this information - and this is the one area where properly sized and well made cables really does make a difference. If you own speakers that present a highly capacitive load, or have deep "notches" in the impedance curve, I would take this information seriously.

Summary

Essentially, the main offenders in speaker leads are resistance and inductance. Of these, inductance is the hardest to minimise, and although usually small, it may still cause problems with some loads (see update, below). Many construction methods have been used, from multiple CAT-5 data cables, with the wires interconnected (usually all the coloured leads are deemed the +ve conductor, and all the white wires - the "mates" - are used as the negative). Because of the tight twist, the inductance is minimised, but at the expense of capacitance. In some cases, the capacitance may be high enough to cause instability in the amplifier, which not only does awful things to the sound, but can damage the amp.

Another popular method of minimising inductance is to use a pair of coaxial leads (e. g.75 Ohm TV/video coax or similar). The inner conductor of one and the outer conductor of the other are joined to make the +ve lead, and vice-versa for the negative. A good quality coax has a relatively low capacitance, and by interconnecting in this way, inductance is also reduced by a very worthwhile margin.

It is widely held that with difficult loudspeaker loads - as presented by many modern speaker systems with complex crossover networks - that reducing inductance can be very beneficial. This is especially true where the crossover causes significant drops in impedance at some frequencies. This also places unusually high demands on the amplifier - one of the reasons that some amplifiers just don't "cut it" with some speakers.

These problems can be reduced or even eliminated entirely by biamping or triamping [3], allowing the use of good quality but not extravagant speaker leads.

Resistance, which is easy to eliminate, reduces the damping factor and wastes power. With even reasonably robust leads, this should not be an issue

Bottom Line on Speaker Leads

Use quality cable, but extravagance will buy no more genuine performance. You will be able to obtain far greater benefits by biamping the system [3] than spending the same

amount on esoteric (read "expensive") speaker leads.

Be willing to experiment, using 3-core mains cable (not the types described above, either), and paralleling two of the conductors for the speaker negative connection (or the positive - the speaker will not care either way). Save yourself a fortune, so you can buy more music instead.

I have seen several references on the web regarding the use of Cat-5 network cable and specially wired coaxial cable for speakers. The idea with network cable is to parallel the wires (these cables are usually 4-pair), and it is claimed that the sonic performance is excellent. I haven't tried it, but Cat-5 is relatively inexpensive, and might work quite well. Try it if you want to. Wiring coaxial cables for speaker use is also not too hard, and it is claimed that this can beat most of the really expensive cables.

Before one even considers the alleged benefits of one cable over another, here is something to think about ...

> "What does "veiled" mean (in reference to high frequency reproduction), and how is it determined that the veiling effect is caused by anything specific, as opposed to everything in general? This includes state of mind (i.e. good day, kids acting up, wife annoyed about something), health (cold or flu, hay fever), position of listening chair (was it moved to vacuum the floor?), etc."

And, no, these are not trivial questions. They are every bit as important as anything else, and all the more so if we have only a subjective interpretation of the sound, without measured results that show the effect. Have a look at the article "Amplifier Sound" for more info.

<PREVIOUS NEXT>

Main Index

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999/2000/2001/2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Last Revision: 07 Apr 02 - changed layout, added additional comments

Elliott Sound Products — Cables, Interconnects & Other Stuff - Part 3

Rod Elliott - Copyright (c) 1999/ 2000/ 2001/ 2002

Page Last Updated - 07 April 2002

Main Index

Articles Index

Contents

- **Part 1**
  - Introduction
  - Preamble
- **Part 2**
  - Speaker Leads
- **Part 3**
  - Interconnects
- **Part 4**
  - Power Leads
- **Part 5**
  - Audiophile Capacitors
  - High Current Amplifiers
  - Monoblock Amplifiers
  - Power Supplies
  - "Special" Designs
  - Opamps
  - Valves (Tubes)
  - Speakers
- **Part 6**

- Conclusion
- Further Reading
- References

- **External Sites**
  - The ABX Data Page

Interconnects

All well designed interconnects will sound the same. This is a contentious claim, but is regrettably true - regrettable for those who have paid vast sums of money for theirs, at least. I will now explain this claim more fully.

The range (and the associated claims) of interconnects is enormous. We have cables available that are directional - the signal passes with less intrusion, impedance or modification in one direction versus the other. I find this curious, since an audio signal is AC, which means that electrons simply rush back and forth in sympathy with the applied signal. A directional device is a semiconductor, and will act as a rectifier, so if these claims are even a tiny bit correct, I certainly don't want any of them between my preamp and amp, because I don't want my audio rectified by a directional cable.

Oxygen free copper (or OFC) supposedly means that there is no oxygen and therefore no copper oxide (which is a rectifier) in the cable, forming a myriad of micro-diodes that affect sound quality. The use of OFC cable is therefore supposed to improve the sound.

Try as I might (and many others before me), I have never been able to measure any distortion in any wire or cable. Even a length of solder (an alloy of tin and lead) introduces no distortion, despite the resin flux in the centre (and I do realise that this has nothing to do with anything - I just thought I'd include it :-). How about fencing wire - no, no distortion there either. The concept of degradation caused by micro-diodes in metallic contacts has been bandied about for years, without a shred of evidence to support the claim that it is audible.

At most, a signal lead will have to carry a peak current of perhaps 200uA with a voltage of maybe 2V or so. With any lead, this current, combined with the lead's resistance, will never allow enough signal difference between conductors to allow the copper oxide rectifiers (assuming they exist at all) to conduct, so rectification cannot (and does not) happen.

What about frequency response? I have equipment that happily goes to several MHz, and at low power, no appreciable attenuation can be measured. Again, characteristic impedance has rated a mention, and just as with speaker cables it is utterly unimportant at audio frequencies. Preamps normally have a very low (typically about 100 Ohms) output impedance, and power amps will normally have an input impedance of 10k Ohms or more. Any cable is therefore mismatched, since it is not sensible (nor is it desirable) to match the impedance of the preamp, cable and power amp for audio frequencies.

Note: There is one application for interconnects where the sound *can* change radically. This is when connecting between a turntable and associated phono cartridge and your preamp. Use of the lowest possible capacitance you can find is very important, because the inductance of the cartridge coupled with the capacitance of the cable can cause a resonant circuit within the audio band.

Should you end up with just the right (or wrong) capacitance, you may find that an otherwise respected cartridge sounds dreadful, with grossly accentuated high frequency performance. The only way to minimise this is to ensure that the interconnects have very low capacitance, and they **must** be shielded to prevent hum and noise from being picked up.

At radio frequencies, Litz wire is often used to eliminate the skin effect. This occurs because of the tendency for RF to try to escape from the wire, so it concentrates on the outside (or skin) of the wire. The effect actually occurs as soon as the frequency is above DC, but becomes noticeable only at higher frequencies. Litz wire will not affect your hi-fi, unless you can hear signals above 100kHz or so (assuming of course that you can find music with harmonics that go that high, and a recording medium that will deliver them to you). Even then, the difference will be minimal.

In areas where there is significant electromagnetic pollution (interference), the use of esoteric cables may have an effect, since they will (if carefully designed) provide excellent shielding at very high radio frequencies. This does not affect the audio per se, but prevents unwanted signals from getting into the inputs or outputs of amps and preamps.

Cable capacitance can have a dramatic effect on sound quality, and more so if you have long interconnects. Generally speaking, most preamps will have no problem with small amounts of capacitance (less than 1nF is desirable and achievable). With high output impedance equipment (such as valve preamps), cable capacitance becomes more of an issue.

For example, 1nF of cable capacitance with a preamp with an output impedance of 1k

will be -3dB at 160kHz, which should be acceptable to most. Should the preamp have an output impedance of 10k, the -3dB frequency is now only 16kHz - this is unacceptable.

I tested a couple of cable samples, and (normalised to a 1 metre length) this is what I found

	Single Core	**Twin - One Lead**	**Twin- Both Leads**	**Twin - Between Leads**
Capacitance	77pF	191pF	377pF	92pF
Inductance	0.7uH	1.2uH	0.6uH	NT
Resistance	0.12 Ohm	0.38 Ohm	0.25 Ohm	NT

NT - Not Tested

These cables are representative of medium quality general purpose shielded (co-axial) cables, of the type that you might use for making interconnects. The resistance and inductance may be considered negligible at audio frequencies, leaving capacitance as the dominant influence. The single core cable is obviously better in this respect, with only 77pF per metre. Even with a 10k output impedance, this will be 3dB down at 207kHz for a 1 metre length.

Even the highest inductance I measured (1.2uH) will introduce an additional 0.75 Ohm impedance at 100kHz - this may be completely ignored, as it is insignificant.

The only other thing that is important is that the cables are properly terminated so they don't become noisy, and that the shield is of good quality and provides complete protection from external interfering signals. Terminations will normally be either soldered or crimped, and either is fine as long as it is well made. For the constructor, soldering is usually better, since proper crimping tools are expensive.

The use of silver wire is a complete waste, since the only benefit of silver is its lower resistance. Since this will make a few micro-ohms difference for a typical 1m length, the difference in signal amplitude is immeasurably small with typical pre and power amp impedances. On the down side, silver tarnishes easily (especially in areas where there is hydrogen sulphide pollution in the atmosphere), and this can become an insulator if thick enough. I have heard of some audiophiles who don't like the sound of silver wire,

and others who claim that solid conductors sound better than stranded. Make of this what you will :-D

The use of gold plated connectors is common, and provides one significant benefit - gold does not tarnish readily, and the connections are less likely to become noisy. Gold is also a better conductor that the nickel plating normally used on "standard" interconnects. The difference is negligible in sonic terms.

There is no reason at all to pay exorbitant amounts of hard earned cash for the "Audiophile" interconnects. These manufacturers are ripping people off, making outlandish claims as to how much better these cables will make your system sound - rubbish! Buy some good quality audio coaxial cable and connectors from your local electronics parts retailer, and make your own interconnects. Not only will you save a bundle, but they can be made to the exact length you want.

Using the cheap shielded figure-8 cable (which generally has terrible shields) is not recommended, because crosstalk is noticeably increased, especially at high frequencies. That notwithstanding, for a signal from an FM tuner even these cheapies will be fine (provided they manage to stay together - most of them fall to bits when used more than a few times), since the crosstalk in the tuner is already worse than the cable. With typical preamp and tuner combinations, you might get some interference using these cheap and nasty interconnects, but the frequency response exceeds anything that we can hear, and distortion is not measurable.

Digital / Optical Interconnects

Recently I have seen adverts and reviews on fibre optic digital interconnects. Some are supposedly far superior to others, despite the fact that 1s and 0s (light present, light not present) are all that is passed. IMHO, it would take truly monumental incompetence to design any digital interconnect that was incapable of passing a digital signal without corruption. Since fibre optics (non-audiophile grade) are used to carry phone calls and data all 'round the world, with very low error rates and over huge distances, it is ludicrous to assume that any commercial digital interconnect will make any difference over a distance of a metre or so.

Bear in mind that the receiver reconstitutes the signal wave shape, it is usually buffered, and will use some form of error correction as well. As for claims that the difference is audible ....

Summary

Aside from interference pickup, capacitance and crosstalk are the only real potential

problem with interconnects. Capacitance can be minimised by selection of the cable. In some cases, even though the impedance of the preamp may be low enough, use of a highly capacitive cable may cause RF instability in the output stages - this will definitely ruin the sound.

Crosstalk is all but eliminated by the use of good quality shielding, which will generally also reduce interference. Keeping lead lengths to the minimum needed will also help reduce any possible negative influences.

Bottom Line on Interconnects

Use home made ones, or buy cables that are well made and reasonably priced. The expensive ones that will "make your system sound better" won't - you are just making some idiot richer, and yourself poorer.

I know that this is heresy to some, but I really don't care. This is factual, and I can prove my claims, while the makers of these fancy cables can't.

I have seen home-made cables, braided from multiple strands of wire-wrap wire. The shielding on some of these can be mediocre (at best), so experiment, but don't expect miracles.

< PREVIOUS NEXT >

Main Index

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999/2000/2001/2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Last Revision: 07 Apr - changed layout, added additional comments

Elliott Sound Products — Cables, Interconnects & Other Stuff - Part 4

Rod Elliott - Copyright (c) 1999/ 2000/ 2001/ 2002

Page Last Updated - 07 April 2002

Main Index

Articles Index

Contents

- Part 1
  - Introduction
  - Preamble
- Part 2
  - Speaker Leads
- Part 3
  - Interconnects
- Part 4
  - Power Leads
- Part 5
  - Audiophile Capacitors
  - High Current Amplifiers
  - Monoblock Amplifiers
  - Power Supplies
  - "Special" Designs
  - Opamps
  - Valves (Tubes)

- Speakers
- Part 6
  - Conclusion
  - Further Reading
  - References
- External Sites
  - The ABX Home Page

Power Leads

The power lead (cord or cable if you prefer) is how mains power at 220V, 240V, 110V at 50 or 60Hz gets to your system. The specifics of the voltage and frequency are determined by where you live, and the available household mains provided by your electricity company.

There are mains cables (power cords) available that defy belief. Would you spend US $3000 for a 2 metre mains lead? You can buy a very nice amplifier indeed for this sort of money, but they are there, and someone must be buying the stupid things.

What possible effect 2 metres of flexible cable can have to counteract the kilometres of power company's wiring is a simple question to answer. None. Or, to more precise, none whatsoever. I am not referring to cables with inbuilt filters or other esoterica here, just perfectly ordinary mains leads.

I have measured the distortion on the mains at my workshop test bench. Last time I did this it was 5.6%, and there is absolutely nothing that a cable can do to change this, regardless of cost. The distortion is caused by a multitude of things completely outside our control, with power supplies for computers and other equipment being only one group of offenders.

These draw power at the peak of the AC waveform, causing it to become flattened (similar to clipping in a power amplifier). The various power company transformers along the way will also introduce some degree of distortion, and there are inductive and capacitive losses within the distribution system. As well, there are large motors being controlled by various speed controllers (most use large solid-state switches), used in industry and commercial centres. Lifts, air compressors, machinery, the list is endless.

Because of the resistance of the supply authority's cabling and transformers (there are

some massive cost considerations they must address), when a high power appliance is turned on, the mains voltage falls. This resistance (actually it is impedance) will cause the voltage to vary from one second to the next, with significant drops at the times when meals are being prepared (electric stoves switched on all over the place), and at other periods. I have measured the impedance at my house at 0.8 Ohms (we use 240V in Australia), so an appliance that draws 10 Amps (such as a heater) will cause the voltage to fall by 8 Volts. This could be reduced by increasing the size of my internal wiring, but the gains would be few and the cost high. In 110V countries such as the US, the wiring impedance must be made lower, since all currents are higher for the same power. It is likely that this causes even greater compromise due to the larger wire sizes that must be used (larger wire means greater cost).

So, given that the mains is distorted, and varies in amplitude from minute to minute throughout the day, and has significant impedance, what can be done to fix this? One method would be to use an Uninterruptable Power Supply (UPS), which (if you get the right type) uses the incoming mains to charge batteries, then uses an inverter to supply power to your equipment. You can buy one of these for $3000, and the emerging mains supply will be as clean as the UPS can make it. No cable can do this, regardless of price.

Using a UPS will ensure that your 100W amplifier can provide 100W, despite the variations in the supply voltage. Whether you can hear any difference is doubtful, because even given that the mains can (and does) vary by up to 10%, your equipment should have a reserve power rating that can accommodate such variance.

I have been advised by a reader (in the US) who reminded me that unless the UPS is specifically designed for sinewave output (and the distortion is quoted), it will probably have greater distortion than the mains anyway. The voltage will be stable, but the switching noise of the UPS may actually make matters worse.

A 100W amp (at nominal supply voltage) should give at least 100W. At 10% low supply, this drops to about 80W, a difference of less than 2dB. If you are operating power amps at close to clipping all the time, you will hear the difference, but this is not the way Hi-Fi gear is meant to be used.

> **Q:**Having a nice clean sinewave from the UPS or power conditioner should make a difference though, shouldn't it?
> **Q:**Why would it not make any difference to the sound from the amp?

The answers are simple and complex, but the result is the same. At the very best,

transformer dissipation might be slightly lower, but the AC from the secondary of the transformer is rectified and filtered, making it into DC, since the amp cannot operate from AC power supplies. The amount of DC "ripple", or superimposed AC signal is determined by the design of the amplifier's power supply, and is completely independent of outside influences other than the mains supply impedance. Even this has a very minor effect in the greater scheme of things.

If I really wanted to be able to supply 100 Amps to my speakers for brief moments, my power supply can already do this. If I wanted to be able to do it all the time for extended periods, my speakers would catch on fire. No mains lead will give my power supply this ability, nor take it away. The limitations are in the supply itself, and include the transformer, rectifier and filter capacitors.

One useful observation is that the mains in the US seems to be basically pretty nasty, and not at all what we are used to in Australia. Interference seems to be a major problem, and if this is the case, it will find its way through the power supply and into the amplifier (or other equipment) if the power supply is not well designed.

Also, because of the lower mains voltage in the US (nominally 120V), the current drawn by power amplifiers in particular can cause real problems with cheap light duty cables. I have already made this point, but it is worth making again. Use of a heavy duty lead (possibly shielded if interference is a problem) will make a difference. Whether the power difference is audible or not is debatable, but elimination (or even reduction) of mains borne interference is likely to result in a worthwhile improvement in sound quality.

Normally I would expect that any interference would be audible all the time, and it seems very unlikely that it only manifests itself when there is music playing. Electricity is not that cunning, and is not by nature vindictive. Having said that, it must be noted that when an amplifier is producing a lot of power, the current spikes on the mains will be much greater, and may have an influence that would normally be un-noticed. The easy way to determine this is to move the leads about to find out if this changes the background noise level. If it does, then re-locating the leads will provide far greater benefit than spending a king's ransom on power leads.

Use of a proper power conditioner (or a "pure" sinewave UPS) will completely eliminate mains interference, and this will naturally be beneficial. This is not something I have encountered, but if there is a problem, then this is probably the best way to fix it.

In some cases, all that may be needed is interference suppressors ("spike arrestors"), to get rid of clicks and pops that get into the system via the mains. These are readily

available and fairly cheap, and might be a good place to start.

I have examined several mains leads I have, and upon inspection I saw that the pins of the US plugs are of thin sheet metal (brass). This is folded over for the flat pins (active and neutral) and rolled into a tube for the earth pin. In contrast, our (Australian) plugs have solid brass pins, and are altogether much more substantial than the US ones (the US lead I have is rated at 13A, and is a very solid cable - but the pins are a weakness IMHO). The standard 13A fused UK plug is even more solid than ours - the pins are probably capable of at least 50A based on their size. One style of European style plug I have is also nice and solid. Elsewhere, I do not have samples, and can't comment.

To give you an idea. Listed below are the pin sizes and materials for 4 mains plugs I have

	Australia	10A	**US**	13A	**UK**	13A	**Europe**	10A - 15A
Active	6.4 x 1.5	Brass	6.2 x 1.1	F-Brass	6.3x 4	Brass	4.8 Dia	NP Brass
Neutral	6.4 x 1.5	Brass	6.2 x 1.1	F-Brass	6.3x 4	Brass	4.8 Dia	NP Brass
Earth	6.4 x 1.5	NP Brass	4.7 dia	T-Brass	8x 4	Brass	NA	

- F-Brass - Folded thin sheet brass
- T-Brass - Tubular thin sheet brass
- NP Brass - Nickel Plated Brass
- NA - Not Applicable (This connector uses a recessed socket, and has two plates for the earth contact, and a socket for the earth pin which is in the wall outlet)

As you can see, even though the US lead has a higher current rating and requirement, the pins are smaller and capable of less current. Where the others use solid brass pins, the US lead has thin sheet brass folded over to give a thicker overall pin - which is still thinner than the others. It is also wibbly - you can quite easily bend the flat pins with your fingers - I did ! Bent them right over at the base, *and* bent them in the middle. This is impossible with any of the others. From what I recall of US wall outlets, they are also fairly wimpy affairs, with relatively poor contact surfaces.

Another point mentioned was that by using a high current shielded cable for the mains may prevent the current spikes caused by the amplifier's rectifier from injecting spurious

noise into interconnects and other equipment. There may be some truth in this, and the effects are certainly measurable - this is why professional audio gear uses balanced connections, to eliminate exactly this sort of problem. Use proper "lead dress" - keep all power leads as far as possible from signal interconnects, and if they must cross, then cross them at right angles.

It also seems that the regulations in the US as to what you may (or may not) use as a power cable are somewhat lax (by Australian standards, anyway), allowing sub-standard connectors at both ends of the cable, and with seemingly minimal control over what may legally be sold as a mains lead.

Many (most?) of the high-end cables that I have seen referred to would not be legal in Australia, and in many other countries. No mains lead is allowed to be sold here without electrical authority approval - this is quite expensive to obtain, and involves voltage drop testing (the lead's resistance) and electrical insulation tests, along with various others. Shielded mains lead is uncommon, but I am sure that it would be available.

I have been taken to task seriously by some for not having tested any of the mains cables - well I can't, because I (like everyone else) have no criteria to base tests on. I know from experience that long (or light duty) leads will reduce the power available, but have no way to create interference of the type that could cause severe sonic degradation so I can verify that a cable eliminates it.

As to blanket claims that "the power cord has more influence than anything else in the chain other than room positioning of the speakers" (and yes, someone did make that claim), what can I say?

One of the respondents is apparently a distributor of high end power leads (so I discovered from someone else's posting), and he had no proof to offer, and nor did anyone else, so I am still left with the same conclusion as before (with some modification based on interference problems).

Other (Cheap) Things You Can Do

I had an e-mail from one of my regular readers, who was telling me that his apartment is wired using aluminium cable. This is (apparently) no longer acceptable in the US, but the fact that it was ever allowed at all is quite amazing. He discovered that he was having mains problems, so rather than "invest" in high-end power cables, he simply decided to replace the wall outlets with new ones, and re-terminate the aluminium house wiring. This in itself is not easy, because aluminium forms an oxide (*very* quickly) which is an insulator, and terminations need to be airtight - literally - to stop this from happening.

Aluminium also "flows" under pressure, so to terminate it properly needs a connector that applies constant pressure over a prolonged time - either that, or the terminations need to be tightened every couple of years. This can even happen with copper - many is the time I have found wall outlets where the connector screws were loose enough to allow the cable to move, this was not through negligence but simply the passage of time.

I quote (verbatim) from the e-mail -

> About the power lead, it's a sad world. Actually, my apartment has aluminum wiring. It is deemed fire-hazard these days, but it's an old building and they're not going to re-do the wiring. I had to replace three receptacles because the contact points of the aluminum wires slowly burnt away and left the sockets unusable. Whenever I plugged in a high-power equipment, it'd crackle, lose power and cause even more contact point to burn away.
>
> In that sense, buying a new $3 socket and getting it freshly connected to the mains wire helps HECK of a lot more than buying a $650 mains cable. As an added bonus, I get fresh copper socket holes. I'd think that helps a lot more than replacing a standard cable with a silver super-duper cable.

I couldn't agree more. This is a sensible approach, and does not cost a great deal. In addition, his apartment is (marginally!) less of a fire hazard than before, and the use of an expensive mains lead would not have fixed the underlying problem. Perhaps a few more people could adopt this sensible attitude and actually get some real (as opposed to imaginary or just "cover-up") improvements.

Bottom Line on Power Cables

I am still waiting for a "high-end" power lead manufacturer to supply me with some scientific proof of the advantages of their cable, and how they improve the sound. I have asked, and have not received the information. Nor do I expect to, since they cannot provide any sort of proof because they don't have any.

The last paragraphs of the above tell more of the truth of the matter than any high-end power lead maker ever will. The same (but to a lesser degree outside the US) benefits can be had from anyone who has old wiring and wall outlets regardless of where they live. Even in my own home, I have completely rewired the mains, because the old wiring had perished insulation, and all the sockets were worn out. The difference was not audible, but at least I know that an electrical fault is unlikely.

< PREVIOUS NEXT >

Main Index

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999/2000/2001/2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Last Revision: 07 Apr 02 - changed layout, added additional comments

Elliott Sound Products — Cables, Interconnects & Other Stuff - Part 5

Rod Elliott - Copyright (c) 1999/ 2000/ 2001/ 2002

Page Last Updated - 07 April 2002

Main Index

Articles Index

Contents

- **Part 1**
  - Introduction
  - Preamble
- **Part 2**
  - Speaker Leads
- **Part 3**
  - Interconnects
- **Part 4**
  - Power Leads
- **Part 5**
  - Audiophile Capacitors
  - High Current Amplifiers
  - Monoblock Amplifiers
  - Power Supplies
  - "Special" Designs
  - Opamps
  - Valves (Tubes)
  - Speakers
- **Part 6**

- Conclusion
- Further Reading
- References

- **External Sites**
  - The ABX Data Page

Audiophile Capacitors

Here we have another bunch of lies - or perhaps half truths is a better description. There are differences between capacitors, but they are not (generally) audible - despite the claims. I have seen reference to dielectric losses, the "sound" of polyester is supposedly inferior to that of polystyrene, and on and on. The stupid part is that all these are true - at radio frequencies - at audible frequencies it is very hard or impossible to measure any difference (or hear a difference, using even a simple blind test).

At the frequencies you and I can hear, there is no audible or measurable difference between most capacitors, unless the equipment builder has done something monumentally idiotic, such as reverse bias an electrolytic. This is (fortunately) rare.

There are some capacitors that are inferior in some regards (but superior in others). For example, many ceramic capacitors have a temperature coefficient that causes the capacitance to vary with temperature (usually negative - N750 or N500 capacitors). NPO ceramic caps have a "negative/positive/zero" temperature coefficient - i.e. close to zero). There are many claims that these should not be used in audio, but they are useful in audio and RF designs for decoupling (bypassing). The values are generally too low to be useful in most audio circuits (although ceramics are made in higher values, but not always easy to get), but otherwise they would almost certainly be fine - after all, the dielectric is a ceramic and not plastic, so they have low loss and very low self inductance. Having said that, ceramic caps usually have poor stability (they should never be used in filter circuits other than for RF suppression), and non-linearities are well documented at high AC voltages. I would not use a ceramic cap in the audio path for this reason.

Many other high stability or low loss/high power RF circuits (but not those using inductors - N500 or N750 ceramics counteract the temperature coefficient of the coils) will use silvered mica or the like - this is great at 400MHz, but quite unnecessary at 20kHz. Mind you, they are far and away the best low value caps you can buy, and if you can tolerate the expense, fine. Just don't expect to hear a difference.

Many modern opamps have such a wide bandwidth that ceramic caps (usually in conjunction with electrolytics) should be the only choice for bypassing, despite the negative comments of some audiophiles.

Then there are electrolytic capacitors. Claim upon claim has been made about their distortion and poor frequency response, particularly at high frequencies. I recently saw an article (which I would give reference to if I could remember where I saw it) where a standard electrolytic and an audiophile grade unit were tested in the same circuit. The standard electrolytic was actually better, having a distortion component at mid and high frequencies that was only marginally worse than the "high end" unit, but was much better at low frequencies.

The audibility of an electrolytic cap is (to my mind) still highly contentious. At low frequencies, all electrolytics will start to introduce some distortion. The levels are quite low, but as the capacitor's reactance becomes significant, distortion rises.

The reactance of any capacitor is determined with the formula

$$\mathbf{Xc = 1 / (2 * \pi * f * C)}$$

where **Xc** is capacitance reactance, **f** is frequency and **C** is capacitance (in Farads)

If Xc is maintained at 1/10 (or 0.1) of the supplied load impedance, then this low frequency distortion will not be an issue, but in any case is far lower than that of a loudspeaker.

High frequency performance is affected by the capacitor's internal inductance and dielectric losses, which causes a rise in impedance as frequency increases. It is very common to see electrolytics bypassed with polyester or similar caps, and for RF this is essential. It is also needed when bypassing the power supply rails of an amp, since at the frequencies that amps like to oscillate at (typically above 1MHz), the electrolytic simply has too much impedance. For audio frequencies a bypass is not needed, but will do no harm.

The combined effects of internal resistance and inductance contribute to the electrolytics' equivalent series resistance, or ESR. This can be measured (I have an ESR tester), and is a good indication that a cap is failing. As electrolytics age, their ESR will rise until a point is reached where the component will be unserviceable.

As a test, I checked a few caps in my workshop. I could not measure any distortion created by an electrolytic passing signal current (as opposed to speaker current, which I did not test at this stage). I also checked the frequency response of a couple of electros, and found zero degradation at 100kHz - even a square wave was passed with no visible deterioriation in rise time (which would indicate frequency limitation).

I then tested the ESR and capacitance of a 220uF and 10uF electrolytic, and a 1uF polyester capacitor.

Type	Value	Meas Val.	ESR
Electro	220uF	207uF	0.17
Electro	10uF	10.1 uF	1.2
Polyester	1.0uF	1.02uF	1.5

I thought that this was quite interesting, personally. If we use 10uF electros where we might have otherwise used a 1uF polyester, the ESR is better. Will it make any difference whatsoever to the sound? Of course not. None of these devices introduced any measurable distortion or anything else that I could see. One thing I know for sure is that if I can't see any change on my distortion meter residual, then there is no change. A complex waveform does not affect the validity of this testing, since I can test distortion at any frequency I like, and the appearance of multiple frequencies at once does not affect any passive device.

Considering that I use the averaging facility on my oscilloscope to eliminate noise completely, I can see the most minute change in a signal waveform. If nothing can be seen here, then no-one, regardless of how good they think their ears are, will be able to hear the difference in a properly conducted test.

I was recently taken to task for not mentioning tantalum capacitors. I hate them! They are unreliable, and many tests have shown that their linearity is highly suspect. The only intermittent short circuit I have ever found in a cap was with a tantalum in a power supply circuit. It would fail for long enough to blow the fuse, and then work again. I strongly suggest that you don't use tantalum caps in anything more advanced than a dustbin.

Bottom Line on Capacitors

Various people have advocated passing pulse signals through two different sorts of

capacitor, and subtracting the result, claiming that the non-zero residue proves that capacitors can introduce audible errors. In fact such tests expose only well-known capacitor shortcomings such as dielectric absorption and series resistance, and perhaps the vulnerability of the dielectric film in electrolytics to reverse biasing.

No-one has been able to show how these imperfections could cause capacitor audibility in an amplifier,and my own tests confirm this.

I must confess though, that perhaps we don't know how to perform the "audibility" tests. I do not believe that there is a significant difference, but many do ... who knows?

Non-polarised electrolytics are a different matter, especially when used in crossover networks. These have a tendency to lose capacitance as they age, shifting the crossover frequency with disasterous results (sonically speaking). Because the loss is gradual, you may possibly not even hear it until the tweeter has almost stopped working, as you get used to the sound over a period of time. Unless all bi-polars age at the same rate (unlikely), you *will* start to notice a difference between the two speakers. This is your cue to head off to the electronics shop and buy some replacements (non electrolytic, preferably). There are (supposedly) some major audible differences between bipolar electrolytics and film dielectric (plastic) types. This is your chance to test the theory.

A reader wrote telling me I was wrong about capacitors, and that the differences are audible. The specifics of this audibility were not discussed, and no measurements were offered, except for the following observation:

> There is an audible and measurable difference between different dielectrics. It's less to do with dissipation factor, and more to do with dielectric absorption. There is no black magic about it - its very well documented throughout the entire range of electronics industries. Here's how to convince yourself that this is possibly the most insidious source of distortion in audio. Get a largish value electrolytic reservoir cap and charge it up to (say) 50 volts for a minute or so. Then, with a DVM attached to the terminals, discharge the cap so the voltage reads zero.
>
> After removing the discharge resistor, watch the voltmeter reading climb back up as the cap miraculously charges itself back up from nowhere. In a signal coupling capacitor this would be bizarre enough behaviour to be a worry, but can you imagine what the effect must be if the cap is in a feedback loop? Give it some thought, and think about how much damage would have to be done to a signal to loose the ambience surrounding a quiet instrument buried in a large ensemble. Think of reverb at -60dB, or lower.

> Dielectric absorption is significant, and ceramic caps suffer from it badly, as do electrolytics. Do some measuring, but more importantly, do some listening.

Fair comment, and it deserves an answer. I also know the "stored charge" phenomenon of electrolytics (I actually demonstrated it to my son recently), and although this is absolutely real, it does not reflect the behaviour of the cap in a real world amplifier.

Most capacitors normally don't charge and discharge in this manner, but remain charged to some DC potential at all times. The charge recovery mechanism should never come into play in a properly designed circuit, regardless of programme material. There are exceptions! A capacitor used as the "dominant pole" or Miller capacitance in the Class-A amplifier section of a power amp is charged and discharged fully with each cycle of the input waveform. Ceramic capacitors are commonly used in this role, but I suggest that polystyrene is probably much better. Will you be able to hear the difference? My own experience is that you should not hear the slightest change in the sound, but it is conceivable that with some amplifiers this may in fact be audible in extreme cases.

The dielectric absorbtion process is present to some degree in all capacitors, and although some are definitely worse than others in this respect, I have conducted some tests with my Sound Impairment Monitor (SIM), which has never been able to detect any degradation of the type you might think should happen.

The claim that there will be an effect similar to reverberation at -60dB is complete nonsense. No such effect is measurable or audible. I do think that after all these years, someone would have worked out a way to prove this effect if it existed at all. No such proof has been offered, but I have seen "proof" that a ceramic capacitor (pushed way beyond its voltage rating if I recall correctly) can introduce some measurable distortion. The solution is easy - don't run any capacitor at above its voltage rating, and all will be well.

Feel free to test the theory. Make sure there is no DC in the signal line, and connect a bipolar cap (say 10uF) in series with the interconnect between preamp and power amp. Wire a switch across the cap, and have someone operate the switch while you are listening. You have to be able to hear a difference at least 75% of the time (and accurately identify whether the cap is in or out of circuit). If you can do this, then the probability is that the capacitor is audible (unless you do something nasty with the switch wiring that gives audible clues - this is cheating :-)

High Current Amplifiers

There are some who insist that the instantaneous current output needs to infinite (or at worst, half this value), and that amplifiers with limited current sound terrible. This is another piece of nonsense.

Let's assume that a nominal 8 Ohm loudspeaker load has an impedance minimum of 1 Ohm at some frequency. This is a bad design, but a valid assumption. This means that the amplifier must be able to supply a maximum of 8 times the normal current. A 100W amplifier would then supply a normal peak current of a little over 3.5 Amps. At the frequency where impedance falls to 1 Ohm, this becomes just over 28A.

So let's have a look at the very worst case possible, where the load is fully reactive and returns all supplied energy 180 degrees out of phase (at this point, the load is performing no work, so if a loudspeaker, is making no sound). The amplifier now has to deal with two lots of current - that supplied to the load, and that returned from the load. Even it it were possible, the worst case above would require a current capacity of 56A, however a loudspeaker that presented such a load to any amplifier will not last long in the market, since it will blow up nearly every amplifier that is attached to it.

There is no audible benefit whatsoever in creating an amp that can supply 100 or 200A, since the load will never need this current and is incapable under any circumstance of drawing more than the applied voltage and minimum impedance will allow (allowing for the reactive component of the load).

Bottom Line on High Current Amps

Most quality amps will be able to supply sufficient current to drive the loudspeaker load. Any more capability than this is a waste of money, since it will never be used. To achieve these extravagant currents, the output stage and power supply must be far larger than will ever be needed in real life.

Class-A amplifiers are generally capable of a very modest current, usually barely above that theoretically needed to drive the speaker. I have not heard anyone claim they are rubbish, because of the low current capability.

The one exception is with extreme crossover networks or other speaker configurations that create a difficult impedance load. It will often be found that some amplifiers cannot drive these speakers well, and others have no problem. An amplifier capable of high current may sound better with these loads, but I suggest that the speaker design is flawed if the designer is incapable of creating a crossover that cannot maintain a respectable impedance.

Monoblock Power Amps (And Preamps)

Someone has managed to convince a sizable segment of the audiophile fraternity that to achieve acceptable channel separation, completely separate amplifiers must be used. Considering that it has been shown [5] that 20dB channel separation is quite sufficient for a full stereo image to be appreciated, it is nonsense to claim that infinite separation is needed or desirable.

It is not at all difficult to design an amplifier with better than 50dB separation, even using valves, and any more than this is of no audible benefit. The "cross-modulation" effect that a shared power supply supposedly introduces is drivel. If an amp is so heroically ill-conceived as to suffer from cross-modulation, then simply sticking it into its own case with a separate power supply certainly won't fix it. I might suggest that it is most ideally suited as a boat anchor, since the design is so seriously flawed that it is beyond salvation.

A common power supply is a sensible (and far cheaper) alternative, and will cause no crosstalk in itself. Most amps have a very high ripple rejection, and if they reject ripple, they will also reject any signal frequencies that happen to get onto the supply line.

In fact, the conventional power supply capacitors will filter out all but the lowest frequencies anyway, and since bass is almost invariably recorded onto disc as mono, a minor amount of crosstalk at low frequencies is of no consequence - even if it *were* possible by this means, which it generally is not.

Bottom Line on Monoblocks

Unless you only need a single channel amp (for a subwoofer, for example), they are a waste of money, and serve no useful purpose. You will get a slight improvement in output power, but the real difference will be inaudible in the majority of cases.

Alternatively, they are useful if you want to have the shortest possible speaker leads. The amp can be installed next to the speaker, and a very short lead used to connect the two. Then we create a problem with the low level interconnects, which will be of significant length. There is far more chance of interference and high frequency loss in long interconnects than in speaker cables, so ideally the interconnects should be low impedance balanced circuits. Sadly, most monoblocks do not offer this essential option.

Power Supplies

I have seen reviews, and claims by amplifier designers, that for this amp to sound good, it must have a fast power supply. A power supply does not have speed - high, low or otherwise ... unless it is regulated.

> **Anecdote:** A regulated supply will have a finite ability to maintain its output voltage as the load varies - this is well known in engineering circles, and can cause problems if it is not fast enough.
>
> I have worked on power supplies (many years ago) that were used to power the head positioning amplifiers in the old washing machine sized computer disk drives. One of the tests that we needed to run was a switched load - varying the load from about 0.1 of the rated current to full current repeatedly.
>
> Detectors were used to measure how far the output voltage dropped when the full load was applied, and I designed the circuit to do this. It was fairly fast, and would latch a "Fail" LED if the output fell below a predetermined limit for more than about 1us. BTW, the low voltage limit was set at only about 1V below the rated output voltage, which was 24V if I remember correctly.

Bear in mind that there is no audio signal that can cause an amplifier to go from quiescent current to full output in anything *near* 1us. A 20kHz signal has a period of 50us, and full power at 20kHz will fry your tweeters in seconds.

So, speed is valid for a regulated supply for a critical application, but is completely meaningless for a power amplifier with an unregulated supply - which is 99.9% of them. There is no audio signal that is so fast that it will demand power from perfectly ordinary electrolytic capacitors faster than they can supply it. It is not necessary (other than for radio frequency bypassing) to use polyester caps in parallel with the main electrolytic filter caps, and nor is there any valid reason to specify that 100,000uF (or any other outlandishly expensive number) is needed to power an amp so the supply will be "fast".

A standard electrolytic (say, 10,000uF) will have an equivalent series resistance (ESR) of perhaps 0.01 Ohm, which means that it has an internal time constant of about 100us. More significant is the fact that discharge current is limited by ESR, so if charged to 50V, the maximum current available is 5,000A peak - this is a lot of current! In fact it is so high that it can destroy the cap itself - this is a very good reason not to use a screwdriver to discharge a power supply, well apart from the fact that a decent amount of

capacitance will take the end off a small screwdriver.

If this 50V supply is connected to a 8 Ohm speaker via an amplifier, the maximum current the speaker will draw is 6.25A (although some speakers will demand more at certain frequencies). In reality it will be far less than this most of the time. I can make a power supply "slow", simply by placing some resistance in series - the caps will no longer be able to discharge at their maximum rate. Will this affect an amplifier? Only in that the maximum power will no longer be achieved, but this will also happen if the AC mains supply is 10% low. Does this somehow degrade the sound of an amplifier? I think not.

Bottom Line on Power Supplies

Fast power supplies are a myth, as all power supplies are inherently "fast" in this context. Regulated supplies are generally only used with Class-A amps to reduce ripple. These do not have to be fast, because the current variations are much lower with the Class-A topology. Most will have a fair sized electrolytic at the regulator output anyway, so they are "fast" again.

Massive capacitance and "audiophile" grade caps are not going to improve the sound of the DC from your supply, regardless of cost or claim. The power supply is a passive part of the amplifier, and has little or no influence on the sound, unless grossly and ingeniously poorly designed. I say "ingeniously", because it would take spectacular incompetence to so badly design a power supply that it audibly affected the sound at any signal level below clipping. From some of the posings I have seen on various bulletin boards, such incompetance may well be rife, since just changing a power lead makes them audibly better :-)

"Special" Designs

The editorial page has a pair of prime examples of "Special Designs", including a more detailed examination of the sample below. See I am as Mad as Hell for more info.

I recently saw information on the web about an amp whose claims to fame (infamy, more like it) were along the following lines (this is taken from the site)

- World's smallest number of parts - 9 parts per channel (excluding attenuators)
- World's shortest signal pass length - 32 m/m (including the length of parts)
- World's shortest NFB loop length - 9 m/m (including the length of the resistor)
- World's smallest filter/condenser - 1000µF

- Powerful voltage regulation with high capacity transformer - 170 VA cut-core transformer +- individual coils
- Dual mono construction with each channel in a separate chassis.
- Rigid and compact aluminum chassis construction to release vibrations smoothly.
- Separate 12-position attenuators for each channel.
- Can be up-graded into a pair of complete mono amplifiers by adding another Model X Stupid Name

Oh, wow, and ... I mean ... like ... who gives a toss! This amplifier sold for some astronomical sum, and as near as I could tell from the advertising blurb, seems to use a couple of power opamps as the entire circuit. The power supply was separate.

So the amp has few components and a short signal path. What about the several hundred metres of standard professional class cable and very long signal paths that are common in the mixers that were used in the recording studio? Is this "magical" short signal path going to somehow make that all go away - somehow I doubt it. Since this amp does use a power opamp, the manufacturer obviously does not count the 30 or so transistors inside the device - why not? They are real, whether you acknowledge them or not.

As for the "world's shortest negative feedback path". So what? The claim was made that by doing this, power supply bypass capacitors that by some mystical process ruin the sound were not needed. What rubbish. My 60W power amp has a negative feedback path that is about 50mm long - in other words, typical. Because of its design, it will operate perfectly happily with no power supply bypass capacitors too - the result is greatly reduced power because of the resistance of the power leads. Do you want that? Does this sound like a good idea? No, I didn't think so either.

The one I liked best (or least, depending on how you think I read this nonsense) was the "world's smallest filter capacitors". What possible benefit - other than profit maximisation - does this infer? I honestly have no idea. I could run my amp with 1000uF caps too. Anyone can. The immediate result is a dramatic reduction in power, as ripple voltage is very high at any reasonable power level, and you start to get clipping as the ripple voltage encroaches on the audio signal. Ah, but we also have "powerful voltage regulation" and a high capacity transformer. Big deal. I can run my amp off a 10kVA transformer if I want to, and it won't change the sound one iota. Anyone can make a voltage regulator (assuming that one is actually used, which I doubt), but why? Extra heatsinks, more stuff to fail, and zero sonic benefit.

I won't even bother discussing the "dual mono construction", but I am intrigued by the "rigid and compact aluminium chassis to release vibrations smoothly". Quite apart from

the fact that being rigid and compact in no way ensures that vibrations will be released smoothly or otherwise, I am at a complete loss as to why anyone might think for an instant that this was important. This is an amplifier, not a speaker cabinet. Left to their own devices, amplifiers don't normally vibrate - this is not one of their characteristics. Are we supposed to believe that a power amp is in some way microphonic?

Try this (if you dare). Place your ear as close as possible to the speaker, and have someone drop the power amp a short distance while powered on and connected. What do you think is the chance that you will hear anything from the speaker (other than if the amp destroys itself when it is dropped)? I will tell you, to save the embarassment of having to explain to the service guy what happened to the amp. Nothing, that's what. If these clowns *have* managed to make an amp that is microphonic, then I definitely don't want one.

Bottom Line on Special Designs

Most are rubbish, but genuinely overpriced, while others are just trying to do something different (which they're not) and desperately attempting to convince (confuse?) us that it makes a difference. It doesn't.

I thought about this one for a while, and it finally made it into my "Hall of Infamy" - the editorial. You can read more about it there (see above).

Opamps

Many is the claim that opamps have a distinctive sound, and can readily be heard in audio equipment. Discrete designs supposedly sound superior, regardless of the fact that in many cases they will measure worse than even a cheap opamp.

I have never been able to measure an opamp's distortion, because it is so far below my equipment's limits that it cannot be detected. Devices are available with distortion as low as 0.00008% - this is close as you can get to the ideal "straight wire with gain". The bandwidth of the better devices is so wide that significant gain is available at 100kHz, so phase irregularities and response problems are non-existent in sensible designs.

Considering the fact (and in the vast majority of cases, it is fact) that the final mixed down signal you get from a CD has passed through up to 100 opamps at various stages of production before you even get to listen to it, it is ludicrous to assume that *not* using opamps in the last 1% of the audio chain will have any audible effect.

Bottom Line on Opamps

Opamps are great. They simplify design, have low distortion and excellent power supply rejection, and good ones are very quiet indeed. There are few areas where a discrete design will be better. This naturally assumes the use of good quality units - the venerable uA741 might be OK in a thermal controller, but you don't want them in audio gear (although you might be surprised at some of the opamps that you might find - some are little better than the 741, but they are still used).

Valves (Tubes)

Valve amplifiers are back, with units in all sorts of configurations selling for astounding sums.

The valve sound is one phenomenon that is real. It has been known for a long time that listeners sometimes prefer to have a certain amount of second-harmonic distortion added in, and most valve amplifiers provide just that, due to huge difficulties in providing good linearity with modest feedback factors.

While this may well sound nice, hi-fi is supposedly about accuracy, and if the sound is to be modified in this manner, it should be set from the preamp front panel by a control (Douglas Self suggests a 'niceness' knob).

Valves offer some advantages - their overload characteristics are smoother than solid state designs, so even when clipping the sound is less harsh. While this is most desirable for a guitar amplifier that will be operating into clipping for much of the time, it is unhelpful for hi-fi, where clipping should be avoided altogether.

Valve amps also have much higher output impedance than transistor amps, and this makes some speakers sound better. It also makes other speaker sound worse, so the results are unpredictable.

There are few modern transistor amps that will measure worse than any valve amp, regardless of cost. Indeed, the vast majority are so superior in all respects that it is difficult to justify using valves in anything other than guitar amps, where, despite much advertising hype, no transistor amp has ever been able to sound exactly the same as a valve unit. Close - but not the same.

The rash of single-ended directly heated triode monoblock amplifiers of late is something that astonishes me. These will typically have a distortion of 1 to 3%, are of low power -

typically less than 10W, and have no redeeming features (IMHO).

Such an amplifier generates large amounts of second-harmonic distortion, due to the asymmetry of single-ended operation, and needs a very large output transformer because the primary carries the full DC anode current, and core saturation must be avoided. The inherent distortion of an iron cored inductor or transformer is ever-present, and only global feedback can remove it.

High values of feedback around a transformer are extremely difficult, because the phase irregularities generally cause the amplifier to oscillate. This may have been the state of the art 50 years ago, but there is no sensible reason to go back. Next we will hear someone extolling the virtues of the wax cylinder as having superior sonics to vinyl or CD (needless to say these superior sonics will be "very subtle" and "only audible with the finest (i.e. most expensive) single ended triode monoblock amplifier").

In one review, a single ended triode amplifier yielded 3% THD at 9 Watts, at a cost of $3400 [4]. This is an appalling result for a very expensive single channel amp. The amplifiers in powered computer speakers are better than that!

Despite all of the above, I have no doubt that many of these amps sound delightful. Not exactly my cup of tea, but having used valve amps of many types over the years (including those I designed and built myself), I still like the sound of them. They also don't blow up with difficult loads - they may stress out a little and give less power than normal, but they survive. The majority of valve amps are far less forgiving of open circuits (no speakers connected), and some will fail if pushed hard into an open circuit. The typical failure mode is a high voltage flashover, which either carbonises the valve socket or base (or both), or causes the insulation in the output transformer to fail.

Bottom Line on Valves

This is one area I shall leave open-ended. There are some valve amps that do sound very good indeed, but are generally very expensive. Valves are also fragile, generate copious amounts of heat, and have a limited life. Correct biasing is essential, and few valve amps provide a simple method of doing this.

The trend towards having these hot "bottles" out in full view, and able to be touched (and / or broken) by age challenged persons (the rug-rats) is a definite safety hazard. I would not like anyone's kids to be able to burn and then electrocute themselves in one small mishap.

.... However - I do (or did until recently) use a valve preamp in my own system, and I

have no idea what that says about me. It does sound nice, but I am probably deluding myself in thinking that it is better than my solid-state preamp. That's fine for me, because I designed and built it, so it didn't cost me a king's ransom.

Speakers

There are many very fine loudspeakers available, and interestingly, although these have a far greater effect on the sound that you hear than the amplifier, there is nowhere near the controversy with loudspeakers as seems to be evident with the rest of the audio chain.

Certainly there are proponents of various crossover alignments, the benefits or otherwise of vented boxes versus sealed, but otherwise this seems to be a reasonably sensible (even if intimidatingly expensive) field of endeavour.

Most audiophiles have their favourite speaker system(s), and these will all have some undesirable characteristics, for such is the state of the art. The perfect loudspeaker does not exist, because of the physics of making electro-mechanical objects with finite mass react in a completely predictable manner at all frequencies. This (of course) is something that speakers cannot do.

A flat frequency response is desirable, and rapid decay of internal resonances means that the loudspeaker contributes a minimum of its own sound to that from the source. Good quality drivers and well braced, non resonant cabinets, combined with high quality components in crossover networks and a sensible approach to ensure that phase irregularities at the crossover frequencies do not cause response or impedance peaks and dips are common in most quality systems.

The listening room and the recorded material has a very great influence on the final sound you hear, vastly more than a few interconnects or a mains lead. No-one is going to make the listening room anechoic, and nor would you want to. The positioning of speakers is one thing that can have a profound effect on the sound, but this is so often completely ignored.

One problem is that the optimum placement of speakers for sound quality will often be completely inappropriate to the layout of the room, meaning that a liveable area is no longer available, and causing much friction between the listener and s/he who must be obeyed.

Bottom Line on Speakers

Buy what sounds good, build your own and experiment, whatever. This is too complex an area to try to offer suggestions or advice (although there are many who will do just that, with no knowledge of your listening room, its furnishings or anything else).

Bear in mind that building a speaker system without measurements is futile. The ear (and attached brain) is easily fooled, and has a very short memory for what you hear. Speakers can have huge anomalies in response, and within a few minutes the brain has made the necessary adjustments, and everything will seem to sound as it should.

Try this experiment. If you have a graphic or parametric equaliser, reduce a band somewhere in the midrange area (say, between 500Hz and 1kHz). Listen to the system for about 15 minutes, then restore the missing frequency range. Suddenly, the system will sound as if it has a hugh peak in the midrange, and for a time will sound awful. Within another 15 minutes or so, everything will have settled back to normal.

< PREVIOUS NEXT >

Main Index

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999/2000,2001/2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Last Revision: 07 Apr 02 - changed layout, added additional comments

Elliott Sound Products — Cables, Interconnects & Other Stuff - Part 6

Rod Elliott - Copyright (c) 1999/ 2000/ 2001/ 2002

Page Last Updated - 07 April 2002

Main Index

Articles Index

Contents

- **Part 1**
  - Introduction
  - Preamble
- **Part 2**
  - Speaker Leads
- **Part 3**
  - Interconnects
- **Part 4**
  - Power Leads
- **Part 5**
  - Audiophile Capacitors
  - High Current Amplifiers
  - Monoblock Amplifiers
  - Power Supplies
  - "Special" Designs
  - Opamps
  - Valves (Tubes)
  - Speakers
- **Part 6**

- Conclusion
- Further Reading
- References

- **External Sites**
  The ABX Data Page

Conclusion

On the one hand, we have respected designers who simulate, build, measure and modify until they are satisfied that the performance is as expected. Then, and only then, the amplifier (or whatever it might be) is auditioned in a proper listening test (as opposed to a lab speaker), and perhaps only the designer listens to it in the first instance. If the sound is as expected, then others will usually be invited to listen as well. Comments are made, and if it is felt that they are valid (a sufficient number of listeners made the same remarks, for example), then further modifications will be made, more tests, more listening, until everyone is satisfied that the measured and audible performance is in agreement. The measurements are available on the colour glossies, and are considered a part of the equipment - this is the specification, against which others can be compared.

Compare this to the snake oil vendors. As an example, they buy perfectly ordinary cable from an established manufacturer, clad it in some fancy heatshrink tubing, write their sales pitch, and sell it. They might actually bother to listen to it as well, but there isn't much point, since it is the same wire as used by others anyway. Do you see specifications, measurements, or other factual data? No! What you see on the colour glossies is a sales pitch, aimed directly at your emotional responses. There are no means for direct comparison, not a mention of anything that will help you to make a reasonable and informed decision as to which "thing" is (or might be) better than the other.

Non blind listening tests are flawed - and especially so when conducted by a dealer. Don't expect that the levels will be precisely matched, but *absolutely* expect the salesthing to tell you *what to hear* - not exactly a fair comparison.

When only emotions are allowed to make the decision on technical equipment, we can be fairly certain that we will make the wrong choice, other than by chance. Having spent all that money, no-one, and I mean no-one, will be willing to admit that they were defrauded, robbed or deceived. The survival instinct takes over, and we hear exactly what we expect to - whether it exists ot not.

In the long term, the subjectivist approach will cost you a lot of money, and possibly yield a system that is less hi-fi than something from a department store. A review without technical tests is without substance or meaning, and nearly all descriptions about amplifier sound should be taken with a large dose of salt (possibly epsom).

Claims that power leads and interconnects will magically transform the sound of your system are false and misleading in the extreme. The various system components may be influenced by some combinations, but a well designed system should not care.

The current impasse between the scientific and subjectivist camps is unlikely to be resolved in the near future, because as politics and religion have shown over the centuries, people will believe what they want to, despite any evidence that may be offered to show that they are misguided or just plain wrong.

There is great difficulty defining the quality of an audio experience - you can't draw a picture to show what something sounded like. In addition, our acoustical memory is far more fleeting and more readily fooled than visual memory. It is much easier to visualise what the Sydney Harbour Bridge looks like than to recall all but the basic details of a musical performance.

> From Douglas Self -
>
> It has been universally recognised for many years in experimental psychology, particularly in experiments about perception, that people tend to perceive what they want to perceive. This is often called the 'experimenter expectancy' effect; it is more subtle and insidious than it sounds, and the history of science is littered with the wrecked careers of those who failed to guard against it. Such self-deception has most often occurred in fields like biology, where although the raw data may be numerical, there is no real mathematical theory to check it against.
>
> When the only 'results' are vague subjective impressions, the danger is clearly much greater, no matter how absolute the integrity of the experimenter. Thus in psychological work great care is necessary in the use of impartial observers, double-blind techniques, and rigorous statistical tests for significance. The vast majority of Subjectivist writings wholly ignore these precautions, with predictable results. In a few cases properly controlled listening tests been done, and at the time of writing all have resulted in different amplifiers sounding indistinguishable. I believe the conclusion is inescapable that experimenter expectancy has played a dominant role in the growth of Subjectivism.
>
> It is notable that in Subjectivist audio the 'correct' answer is always the more expensive or inconvenient one. Electronics is rarely as simple as that. A major improvement is

more likely to be linked with a new circuit topology or new type of semiconductor, than with mindlessly specifying more expensive components of the same type; cars do not go faster with platinum pistons.

All the above notwithstanding, most audio designers will still tend to accept (however reluctantly) some of the subjectivist propaganda, if only to be able to extract some of the obviously serious money that would otherwise go elsewhere. There is nothing wrong with this in principle, but where this happens, you will almost invariably get what you pay for, and the equipment's performance will be (hopefully) satisfying to both camps.

Just as likely is that the subjectivists will determine that this same piece of equipment is hopelessly inadequate in all respects, despite the fact that it has zero distortion of any kind, and a frequency response from DC to daylight. (A good quality standard interconnect comes to mind!)

Further Reading

For further reading, have a look at "Amplifier Sound", an article that tries to rationalise some of the misunderstandings and differences of opinion that abound in the audio field.

To help gain an understanding of how we form belief systems, have a read of the article "The Belief Engine". It is a fascinating look into the way our minds work, and helps to explain how we can perceive very obvious differences that don't actually exist.

The articles listed in the References are an additional source for information on these topics.

References

1. Wireless World, July 1988 - D. Self 'Science and Subjectivism in Audio' (See also The Self Site)
2. The Audio Pages, ESP, Impedance
3. The Audio Pages, ESP, Bi-amplification - Not Quite Magic (But Close)
4. Stereophile, Sept 1995 - R. Harley 'Review of Cary CAD-300SEI Single-Ended Triode Amplifier'
5. BBC Engineering Monograph No 52 - 'Stereophony & The effect of crosstalk between left and right channels'

Main Index

Articles Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999/2000/2001/2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Last Revision: 07 Apr 02 - changed layout to separate pages, and added extra comments./ 31 Mar 02 - changed layout, added additional comments./ 24 Aug 2001. Added new intro, and merged some material into the main text./ 22 Dec - Added various comments and clarifications./ 15 Aug - added speaker cable update & bipolar electrolytic info./ 09 Apr-Added a couple of minor points, and one correction./ 04 Mar-corrected errors./ 26 Feb-Corrected section on capacitors, added "Amp Sound" reference./ 28 Jan-corrected some mistakes and typos./ 22 Jan 2000-Added mains plug pin data, and cable capacitance./ 17 Dec-added Power Supply section, and a few small corrections./ Page Created-10 Dec 1999

ABX Double Blind Comparator Data

HOME PAGE | HARDWARE | DATA | LITERATURE

May 7, 1977
The Original SMWTMS ABX Test Results

Power Amplifier Comparison	Result	Correct	p less than	Listeners
10 Watt Tubes vs. Dyna 400	DIFFERENT ≠ DIFFERENT	63 / 80 = 79%	0.0005	10
Paoli 60M vs. Dyna 400	DIFFERENT ≠ DIFFERENT	78 / 126 = 62%	0.005	9
Swartz 40* vs. Dyna 400	SAME	41 / 84 = 49%	-	6

These were the first three audio double blind tests ever done.

* The Swartz 40 is a 20 Watt per channel power amplifier with quasi-complementary ouputs built exactly to the five transistor per channel design in the RCA Transistor Manual. By the espoused standards of subjectivists, nothing should sound worse than this amplifier.

DIFFERENT ≠ DIFFERENT	The different symbol means it is unlikely the percent correct score occurred by chance and thus the null hypothesis was disproven, which substantiates a real difference in sound quality.

The same symbol means that it is likely the correct score occurred by chance, the null hypothesis is not rejected, which means no difference was found. It does remain possible a difference may be substantianted with further testing.

ABX Test Data

Capacitors	CD Players
DA Converters	Digital Delay
Distortion	Filters, First Order
Filters, Fourth Order	Interconnects
Level Difference	Mini-Disc
Polarity	Phono Cartridges
Phono PreAmps	Power Amplifiers
Speakers	Speaker Wires
Tape Decks	Tape Generations
Video Cables	

ABX HOME PAGE ABX HARDWARE ABX DATA ABX LITERATURE

Send questions or comments on the ABX Web Page to David Carlstrom.

David Carlstrom
P O Box 721464
Berkley MI 48072-0464
(248) 544-8453

Web Page Created by David Carlstrom. Last Modified: 7/1/2001

PayPal DONATE
You can help by making a tax-deductible donation!

Skeptical Inquirer

- **Subscribe** to the magazine for science and reason - for the Internet price of $19.95
- Index of Articles
- List of Online Articles

Skeptical Briefs

- **Join CSICOP** as an associate member to receive the newsletter

CSICOP

- About CSICOP
- Events
- Join the Mailing List
- Giftshop

The Belief Engine

Our brains and nervous systems constitute a belief-generating machine, a system that evolved to assure not truth, logic, and reason, but survival. The belief engine has seven major components.

James Alcock

The following beliefs are strongly held by large numbers of people. Each of them has been hotly disputed by others:

- Through hypnosis, one can access past lives.
- Horoscopes provide useful information about the future.
- Spiritual healing sometimes succeeds where conventional medicine fails.
- A widespread, transgenerational Satanic conspiracy is afoot in society.
- Certain gifted people have been able to use their psychic powers to help police solve crimes.
- We can sometimes communicate with others via mental telepathy.
- Some people have been abducted by UFOs and then returned to earth.
- Elvis lives.
- Vitamin C can ward off or cure the common cold.
- Immigrants are stealing our jobs.
- Certain racial groups are intellectually inferior.
- Certain racial groups are athletically superior, at least in some specific sports.
- Crime and violence are linked to the breakdown of the traditional family.
- North Korea's developing nuclear capability poses a threat to world peace.

Despite high confidence on the part of both believers and disbelievers, in most instances, neither side has much -- if any -- objective evidence to back its position. Some of these beliefs, such as telepathy and astrology, stand in

The following selections do not necessarily represent the views of CSICOP:

amazon.com and you're done.

No image available

The Biology Of Belief
Bruce H. Lipton
New $15.75!

Beyond Belief
Elaine Pagels

Practicing Greatness
Reggie McNeal

A Hidden Wholeness
Parker J. Palmer

A Code of Jewish Ethics
Joseph Rabbi Telushkin

Mindset
Carol Dweck

The Anger Trap
Les Carter

(Prices may change)
Privacy Information

Special Features

- Chain Letters Anonymous
- Council for Media Integrity
- HoaxWatch
- Intelligent Design Watch
- Inquiring Minds
- Klass Files
- Pensar Magazine
- SKEPTIC Bibliography
- Skeptic's Toolbox
- Skeptiseum
- Superstition Bash

Web Columns

- Special Articles
- Doubt and About
- Generation sXeptic
- Science and the Media

Center for Inquiry

- Center for Inquiry
- CFI Libraries
- Employment Listings

Resources

- International

contradiction to the current scientific worldview and are therefore considered by many scientists to be "irrational." Others are not at all inconsistent with science, and whether or not they are based in fact, no one would consider them to be irrational.

Nineteenth-century rationalists predicted that superstition and irrationality would be defeated by universal education. However, this has not happened. High literacy rates and universal education have done little to decrease such belief, and poll after poll indicates that a large majority of the public believe in the reality of "occult" or "paranormal" or "supernatural" phenomena. Why should this be so? Why is it that in this highly scientific and technological age superstition and irrationality abound?

It is because our brains and nervous systems constitute a belief-generating machine, an engine that produces beliefs without any particular respect for what is real or true and what is not. This belief engine selects information from the environment, shapes it, combines it with information from memory, and produces beliefs that are generally consistent with beliefs already held. This system is as capable of generating fallacious beliefs as it is of generating beliefs that are in line with truth. These beliefs guide future actions and, whether correct or erroneous, they may prove functional for the individual who holds them. Whether or not there is really a Heaven for worthy souls does nothing to detract from the usefulness of such a belief for people who are searching for meaning in life.

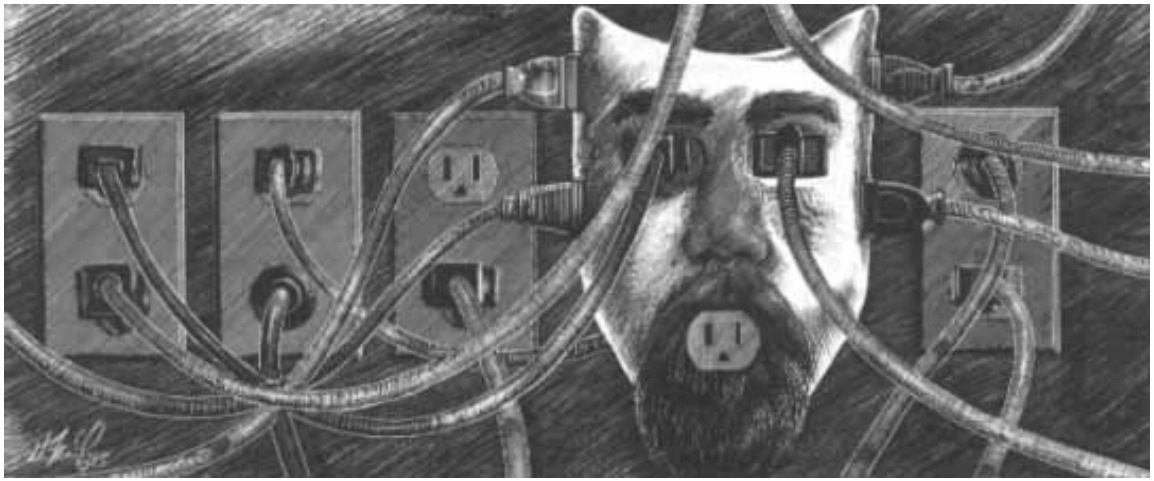
Nothing is fundamentally different about what we might think of as "irrational" beliefs -- they are generated in the same manner as are other beliefs. We may not have an evidential basis for belief in irrational concepts, but neither do we have such a basis for most of our beliefs. For example, you probably believe that brushing your teeth is good for you, but it is unlikely that you have any evidence to back up this belief, unless you are a dentist. You have been taught this, it makes some sense, and you have never been led to question it.

If we were to conceptualize the brain and nervous system as a belief engine, it would need to comprise several components, each reflecting some basic aspect of belief generation. Among the components, the following units figure importantly:

1. The learning unit
2. The critical thinking unit
3. The yearning unit

Network of Skeptical Organizations
- Links

4. The input unit
5. The emotional response unit
6. The memory unit
7. The environmental feedback unit.



The Learning Unit

The learning unit is the key to understanding the belief engine. It is tied to the physical architecture of the brain and nervous system; and by its very nature, we are condemned to a virtually automatic process of magical thinking. "Magical thinking" is the interpreting of two closely occurring events as though one caused the other, without any concern for the causal link. For example, if you believe that crossing your fingers brought you good fortune, you have associated the act of finger-crossing with the subsequent welcome event and imputed a causal link between the two.

Our brain and nervous system have evolved over millions of years. It is important to recognize that natural selection does not select directly on the basis of reason or truth; it selects for reproductive success. Nothing in our cerebral apparatus gives any particular status to truth. Consider a rabbit in the tall grass, and grant for a moment a modicum of conscious and logical intellect to it. It detects a rustling in the tall grass, and having in the past learned that this occasionally signals the presence of a hungry fox, the rabbit wonders if there really is a fox this time or if a gust of wind caused the grass to rustle. It awaits more conclusive evidence. Although motivated by a search for truth, that rabbit does not live long. Compare the late rabbit to the rabbit that responds to the rustle with a strong autonomic nervous-system reaction

and runs away as fast as it can. It is more likely to live and reproduce. So, seeking truth does not always promote survival, and fleeing on the basis of erroneous belief is not always such a bad thing to do. However, while this avoidance strategy may succeed in the forest, it may be quite dangerous to pursue in the nuclear age.

The learning unit is set up in such a way as to learn very quickly from the association of two significant events -- such as touching a hot stove and feeling pain. It is set up so that significant pairings produce a lasting effect, while nonpairings of the same two events are not nearly so influential. If a child were to touch a stove once and be burned, then if the child were to touch it again without being burned, the association between pain and stove would not automatically be unlearned. This basic asymmetry -- pairing of two stimuli has an important effect, while presenting the stimuli unpaired (that is, individually) has a much lesser effect -- is important for survival.

This asymmetry in learning also underlies much of the error that colors our thinking about events that occur together from time to time. Humans are very poor at accurately judging the relationship between events that only sometimes co-occur. For example, if we think of Uncle Harry, and then he telephones us a few minutes later, this might seem to demand some explanation in terms of telepathy or precognition. However, we can only properly evaluate the co-occurrence of these two events if we also consider the number of times that we thought of Harry and he did not call, or we did not think of him but he called anyway. These latter circumstances -- these nonpairings -- have little impact on our learning system. Because we are overly influenced by pairings of significant events, we can come to infer an association, and even a causal one, between two events even if there is none. Thus, dreams may correspond with subsequent events only every so often by chance, and yet this pairing may have a dramatic effect on belief. Or we feel a cold coming on, take vitamin C, and then when the cold does not get to be too bad we infer a causal link. The world around us abounds with coincidental occurrences, some of which are meaningful but the vast majority of which are not. This provides a fertile ground for the growth of fallacious beliefs. We readily learn that associations exist between events, even when they do not. We are often led by co-occurring events to infer that the one that occurred first somehow caused the one that succeeded it.

We are all even more prone to error when rare or emotionally laden events are involved. We are always looking for causal explanations, and we tend to infer causality even when none exists. You might be puzzled or even distressed if you heard a loud noise in your living room but could find no

source for it.

The Critical-Thinking Unit

The critical-thinking unit is the second component of the belief engine, and it is acquired -- acquired through experience and explicit education. Because of the nervous-system architecture that I have described, we are born to magical thinking. The infant who smiles just before a breeze causes a mobile above her head to move will smile again and again, as though the smile had magically caused the desired motion of the mobile. We have to labor to overcome such magical predisposition, and we never do so entirely. It is through experience and direct teaching that we come to understand the limits of our immediate magical intuitive interpretations. We are taught common logic by parents and teachers, and since it often serves us well, we use it where it seems appropriate. Indeed, the cultural parallel of this developmental process is the development of the formal method of logic and scientific inquiry. We come to realize that we cannot trust our automatic inferences about co-occurrence and causality.

We learn to use simple tests of reason to evaluate events around us, but we also learn that certain classes of events are not to be subjected to reason but should be accepted on faith. Every society teaches about transcendental things -- ghosts, gods, bogeymen, and so on; and here we are often explicitly taught to ignore logic and accept such things on faith or on the basis of other people's experiences. By the time we are adults, we can respond to an event in either a logical, critical mode or in an experiential, intuitive mode. The events themselves often determine which way we will respond. If I were to tell you that I went home last night and found a cow in my living room, you would be more likely to laugh than to believe me, even though there is certainly nothing impossible about such an event. If, on the other hand, I were to tell you that I went into my living room and was startled by an eerie glow over my late grandfather's armchair, and that the room went cold, you may be less likely to disbelieve and more likely to perk up your ears and listen to the details, possibly suspending the critical acumen that you would bring to the cow story. Sometimes strong emotion interferes with the application of critical thought. Other times we are cleverly gulled.

Rationality is often at a disadvantage to intuitive thought. The late psychologist Graham Reed spoke of the example of the gambler's fallacy: Suppose you are observing a roulette wheel. It has come up black ten times in a row, and a powerful intuitive feeling is growing in you that it must soon come up red. It cannot keep coming up black forever. Yet your rational mind

tells you that the wheel has no memory, that each outcome is independent of those that preceded. In such a case, the struggle between intuition and rationality is not always won by rationality.

Note that we can switch this critical thinking unit on or off. As I noted earlier, we may switch it off entirely if dealing with religious or other transcendental matters. Sometimes, we deliberately switch it on: "Hold it a minute, let me think this out," we might say to ourselves when someone tries to extract money from us for an apparently worthy cause.

The Yearning Unit

Learning does not occur in a vacuum. We are not passive receivers of information. We actively seek out information to satisfy our many needs. We may yearn to find meaning in life. We may yearn for a sense of identity. We may yearn for recovery from disease. We may yearn to be in touch with deceased loved ones.

In general we yearn to reduce anxiety. Beliefs, be they correct or false, can assuage these yearnings. Often beliefs that might be categorized as irrational by scientists are the most efficient at reducing these yearnings. Rationality and scientific truth have little to offer for most people as remedies for existential anxiety. However, belief in reincarnation, supernatural intervention, and everlasting life can overcome such anxiety to some extent.

When we are yearning most, when we are in the greatest need, we are even more vulnerable to fallacious beliefs that can serve to satisfy those yearnings.

The Input Unit

Information enters the belief engine sometimes in the form of raw sensory experience and other times in the form of organized, codified information presented through word of mouth, books, or films. We are wonderful pattern detectors, but not all the patterns we detect are meaningful ones. Our perceptual processes work in such a way as to make sense of the environment around us, but they do make sense -- perception is not a passive gathering of information but, rather, an active construction of a representation of what is going on in our sensory world. Our perceptual apparatus selects and organizes information from the environment, and this process is subject to many well-known biases that can lead to distorted beliefs. Indeed, we are less likely to be influenced by incoming information if it does not already correspond to

deeply held beliefs. Thus, the very spiritual Christian may be quite prepared to see the Virgin Mary; information or perceptual experience that suggests that she has appeared may be more easily accepted without critical scrutiny than it would be by someone who is an atheist. It is similar with regard to experiences that might be considered paranormal in nature.

The Emotional Response Unit

Experiences accompanied by strong emotion may leave an unshakable belief in whatever explanation appealed to the individual at the time. If one is overwhelmed by an apparent case of telepathy, or an ostensible UFO, then later thinking may well be dominated by the awareness that the emotional reaction was intense, leading to the conclusion that something unusual really did happen. And emotion in turn may directly influence both perception and learning. Something may be interpreted as bizarre or unusual because of the emotional responses triggered.

Evidence is accumulating that our emotional responses may be triggered by information from the outside world even before we are consciously aware that something has happened. Take this example, provided by LeDoux (1994) in his recent article in Scientific American (1994, 270, pp. 50-57):

> An individual is walking through the woods when she picks up information -- either auditory, such as rustling leaves, or visual, such as the sight of a slender curved object on the ground -- which triggers a fear response. This information, even before it reaches the cortex, is processed in the amygdala, which arouses the body to an alarm footing. Somewhat later, when the cortex has had enough time to decide whether or not the object really is a snake, this cognitive information processing will either augment the fear response and corresponding evasive behaviour, or will serve to bring that response to a halt.

This is relevant to our understanding of paranormal experience, for very often an emotional experience accompanies the putatively paranormal. A strong coincidence may produce an emotional "zing" that points us toward a paranormal explanation, because normal events would not be expected to produce such emotion.

Our brains are also capable of generating wonderful and fantastic perceptual experiences for which we are rarely prepared. Out-of-body experiences (OBEs), hallucinations, near-death experiences (NDEs), peak experiences -- these are all likely to be based, not in some external transcendental reality, but rather in the brain itself. We are not always able to distinguish material

originating in the brain from material from the outside world, and thus we can falsely attribute to the external world perceptions and experiences that are created within the brain. We have little training with regard to such experience. As children, we do learn to distrust, for the most part, dreams and nightmares. Our parents and our culture tell us that they are products of our own brains. We are not prepared for more arcane experiences, such as OBEs or hallucinations or NDEs or peak experiences, and may be so unprepared that we are overwhelmed by the emotion and come to see such experience as deeply significant and "real" whether or not it is.

Ray Hyman has always cautioned skeptics not to be surprised should they one day have a very strong emotional experience that seems to cry out for paranormal explanation. Given the ways our brains work, we would expect such experiences from time to time. Unprepared for them, they could become conversion experiences that lead to strong belief. When I was a graduate student, another graduate student who shared my office, and who was equally as skeptical as I was about the paranormal, came to school one day overwhelmed by the realism and clarity of a dream he had had the night before. In it, his uncle in Connecticut had died. It had been a very emotional dream, and was so striking that Jack told me that if his uncle died anytime soon, he would no longer be able to maintain his skepticism about precognition -- the dream experience was that powerful. Ten years later, his uncle was still alive, and Jack's skepticism had survived intact.

The Memory Unit

Through our own experience, we come to believe in the reliability of our memories and in our ability to judge whether a given memory is reliable or not. However, memory is a constructive process rather than a literal rendering of past experience, and memories are subject to serious biases and distortions.

Not only does memory involve itself in the processing of incoming information and the shaping of beliefs; it is itself influenced strongly by current perceptions and beliefs. Yet it is very difficult for an individual to reject the products of his or her own memory process, for memory can seem to be so "real."

The Environmental Feedback Unit

Beliefs help us to function. They guide our actions and increase or reduce our

anxieties. If we operate on the basis of a belief, and if it "works" for us, even though faulty, why would we be inclined to change it? Feedback from the external world reinforces or weakens our beliefs, but since the beliefs themselves influence how that feedback is perceived, beliefs can become very resistant to contrary information and experience. If you really believe that alien abductions occur, then any evidence against that belief can be rationalized away -- in terms of conspiracy theories, other people's ignorance, or whatever.

As mentioned earlier, fallacious beliefs can often be even more functional than those based in truth. For example, Shelley Taylor, in her book Positive Illusions, reports research showing that mildly depressed people are often more realistic about the world than are happy people. Emotionally healthy people live to some extent by erecting false beliefs -- illusions -- that reduce anxiety and aid well-being, whereas depressed individuals to some degree see the world more accurately. Happy people may underestimate the likelihood of getting cancer or being killed, and may avoid thinking about the ultimate reality of death, while depressed people may be much more accurate with regard to such concerns.

An important way in which to run reality checks on our perceptions and beliefs is to compare them with those of others. If I am the only one who interpreted a strange glow as an apparition, I am more likely to reconsider this interpretation than if several others share the same view. We often seek out people who agree with us, or selectively choose literature supporting our belief. If the majority doubts us, then even if only part of a minority we can collectively work to dispel doubt and find certainty. We can invoke conspiracies and coverups to explain an absence of confirmatory evidence. We may work to inculcate our beliefs in others, especially children. Shared beliefs can promote social solidarity and even a sense of importance for the individual and group.

In Conclusion

Beliefs are generated by the belief engine without any automatic concern for truth. Concern for truth is a higher order acquired cognitive orientation that reflects an underlying philosophy which presupposes an objective reality that is not always perceived by our senses.

The belief engine chugs away, strengthening old beliefs, spewing out new ones, rarely discarding any. We can sometimes see the error or foolishness in other people's beliefs. It is very difficult to see the same in our own. We

believe in all sorts of things, abstract and concrete -- in the existence of the solar system, atoms, pizza, and five-star restaurants in Paris. Such beliefs are no different in principle from beliefs in fairies at the end of the garden, in ghosts in some deserted abbey, in werewolves, in satanic conspiracies, in miraculous cures, and so on. Such beliefs are all similar in form, all products of the same process, even though they vary widely in content. They may, however, involve greater or lesser involvement of the critical-thinking and emotional-response units.

Critical thinking, logic, reason, science -- these are all terms that apply in one way or another to the deliberate attempt to ferret out truth from the tangle of intuition, distorted perception, and fallible memory. The true critical thinker accepts what few people ever accept -- that one cannot routinely trust perceptions and memories. Figments of our imagination and reflections of our emotional needs can often interfere with or supplant the perception of truth and reality. Through teaching and encouraging critical thought our society will move away from irrationality, but we will never succeed in completely abandoning irrational tendencies, again because of the basic nature of the belief engine.

Experience is often a poor guide to reality. Skepticism helps us to question our experience and to avoid being too readily led to believe what is not so. We should try to remember the words of the late P. J. Bailey (in *Festus: A Country Town*): "Where doubt, there truth is -- 'tis her shadow."

This site was created by BarelyFitz Designs (designer Patrick Fitzgerald), except where otherwise noted. Content copyright by CSICOP or the respective copyright holders. Do not redistribute without obtaining permission.

Feedback | Reverse links for this page | Translate this page

Integrated Publishing
The Most Informative Site on the Internet
Electrical Engineering Training Series

Order this information in Print | Order this information on CD-ROM | Download in PDF Format

Click here to make tpub.com your Home Page

Page Title: Principles of transmission lines
Back | Up | Next

tpub.com Updates

[Back] [Home] [Up] [Next]

Google

Search
Web
www.tpub.com

Home

Information Categories
Administration
Advancement
Aerographer
Automotive
Aviation
Combat
Construction
Diving
Draftsman
Engineering
Electronics
Food and Cooking
Math
Medical
Music
Nuclear Fundamentals
Photography
Religion
USMC

Products
Educational CD-ROM's
Printed Manuals
Downloadable Books

PRINCIPLES OF TRANSMISSION LINES

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to:

- State what a transmission line is and how transmission lines are used.
- Explain the operating principles of transmission lines.
- Describe the five types of transmission lines.
- State the length of a transmission line.
- Explain the theory of the transmission line.
- Define the term LUMPED CONSTANTS in relation to a transmission line.
- Define the term DISTRIBUTED CONSTANTS in relation to a transmission line.
- Define LEAKAGE CURRENT.
- Describe how the electromagnetic lines of force around a transmission line are affected by the distributed constants.
- Define the term CHARACTERISTIC IMPEDANCE and explain how it affects the transfer of energy along a transmission line.
- State how the energy transfer along a transmission line is affected by characteristic impedance and the infinite line.
- Identify the cause of and describe the characteristics of reflections on a transmission line.
- Define the term STANDING WAVES as applied to a transmission line.
- Describe how standing waves are produced on a transmission line and identify the types of terminations.
- Describe the types of standing-wave ratios.

INTRODUCTION TO TRANSMISSION LINES

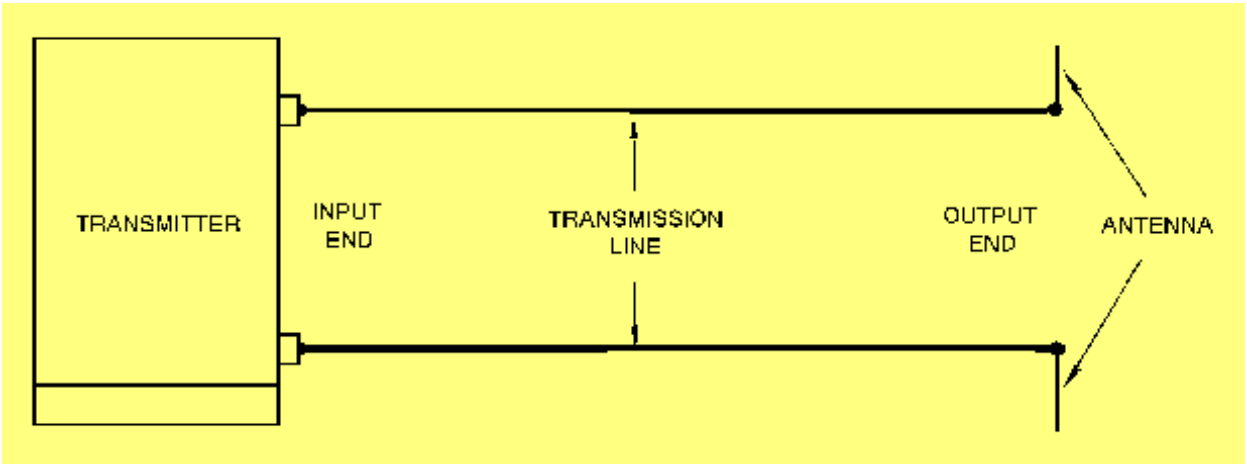
A TRANSMISSION LINE is a device designed to guide electrical energy from one point to another. It is used, for example, to transfer the output rf energy of a transmitter to an antenna. This energy will not travel through normal electrical wire without great losses. Although the antenna can be connected directly to the transmitter, the antenna is usually located some distance away from the transmitter. On board ship, the transmitter is located inside a radio room and its associated antenna is mounted on a mast. A transmission line is used to connect the transmitter and the antenna.

The transmission line has a single purpose for both the transmitter and the antenna. This purpose is to transfer the energy output of the transmitter to the antenna with the least possible power loss. How well this is done depends on the special physical and electrical characteristics (impedance and resistance) of the transmission line.

TERMINOLOGY

All transmission lines have two ends (see figure 3-1). The end of a two-wire transmission line connected to a source is ordinarily called the INPUT END or the GENERATOR END. Other names given to this end are TRANSMITTER END, SENDING END, and SOURCE. The other end of the line is called the OUTPUT END or RECEIVING END. Other names given to the output end are LOAD END and SINK.

Figure 3-1. - Basic transmission line.



You can describe a transmission line in terms of its impedance. The ratio of voltage to current (E_{in}/I_{in}) at the input end is known as the INPUT IMPEDANCE (Z_{in}). This is the impedance presented to the transmitter by the transmission line and its load, the antenna. The ratio of voltage to current at the output (E_{out}/I_{out}) end is known as the OUTPUT IMPEDANCE (Z_{out}). This is the impedance presented to the load by the transmission line and its source. If an infinitely long transmission line could be used, the ratio of voltage to current at any point on that transmission line would be some particular value of impedance. This impedance is known as the CHARACTERISTIC IMPEDANCE.

Q.1 What connecting link is used to transfer energy from a radio transmitter to its antenna located on the mast of a ship? Answer
Q.2 What term is used for the end of the transmission line that is connected to a transmitter? Answer
Q.3 What term is used for the end of the transmission line that is connected to an antenna? Answer

[Back] [Home] [Up] [Next]

Order this information in Adobe PDF Printable Format

Privacy Statement - Press Release - Copyright Information. - Contact Us -

AUDIOHOLICS.com
Online A/V Magazine
Pursuing the Truth in Audio & Video...

Where audio and video equipment undergoes rigorous objective and subjective tests by our staff, ensuring that marketing slogans aren't the only guidelines for your home theater choices.

May 29 2006

Become an AH Member!

Heard something cool? **Send Us News!**

SVS
EXPERIENCE THE NEW
SBS-1 5.1 SYSTEM from SVS

Search Audioholics

Forums: Talk With Others | Professional Product Reviews | Consumer Product Reviews | Learn About Cables & Power | Set Up Your AV System | Our Picks on What to Buy | Find the Lowest Online Prices!

Get an A/V Education
What to Buy - Our Picks
Set-Up Your System
How to Shop
Research AV Topics
Learn About Specs & Formats
Fix Your Room Acoustics
Get the Latest Industry News
Read A/V Editorials
Read Press Releases
Try a DIY Project
Latest Trade Show News
Frequently Asked Questions
Forums: Talk With Others

Read Reviews
Receivers & Amplifiers
DVD & Media Players
HDTV, Projectors & Displays
Cables & Power Conditioners
Loudspeakers
Misc Equipment & Software
DVD Movie Reviews
CD/SACD/DVD-Audio Discs

Buy & Sell Online
Buying Guides - How to Shop
Recommended Online Stores
Find the Lowest Price Online
Sell your Gear - FREE!

About Audioholics
Our Mission Statement
Our Reference Systems
Read Staff Bios
Recommended Manufacturers

Audioholics Featured Articles

Axiom M80s "WOW" Audioholics

"The Axioms sounded big and bold. The bass was tight and fast. It was almost incomprehensible the amount of bass radiating from the small diameter woofers. The midrange had a fundamental rightness, never sounding over-exaggerated, compressed or dry. It

Get the Optoma HD72 - w/FREE Lamp!

Visual Apex, Inc - the AV world's best kept secret for the past five years - is offering an incredible price on the Optoma HD72 home theater projector.Long awaited by HT enthusiasts, the Optoma HD72 has not disappointed. It has enjoyed rave reviews across the board for its stunning 5000:1 contrast ratio and 1280x728 resolution. Visual

Onix x-ls Speakers!

"If you're in the market for a 2-way system and planning on including a subwoofer in your purchase, the Onix x-ls is definitely worthy of serious consideration. Given the remarkable-for-its-class collection of strengths and the budget-respecting price tag, AV123 has crafted a formidable product that's musically satisfying and sure to provide years of listening enjoyment..."

Read the Review

Buy the Speakers!

Contact the Staff
Advertise With Us
Get Cool Audioholics Stuff

Audioholics Supports

Audioholics Editorials

Reaping what you Sow
Monday, May 29 2006

World's Smallest Projector - In Your Eye!
Wednesday, May 24 2006

Exploration of the Best Rock Saxophone Solos
Wednesday, May 24 2006

Pioneer's Inno Rocks the RIAA Boat
Wednesday, May 17 2006

consistently sounded lush and full. I loved the sound of the M80tis and could not find any fault with the sound of these speakers. These speakers have wow and then some WOW!"

Read the Review | Buy from Axiom

Apex as an authorized DIRECT Optoma dealer will ship a free lamp with your HD72 purchase while supplies last (a $399 value) AND include an additional 1yr warranty!

Read the Review | Buy It from Visual Apex

Get our Newsletter!

Sign Up!

Visit our HTA Partners:
HomeTheater Forum
AudioREVIEW

Audioholics Daily Dose of News & Reviews

Editorial: Reaping What You Sow

Posted Monday, May 29 2006, by tom

Being a fairly sarcastic person (and by fairly I mean very), one has to be aware of the karmic toll one's actions are sure to elicit. You can only go around saying whatever you want in the form of a "joke" for so long before it catches up with you. I've spent my life watching words come out of my mouth thinking, I shouldn't have said that, while the retribution begins to crush me under its imposing and fully justified wrath. Examples? How 'bout the time I told a girlfriend that her "signature" dish that she worked on all day tasted just like that taste you get in your mouth after throwing up. Or the time some friends came over with a guy from another country and after a few hours of saying nothing to anyone, I looked over at him and said, "So why are you here anyways?"

[Read More] [Discuss]

ProjectorPeople.com Free Lamp Offer!

Posted Sunday, May 28 2006, by clint

Get a free replacement lamp with select Optoma home theater projectors including the HD78, HD72, and even the brand new

Popular Articles

Denon AVR-5805 Receiver

Fixing Ground Loops

Recommended System Guides

AV123 Rocket UFW-10 Subwoofer Review

Denon DVD-3910 Review

Carada Criterion Projector Screen Review

Yamaha YSP-1 Digital Sound Projector Review

Format Wars: Feel the Angst
Wednesday, May 17 2006

Flushing Out Myths In Consumer Audio
Monday, May 15 2006

IPTV - The Sky is Falling!
Friday, May 5 2006

Industry Press Releases

ELAN Home Systems Ships New A1240 Amplifier
Tuesday, May 16 2006

ReplayTV to Debut "PC Edition" at Digital Experience in New York
Tuesday, May 16 2006

Niles Introduces Subwoofers
Thursday, May 4 2006

VESAs DisplayPort to Compete with HDMI
Wednesday, May 3 2006

Mitsubishi XL5980U Anti-Theft Projector
Tuesday, May 2 2006

Belkin N1 802.11n Wireless Networking in June
Monday, May 1 2006

Pixel Magic Intros 1080p HD MediaBox
Saturday, April 29 2006

HD7100! ProjectorPeople.com brings it to you. That is a savings of up to $300 and makes getting into a front projection system all the more easier. This offer ends soon. Click below for details.

[See the Deal]

Radiient Elara Theater Speaker Package Review

Posted Thursday, May 25 2006, by gds

What do you say about speakers at this price range ($499 for a set of five)? They sound pretty darn good to my ears and I think that many would be pleased as punch to own them. The aesthetics are way above average though a bit higher quality veneer would have been appreciated. I can definitely see these working in a bachelor pad where you'd like to impress your friends without breaking the bank and you don't have to share the sweet spot with anyone.

[Read the Review] [Discuss]

Onix SP3 Tube Integrated Amplifier Review

Posted Thursday, May 25 2006, by gds

What's this, a Tube amplifier review on Audioholics??? Yes for all of you audiophiles that love chocolatey midrange, this product is just for you. The quality of the sound from the SP3 has everything that draws aficionados to prize tubes; which is uncommon at this price. Silky smooth sound that remains composed even during hard clipping. An art deco style chassis makes for one enjoyable audiophile product and conversational piece that will get noticed by Audiophiles and Audioholics alike.

[Read the Review] [$1,199 SP3 + Ref1 Promo Deal] [Discuss]

World's Smallest Projector - In Your Eye!

Posted Wednesday, May 24 2006, by clint

We all like small. I mean, come on - thinner TV screens, smaller parts for a more compact design... About the only thing we don't want small is our diagonal viewing area - assuming I'm not talking

RBH Sound CinemaSITE

Understanding Widescreen, Letterbox and Pan & Scan

Principles of Cables & Interconnects

Component Video Cables A Guide

Hot Forum Topics

Samsung HL-S4666W
posted by godfatherofsoul *05/29/06 11:37 AM*

DENON 3805 - IPod and XM
posted by zildjian *05/29/06 11:37 AM*

Upgradeitis.... Should I or Shouldn't I?????
posted by RMK! *05/29/06 11:27 AM*

semi advanced question
posted by t3031999 *05/29/06 11:23 AM*

Enter to Win: Music from Impact Acoustics
posted by Rickster71 *05/29/06 11:12 AM*

Axiom alternatives
posted by littleb *05/29/06 11:03 AM*

Doubling up unmatched subs
posted by Geoff *05/29/06 10:55 AM*

about certain spouses that is. Well, they've finally gone and done it. The world's smallest projector was successfully created and tested - and it fits in your eye. This machine lets partially blind patients read and explore virtual buildings by projecting images directly onto their retinal cells.

[Read the Editorial] [Discuss]

Exploration of the Best Rock Saxophone Solos

Posted Wednesday, May 24 2006, by gds

Impact Acoustics takes us on a tour of the most memorable saxophone solos in rock. From Steely Dan to Pink Floyd to Inxs and much more. Peruse their trip down saxophone memory lane and let us know other notable artists and songs they may have missed? What songs have they omitted that you just can't understand how they could have forgotten? Hurry up and let them know by May 31st to be eligible to win free music.

[Read the Editorial] [Enter the Contest to Win Free Music]

Jeff Lorber: Worth Waiting For (!993) CD Review

Posted Wednesday, May 24 2006, by gds

Fans of the 1970's group Jeff Lorber Fusion (like me) were thrilled when, after a seven year hiatus from solo recordings, he returned to his fusion jazz roots with this release that was, well, *Worth Waiting For*. Sure, he was busy writing and producing for other (mostly vocal) artists during the late 1980's and early 1990's, but even his own break-through commercial success *Private Passion* (with female vocalist Karyn White) I think belied HIS true passion: fusion jazz.

[Read the Review] [Discuss]

Sony does not think they can meet the Blu-ray demand
posted by MBauer *05/29/06 10:44 AM*

DVI vs COMPONENT
posted by hd newbie *05/29/06 10:26 AM*

Speaker Set-up
posted by shokhead *05/29/06 09:42 AM*

Read Archived News...

designed & hosted by CD MEDIA INCORPORATED

...to Perfect Bass
Velodyne

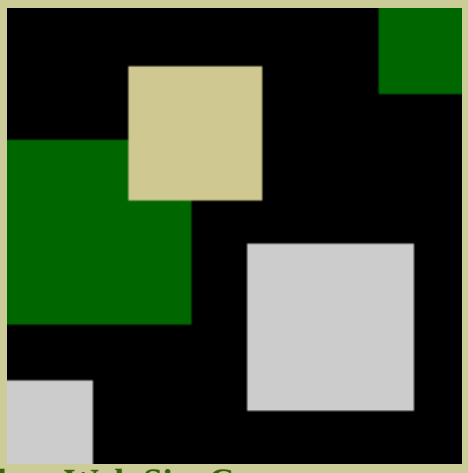
SPL-R

Join the Home Theater Alliance Newsletter! Sign Up!

Copyright ©2005 Audioholics.com. All Rights Reserved. Reproduction rights, terms & conditions. Privacy Policy.

Home | Forums | Pro Product Reviews | Get an AV Education | Cables & Power Conditioners | About Audioholics | What to Buy - Our Picks | Contact Audioholics
Advertise With Us | Consumer Reviews | Recommended Stores | FREE Audio Classifieds | Trade Show News | Find the Lowest Price Online | Get Audioholics Stuff

Jon Risch's Audio DIY Notes & Tweak Info



Link to Web Site Content:
Main Link Page

Contact Info:

Name: **Jon Risch**

E-mail: **j_risch@bellsouth.net**

NOTE: NEW E-MAIL ADDRESS, NOT ALL PAGES HAVE BEEN UPDATED YET!

This is the start page.
NOTE: NEW E-MAIL ADDRESS, NOT ALL PAGES HAVE BEEN UPDATED YET!

j_risch@bellsouth.net

(Not a LINK)

Goertz Intro
Alpha-Core's Serpents
Goertz MI Speaker Cables
Alpha Core Interconnects
BP-30 Balanced, Symmetrical Power Supply
Reviews and Testimonials
Cables FAQ
Retail Price List
Alpha-Core Audio Store
Locate a Dealer
Alpha-Core Home
Send e-Mail to Alpha-Core

Goertz MI™ Speaker Cables

Click here to buy on line!

New! @

The Lowest Inductance In The Business

The patented Goertz MI cable, invented by Ole Goertz, a Danish electronics and audio engineer, eliminates distortion between amplifier and loudspeakers. Patented in 20 countries, Goertz wire is flat by design. It's form follows function, in the best design tradition. We set out to create an electrically correct pathway between amplifier and loudspeaker, which required replacing typical cable psuedo-science with solid physics. Research showed that the perfectly flat shape, with minimum space between conductors, was the best way to achieve this.

But developing such a wire is a different proposition. Many audio companies do not even manufacture their own wire; packaging and marketing the products of others instead. As a primary maker of high-tech laminated conductors for industrial electronics applications, Alpha-Core was uniquely suited to transform the promising Goertz design into reality. Using the dielectrics Polyester Terapthalate and Teflon applied around each solid rectangular conductor, we achieved an extraordinarily thin profile between the two stacked wires. To provide maximum protection while maintaining thinness, the cable sandwich is then sheathed in tough, high grade polymer.

The result is a wire with a characteristic impedance that closely matches the impedance of loudspeakers. Alpha-Core's Goertz MI wire presents 2 to 4 ohms characteristic impedance as opposed to the 50 to 200 ohms presented by even the most expensive speaker cables. This correct electrical matching results in improved stereo imaging, better clarity, less high-frequency roll off and better low frequency definition.

Goertz Improves all Frequency Ranges:
LOW . . . MIDRANGE . . . HIGH
The degree of distortion along a loudspeaker cable depends on its resistance (**R**), inductance (**L**) and capacitance (**C**). Resistance is dependent on length and conductor cross section. A low resistance provides superior damping between the amplifier and the speaker diaphragm, governing its ability to convey a true, powerful bass and lower mid-range. A high cable inductance causes loss of signal strength towards the high end of the audible range (roll-off). It also causes frequency dependent phase-shift which disturbs true stereo imaging. The square root of the ratio L/C is termed Characteristic Impedance, and is a quality inherent in a cable, dependent on its geometry and materials, but not dependent on its length.

Goertz Virtually Eliminates Mismatch
The characteristic impedance of the Goertz MI cables in the order of 2 to 4 ohms closely matches the impedance of loudspeakers. Almost all other speaker cables have characteristic

impedance ranging from 50 to 200 ohms, a mismatch which causes distortion due to signal reflections. Impedance matching primarily improves the clarity of the highs and upper mid-range, but many users have also experienced improvements in the lower mid-range. The cause seems to be that signal reflections caused by impedance mismatch enter the feedback loop of many amplifiers and disturb their ability to reproduce faithfully even lower frequency signals.

Thick on Benefits, Thin as a Penny
The low inductance and characteristic impedance of Goertz cables are due to their flat construction and the nature of their insulation and jacketing material. The configuration places the heavy, solid conductors only 0.003 inches apart. The cables are 0.4 or 0.8 inches wide, and only 0.04 inches thick. As thin as a penny.

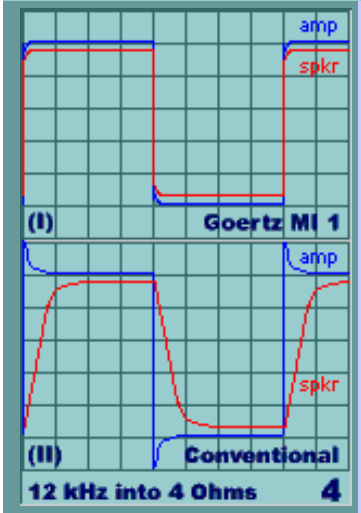
Perfect for Home Theater
Goertz MI speaker cable and interconnects are perfect for home theater applications. Due to the low inductance characteristics, Goertz audio products are ideal for long runs. No more positioning your equipment based on the limitations of your cable system. Dolby Digital home theater design requires full range audio frequencies be sent to rear speakers; Goertz cable allows extreme accuracy at any length. The space ratio, the relationship between conductor cross section, is about 50% in Goertz MI cables versus 5-20% in most other speaker cables. In other words, Goertz cables pack heavy conductor cross sections into an extremely slim package yielding greatly improved highs and mid-range, plus a truly powerful bass. Thanks to the flat construction and extreme compactness, they can be drawn anywhere and routed under thin carpets and wallpaper. If desired, paint with oil or water based paint.

What makes them Special?

Figure 1 is Goertz MI 1 (MI 2 is almost the same thickness, but twice as wide). Each conductor is a 0.375 by 0.010 inch solid band of high purity oxygen free copper (OFC) surrounded by a 0.0051 inch thick sheath of the space age insulation polyester terapthalate. The conductors are placed in close mutual contact within a 0.005 inch thick clear jacket. The mutual distance between the conductors is 0.003 inches and the overall dimensions of the cable are approximately 0.4 by 0.040 inches. The space ratio is close to 50%. The conductor cross section approximates 13 AWG, shown approximately twice actual size.

Figure 2 is a section through a competitor's cable extruded from clear PVC, having two stranded conductors of many fine copper wires. The space ratio of this style cable varies between 10% and 20%.

Figure 3 is a so called "multiple bandwidth" or "time compensated" cable with two larger and two smaller conductors twisted together inside a common jacket. The two differently sized conductors are used in parallel. The manufacturer claims that the treble components will travel along the lighter conductor, and the bass frequencies along the heavier conductors. The space ratio is 5% to 10%.



The Proof: Oscilloscope readings comparing the performance of a Goertz cable with conventional designs.

Figure 4 illustrates comparative measurements using a 12 kHz square wave transmitted via (I) a 25 foot Goertz cable, and (II) a 25 foot speaker cable of conventional bi-wire construction, both driving a 4 ohms loudspeaker. The characteristic impedance of the two cables are (I) 4 ohms and (II) 100 ohms. (A) is the signal at the amplifier, and (S) is the signal at the speaker.

Time Base = 10µS
Channel 1 V/DIV = 0.5 V
Channel 2 V/DIV = 0.5 V

It is evident that (I) is entirely distortion free, showing only the unavoidable power loss due to cable resistance. (II) in contrast indicates a clear leading edge spike at the amplifier, and significant distortion or roll off at the speaker terminals.

Top to Bottom:
AG 3 with Silver Spades
MI 3 with Silver Spades

View 48k Family Picture

Goertz MI 1(Copper)/AG1 (Solid Silver)
The original Goertz cable, MI 1 in 13 gauge (AWG) *solid Oxygen Free Copper* (OFC) or 12 gauge *AG 1 in solid High Purity Silver*. The heavy, high purity Oxygen Free Copper (OFC) or solid high purity silver ribbons are sandwiched together, separated only by micro-thin layers of film insulation. The minute amount of active dielectric also provides rapid break-in; now audiophiles don't have to wait tens of hours to break in their cables, Goertz cables produce immediate results!

Goertz MI 2(Copper)/AG2 (Solid Silver)
MI 2 Heavy 10 gauge OFC copper or 9 gauge AG 2 solid silver cabling for improved bass and power handling over the remarkable Goertz MI 1. Great reviews confirm that this remarkable cable really delivers; superb bass control, vivid mid-range, and excellent high frequency extension combine with rejection of interference and no break-in period. Simply great value.

Goertz MI 3 (Cooper)/AG3 (Solid Silver)
Alpha-Core's "Big Boy" 7 gauge speaker cable; the answer for powerful audiophile systems. MI 3 was designed for high power amplifiers driving full range speakers. Perhaps the best speaker cable made today, the MI 3 delivers sublime clarity, dynamics, and definition throughout the audio spectrum. The extreme low DC resistance ensures diaphragm control for tight, powerful bass and mid-range, and the low inductance carries distortion free highs frequencies and crucial harmonics for outstanding imaging and depth perception.

Goertz 7 gauge MI 3 cables were developed at the request of Dan D'Agostino of Krell Industries and other customers who were seeking a speaker cable to complement today's powerful amplifiers. Double the cross section of Goertz MI 2 and AG 2, this newcomer features two solid conductors each measuring 1.25" by 0.012". As in Goertz MI 1 and MI 2, the micro-thin film dielectric means minimum break-in time.

The 7 AWG cable is available in two versions: The MI 3 'Divinity' in OFC copper, and the AG 3 'Divinity' in solid fine silver. Both are supplied with four inch user-friendly pigtails, and both feature Alpha-Core's unique solid fine silver spade connectors to fit all high-end binding posts. Goertz AG 3 is arguably the heaviest solid fine silver cable ever made. An eight foot pair contains more than a pound of 99.99% pure silver. So, for that extra muscle to drive the really powerful systems, Goertz MI 3 or MI AG 3 is the answer, The Alpha-Core "Big Boys" deliver punch in the bass and a sound so revealing that we maybe ought to mark them "Adults Only".

Specifications

	Conductors/ Cross Section/ Equivalent Gauge	Resistance/ Capacitance/ Inductance/ Impedance	Dielectric Material/ Outside Dimensions	Lengths/ Pricing Links
Goertz MI 1: 'Center Stage'	High Purity Oxygen Free Copper .375 x .010" 13 AWG	0.0044 ohms/ft 0.5nF/ft 10nH/ft ~ 4 ohms	Polyester Terepthalate/ Polycarbonate 0.4 x 0.04"	**4 - 20 feet terminated / 25/150/500 foot bulk spools.**
Goertz MI 2: 'Veracity'	High Purity Oxygen Free Copper .750 x .010" 10 AWG	0.0022 ohms/ft 0.95nF/ft 6nH/ft ~ 2.5 ohms	Teflon/ Polyester Terepthalate/ Polycarbonate 0.8 x 0.04"	**4 - 20 feet terminated / 25/75/250 foot bulk spools.**
Goertz MI 3: 'Divinity'	High Purity Oxygen Free Copper .1.25 x .012" 7 AWG	0.0011 ohms/ft 1.5nF/ft 4nH/ft ~ 1.7 ohms	Teflon/ Polyester Terepthalate/ Polycarbonate 1.3 x 0.08"	**4 - 20 feet terminated.**
Goertz AG 1: 'Center Stage'	High Purity Solid Fine Silver .375 x .010" 12 AWG	0.0034 ohms/ft 0.5nF/ft 10nH/ft ~ 4 ohms	Polyester Terepthalate/ Polycarbonate 0.4 x 0.04"	**4 - 20 feet terminated / 25/150/500 foot bulk spools.**
Goertz AG 2: 'Veracity'	High Purity Solid Fine Silver .750 x .010" 9 AWG	0.0017 ohms/ft 0.95nF/ft 6nH/ft ~ 2.5 ohms	Teflon/ Polyester Terepthalate/ Polycarbonate 0.8 x 0.04"	**4 - 20 feet terminated / 25/75/250 foot bulk spools.**
Goertz AG 3: 'Divinity'	High Purity Solid Fine Silver 1.25 x .012" 7 AWG	0.00098 ohms/ft 1.5nF/ft 4nH/ft ~ 1.7 ohms	Teflon/ Polyester Terepthalate/ Polycarbonate 1.3 x 0.08"	**4 - 20 feet terminated.**
Temperature Rating:	**-40°F to +270°F | Maximum Operating Voltage: 300 Volts**			
Termination Styles:	**Bananas/Copper Spades/Pins/Silver Spades/Hybrid/T (auto-style) in Single and Bi-wire Configurations**			
	Goertz speaker cables are available in pre-terminated pairs in standard lengths (2 foot to 20 in 1 foot increments). The electrical characteristics of Goertz cables make them ideally suited for even longer runs.			

209

The Self Site

Click here for direct access to page.

THE AMPLIFIER INSTITUTE

17 Nov 2005 UPDATED:
A New Failed Project...

THE MUSEUM OF RETROTECH

7 May 2006 UPDATED
Monowheels

8 May 2006 UPDATED
Diwheels

15 May 2006 UPDATED AGAIN
Thermo-Electric Generators

UNUSUAL STEAM LOCOMOTIVES

21 May 2006 UPDATED
The Heilmann Locos

25 May 2006 UPDATED
Petiet's Locos

Comments to:
selfsite@dsl.pipex.com

Audio Design & Technology

The Museum of RetroTech

Unusual Steam Locomotives

Genealogy: Selfs & Perfects

Roswell Incidents: Goddard

News Links Log

SPONSOR LINKS

Mountain Warehouse

Racing Green

Fiorelli

Copyright © Douglas Self 1998-2006. All rights reserved. Tested on IE6.0 only.

Elliott Sound Products **Class A Amplifiers - Part 2**

Copyright © 2005 Rod Elliott
Page Created 04 Apr 2005

Main Index

Articles Index

Class-A Part 1

Introduction

As discussed in Part 1, there has been a resurgence of two "ancient" technologies - vacuum tube (valve) amplifiers and Class-A systems. The big question is .... is there a difference? This part of the discussion looks at the differences in the output stages only - no feedback is used in any of the following, so the performance of the final stage can be assessed.

Class Comparison

The Class-AB output stage is essentially the same as that used in Project 3A), and this is compared with a Class-A emitter follower circuit using exactly the same circuitry. Both were operated with zero feedback, and the Class-AB stage was run at a quiescent current of 14mA vs. 2A for the Class-A circuit.

The Class-A version was operated using an ideal current source - a real current source will not be quite as good, however the difference in practical terms is very small (this was verified in a simulation).

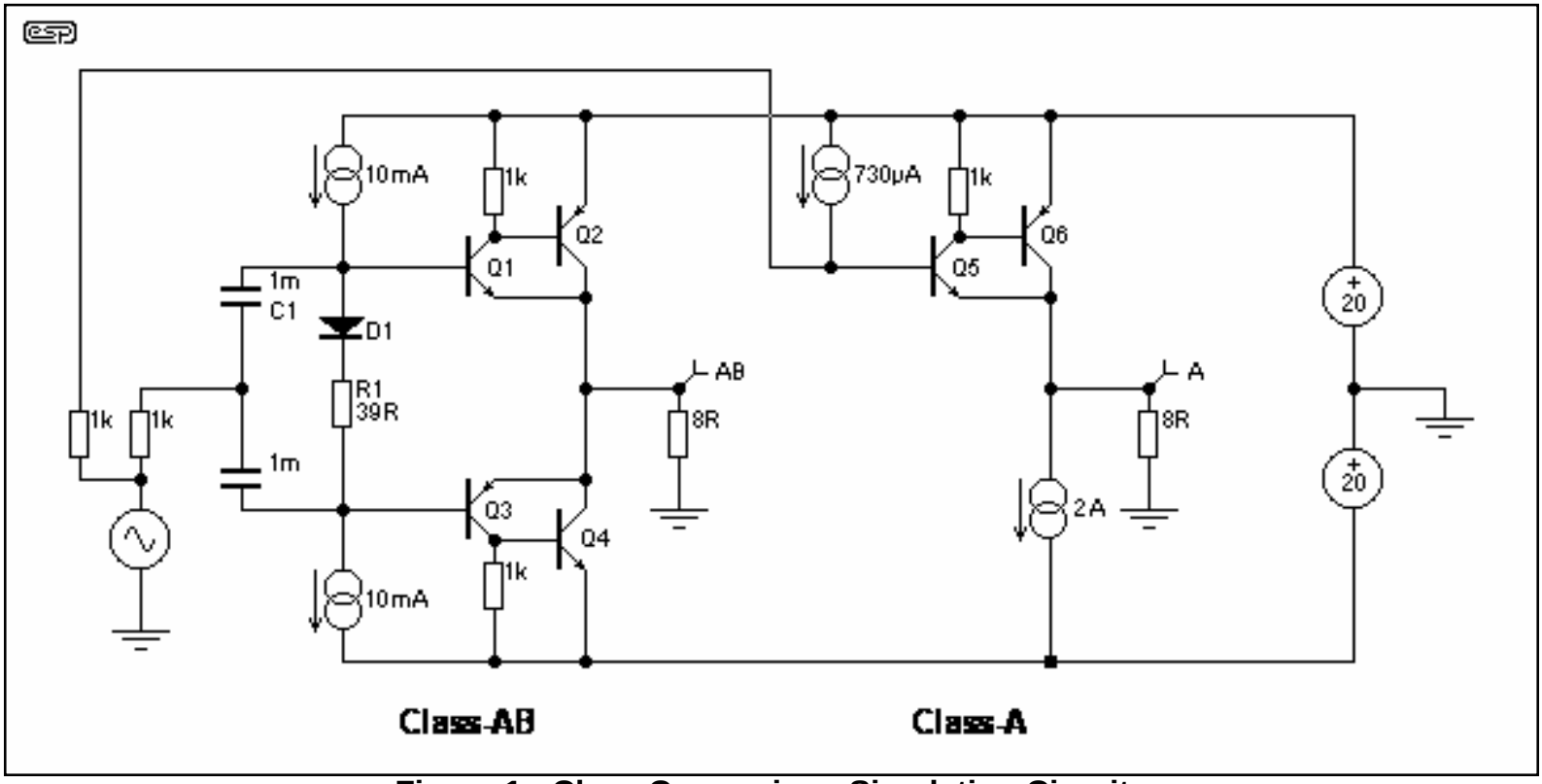


Figure 1 - Class Comparison Simulation Circuits

A description of the circuits is in order before we continue. The Class-AB output stage uses current sources to bias the bases, and 1,000uF coupling caps to the signal source. This is simply to allow the circuit to be biased correctly for the simulation, and has no effect on distortion. Because there is a small difference between the positive and negative halves of the circuit, the current sources were actually very slightly unbalanced to get less than 10mV DC offset at the output.

Quiescent current is set by the diode and 39 Ohm resistor, and was just over 14mA for the simulations used in this article.

The Class-A amp uses the same transistor configuration, but the bias is supplied by a much smaller current source - again, to obtain less than 10mV DC offset. I used a 2A current source rather than semiconductors to ensure that the results were not adversely affected by an imperfect current sink. Naturally, this choice is removed for a real amplifier.

Different results would be obtained using a Darlington emitter follower, however the difference in real terms is far less than will be caused by actual (as opposed to simulated) transistors. A quick check indicates that the Darlington configuration has slightly less second harmonic distortion, and less distortion overall. This will not necessarily be the same in a real amplifier.

It is worth noting that although the Class-A amp performs better with a Darlington, the

Class-AB circuit is significantly worse - even with higher quiescent current - this has been known for quite some time, but is still ignored by many designers.

Test Results

The first test was at close to full power - 15V peak (30V P-P) was used to ensure that both circuits were well away from clipping, as this would affect the results making a proper comparison impossible.

Logically, one would expect the Class-A emitter follower to be better (have lower distortion) but it doesn't! Distortion is 0.41% for the Class-A version, and 0.098% for the Class-AB. Spectral content shows that the Class-A circuit has higher harmonic levels than the Class-AB circuit up to about 6kHz, and it is only after this frequency that the Class-A amp shows an advantage. Even in this form (widely believed to have predominantly second harmonic distortion), there is considerable third harmonic distortion, as well as 4th, 5th, etc. Distortion falls to very low levels (-100dB) at 6kHz (6th harmonic).

By contrast, the Class-AB circuit has lower distortion overall, with low order harmonics almost an order of magnitude lower than Class-A. However, the spectral graphs show that high order harmonics are at a much greater level. Where the eighth harmonic was 100dB down in the Class-A circuit, the 31st harmonic was 100dB down in Class-AB. Which is preferable? I shall leave that up to the individual reader. I know for a fact that I can't hear distortion that is -100dB ... perhaps some people can, but I've not met any.

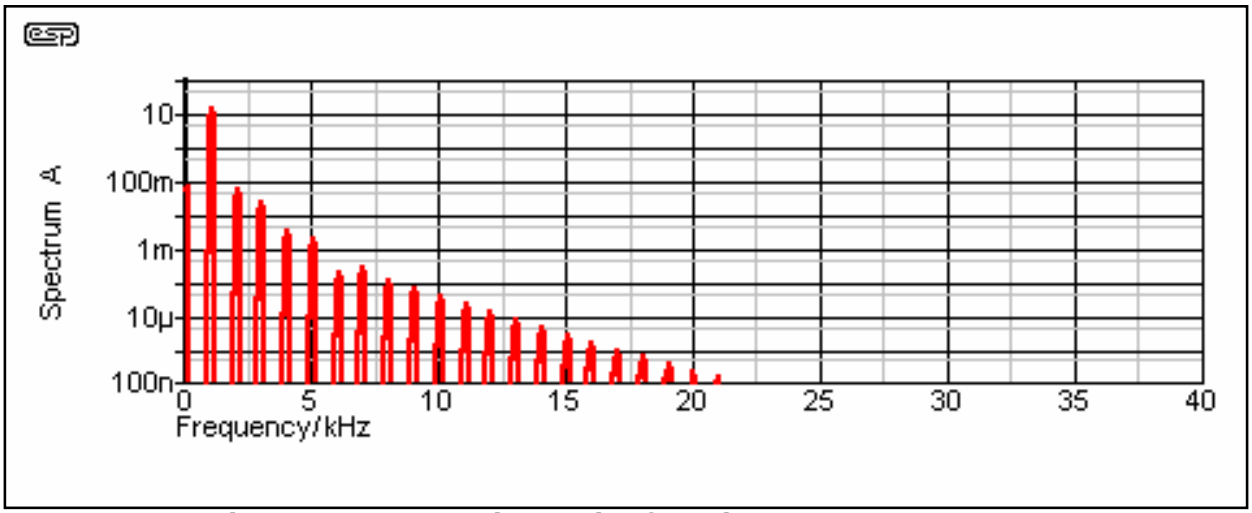


Figure 2 - Spectral Results for Class-A Output Stage

As you can see, the second harmonic (2kHz) is about 48dB below the fundamental, and the remaining harmonics are below the -100dB level by 8kHz (the 8th harmonic). The level at 8kHz is actually -102dB referred to the fundamental. This cannot be considered a bad

result. With the addition of feedback, these components will be reduced further, but it is also to be expected that any 'real-world' semiconductors will perform worse than their simulated counterparts.

Note that this is a single-ended Class-A stage, and it is this very configuration that is said to have predominantly even order harmonics (second, fourth, etc.). Instead of what is normally expected, we have an even progression of harmonics, with both odd and even represented in an orderly progression. To obtain predominantly even order harmonics, common 'wisdom' claims that the output stage should be operated as a common emitter amplifier (an emitter follower as shown is common collector). This is analogous to a common cathode stage as used in most single-ended triode (SET) amps.

Use of a common emitter stage will increase the levels of distortion dramatically - a quick test indicates that using the same topology for the stage but operating it as common emitter increases distortion to 8.56% at 30V P-P output, and it is still 1.61% at 2V P-P (these are the same levels used for all tests described in this article). This is hardly an encouraging result, but it does match the typical distortion levels of a SET Class-A amp. Adding some emitter feedback will reduce this very high distortion to something more tolerable at low levels (I measured 0.23%), and will also reduce distortion at higher voltages/ powers. At a little under 30V P-P, distortion was 3.8% - hardly awe inspiring. There is still a significant number of odd order harmonics at all power levels, and it is probably fair to say that very few (if any) amplifiers actually produce *only* even order harmonics.

It is easy to eliminate even order harmonics, but very difficult to eliminate odd ordered ones. This is despite any advertising material that claims the contrary.

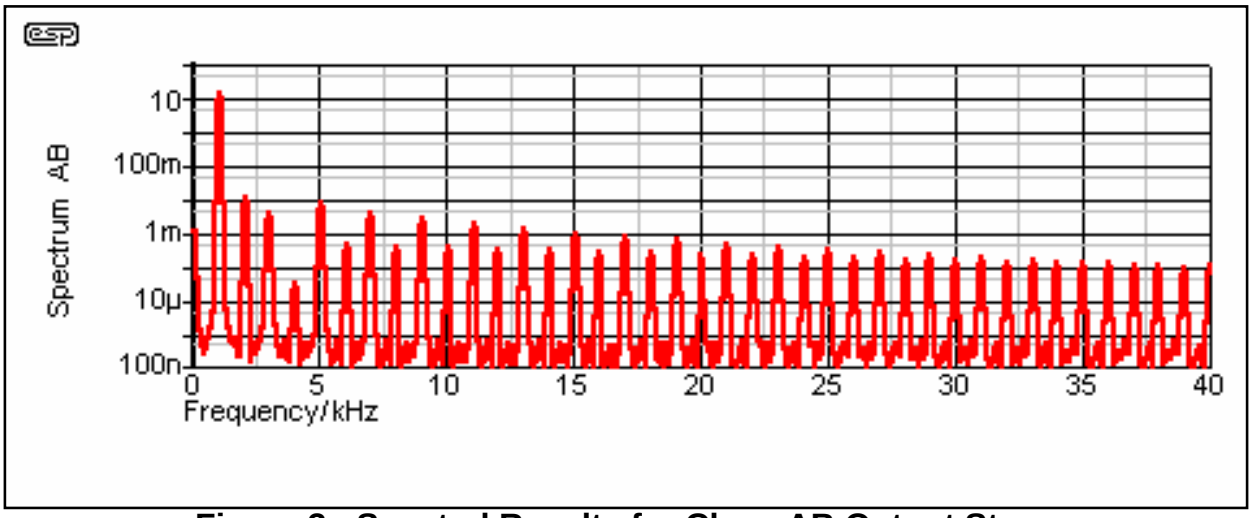


Figure 3 - Spectral Results for Class-AB Output Stage

For Class-AB, the second harmonic is 63dB below the fundamental, but there are higher order harmonics present at ever diminishing levels. Although it is unlikely that any of these would be audible, it is just this effect that some people claim ruins the sound. Again, feedback will reduce these levels, and in both cases shown, the system noise will help to mask any distortion products. Do not expect real power transistors to equal these results,

because it is almost a certainty that they won't (as with the Class-A example above).

Bear in mind that the 1mV level shown represents -83dB, and the 100nV minimum level equates to -163dB referred to the 15V peak level at 1kHz. A great many of the harmonics displayed are so far below the noise floor that no affordable test instrument would be capable of resolving them (and this includes the ears of the vast majority of people).

Low Power Operation

It is at low levels that we see some differences, and the performance differences are more like we might expect. For the Class-AB circuit, it is important to ensure that the test level is high enough to ensure that the transistors will turn off completely - at very low levels the circuit will operate in Class-A, and this would give us an unrealistic distortion level.

With 1V peak output (707mV RMS, or 2V P-P), this ensures that the Class-AB circuit is working as Class-AB. Now we see the effects of crossover distortion - even though it is not visible on a waveform graph. Total Harmonic Distortion (THD) is now 0.245% for Class-AB, but a minuscule 0.0195% for Class-A.

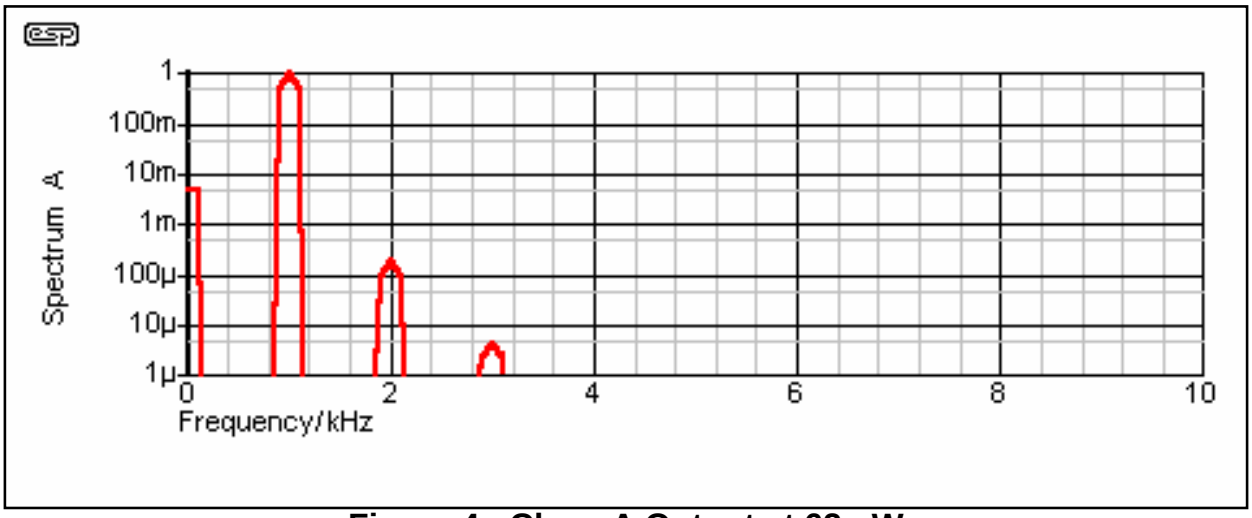


Figure 4 - Class-A Output at 62mW

Here we see the harmonic content at 62mW output level. There are only a couple of low order harmonics present, with everything beyond 4kHz well below audibility. The minimum level of 1uV is 120dB below the signal level, and can be considered so far below audibility that it is not worth considering.

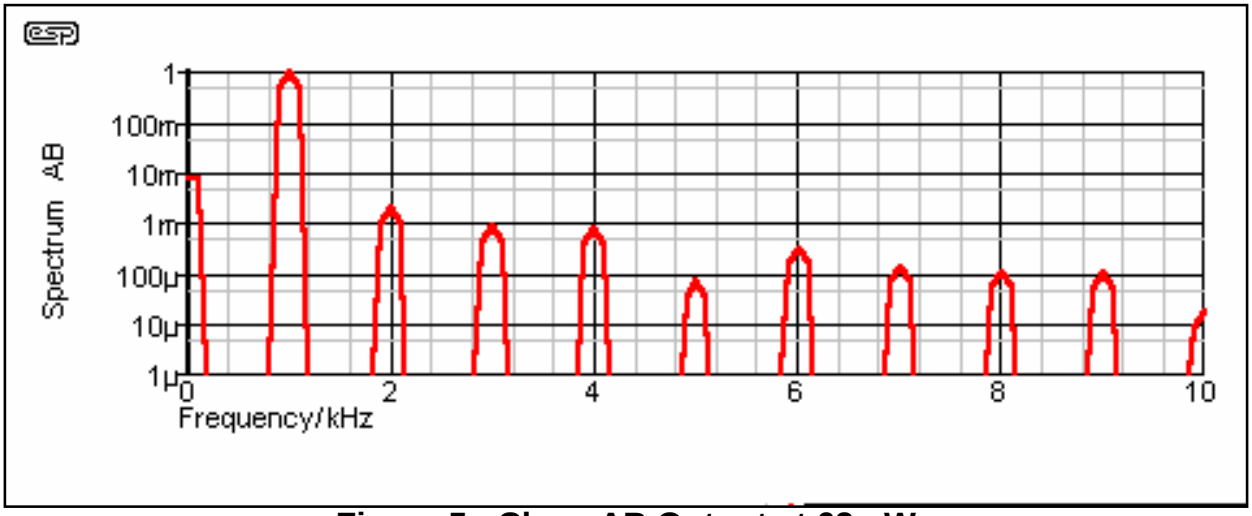


Figure 5 - Class-AB Output at 62mW

Things are not quite so rosy here. However, with an overall THD of 0.245% it still cannot be considered a bad result. Everything beyond 4kHz is more than 60dB below the signal level, and that will be close to the amplifier noise floor, and well below the listening room noise.

To explain that last point, assume a typical high-efficiency loudspeaker, with 100dB/W/m (very high efficiency by today's standards). 62mW is 12dB below 1W (close enough), so the peak level at 1 metre in the room is 88dB SPL. Since the upper harmonics are all more than 60dB below the fundamental, that makes them less than 28dB SPL. A very quiet room indeed would be needed to be able to hear a signal at only 28dB SPL, and it would probably be impossible in the presence of the fundamental.

Conclusion

As shown above, Class-A will give vanishingly low levels of distortion at low levels. Based on many amplifier tests over the years, I also know that a Class-AB amp can do exactly the same, but at all levels below clipping (not just low levels). Whether the overall advantages and disadvantages of a Class-A amp are worth the effort is entirely up to you - if having one makes you feel better (and especially so if you build it yourself), then there is no reason to avoid having one.

Of course you will need sensitive speakers, otherwise the available power at any sensible dissipation is just too low to be useful. In a biamped or triamped system, Class-A is ideal for tweeters. Will you hear a difference? I don't know - it depends on the amp, your hearing, and whether you perform a true blind test so you are completely unaware which amplifier you are listening to at any time during the test.

Main Index

Articles Index

Class-A Part 1

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created as an addition to the original Class-A article - 05 Apr 2005

The Audio Pages

Elliott Sound Products **Project 101**

High Power, High Fidelity MOSFET Power Amplifier

Rod Elliott (ESP)

Updated 01 Feb 2004

PCBs are available for this project. Click the PCB image for details.

Introduction

In various parts of The Audio Pages, I have said that I am not a fan of MOSFET power amplifiers. Well, this amp has changed my views, and I consider this to be a "reference" system in all respects. The performance is extremely good, with vanishingly low distortion levels, plenty of power, very wide full power bandwidth, and the "self protecting" nature of the MOSFETs themselves.

This is not to suggest that the amp is indestructible (no amplifier can make that claim successfully), but it is much more tolerant of faults than a bipolar transistor amp, and requires nothing more than a pair of zener diodes to limit the current. Having said that, I would still recommend that you avoid shorted output leads and the like - i.e. Don't push your luck ;-)

One thing that has emerged that is absolutely critical is the PCB layout. The layout of this new amplifier is similar to that used for the P68 Subwoofer amplifier, and this has some major benefits. P68 has no right to sound as good as it does, and although designed for subwoofer use, it has proven during listening and testing to be a very low distortion design - despite the Class-B output stages. All PCB tracks in the input and driver section are as short as possible, minimising the chance of noise pickup from other sections of the circuit.

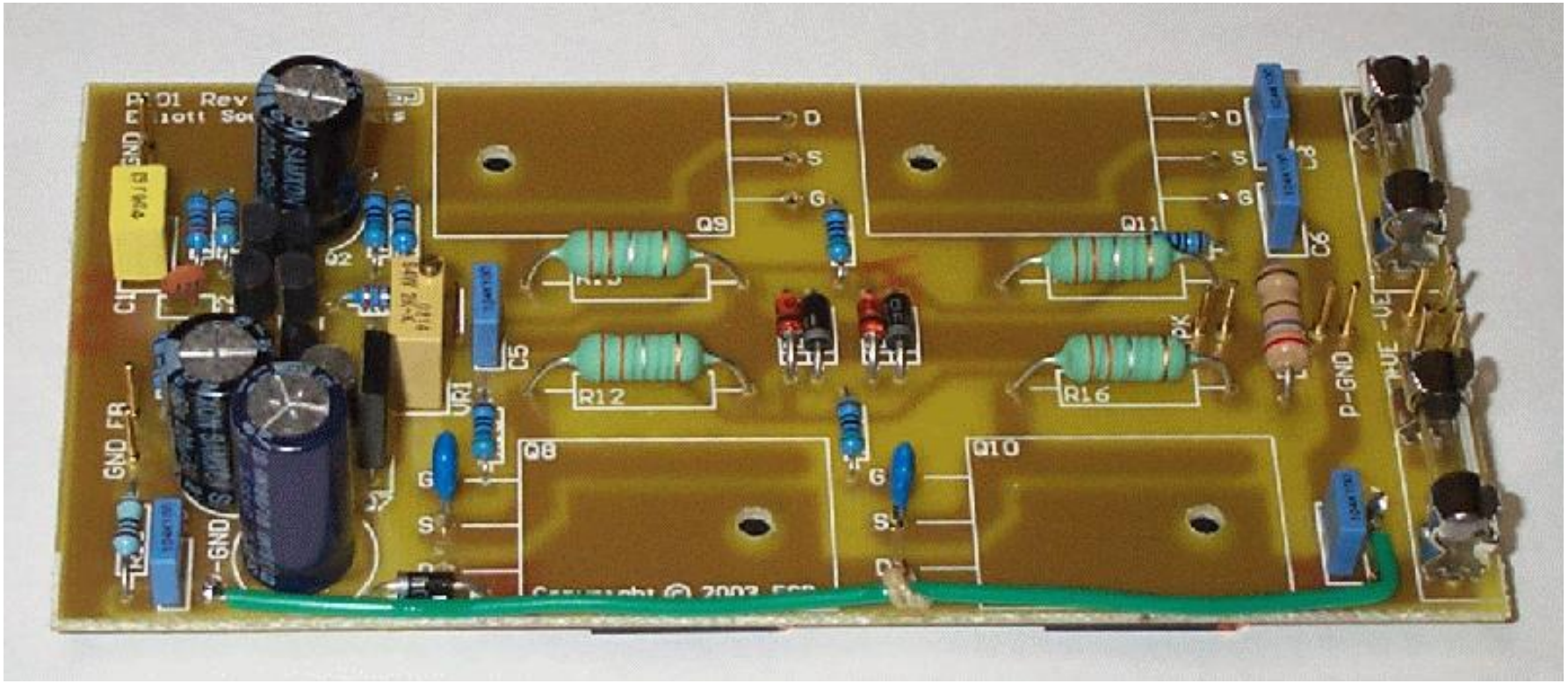
This new amplifier is designed to be as flexible as possible, with no bad habits. Indeed, it will operate stably with supply voltages as low as +/-5V (completely pointless, but interesting), all the way to the maximum supply voltage of +/-70V. The only change that is needed is to trim the MOSFET bias pot!

With the full supply voltage of +/-70V (which must not be exceeded!), RMS power is around 180W into 8 ohms, or 250W into 4ohms.

Short term (or "music") power is typically about 240W into 8 ohms and 380W into 4 ohms. Note that depends to a very great degree on the power supply, and a very robust supply is an absolute requirement for the maximum output.

Since this amp probably has more power than you will normally ever need, even if you do skimp a little on the transformer, the loss will be very small.

It is worth noting that a MOSFET amp will always produce less power than a bipolar transistor version using the same supply voltage. Even using an auxiliary supply will make only a small difference (one reason I elected not to add the extra complexity). A bipolar design using a ±70V supply can be expected to produce something in the order of 270W into 8 ohms, and well over 500W into 4 ohms. The specified MOSFETs have a rated Vds (saturated voltage, Drain to Source) of 12V at full current, and that is simply subtracted from the DC value of the supply voltage. Using the same ±70V supply with a MOSFET amp will give less power than quoted above (see below for measured figures).



Photograph of Completed Amplifier Board

The photo shows the simplicity of the PCB. The MOSFETs are mounted below the board, and are bolted down in the same way as with the P3A and P68 boards. No other mounting is needed. PCB pins are used as anchor points for the power ground link (the green wire along the front edge), so that the main current carrying tracks were not compromised by running a separate track (which would have required a reduction in size of the positive supply rail).

The entire front-end section is between the electrolytic caps, and is deliberately as compact as possible. This improves performance, by ensuring that there are no long tracks for the input stage, which will pick up noise and can seriously degrade the sound of the amplifier.

Performance Figures

The performance of this amp is such that many measurements are very difficult. Some of the more basic measurements are as shown below, based on my custom made transformers which provide ±65V unloaded ...

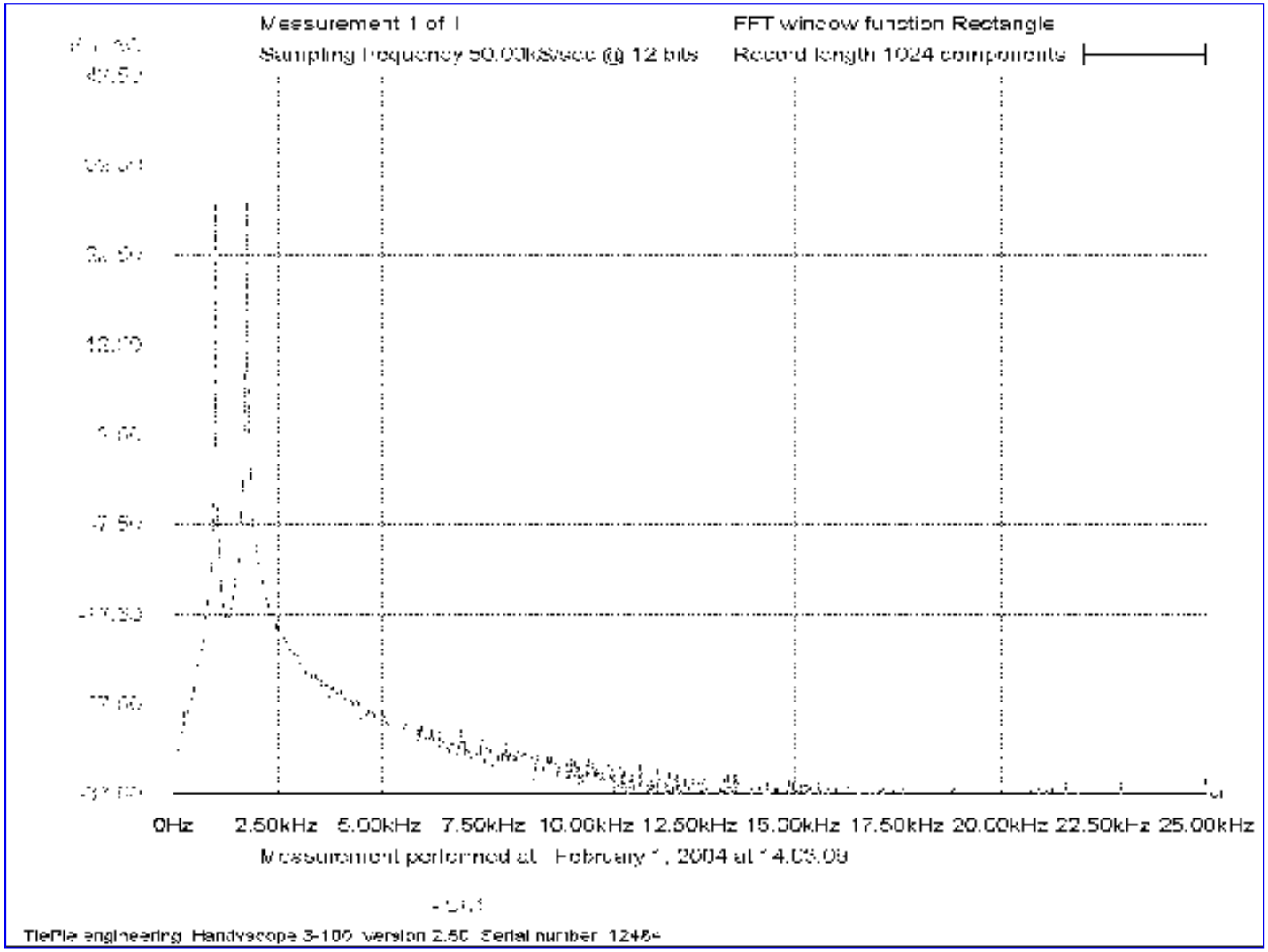
Parameter	Measurement	Conditions
Output Power	> 180W	< 1% THD, 8Ω
	> 275W	< 1% THD, 4Ω
DC Offset	< 20mV	Typical
Noise	< 2mV RMS	Unweighted (-54dBV)
THD	0.015%	No load, 30V RMS output, 1kHz
	0.017%	8 Ohms, 30V RMS output, 1kHz
	0.02%	4 Ohms, 30V RMS output, 1kHz
Output Impedance	< 10 mΩ	1kHz, 4Ω load
	< 25 mΩ	10kHz, 4Ω load
Frequency Response	10Hz to 50kHz	At 1W, -1.5dB

Basic Performance Figures

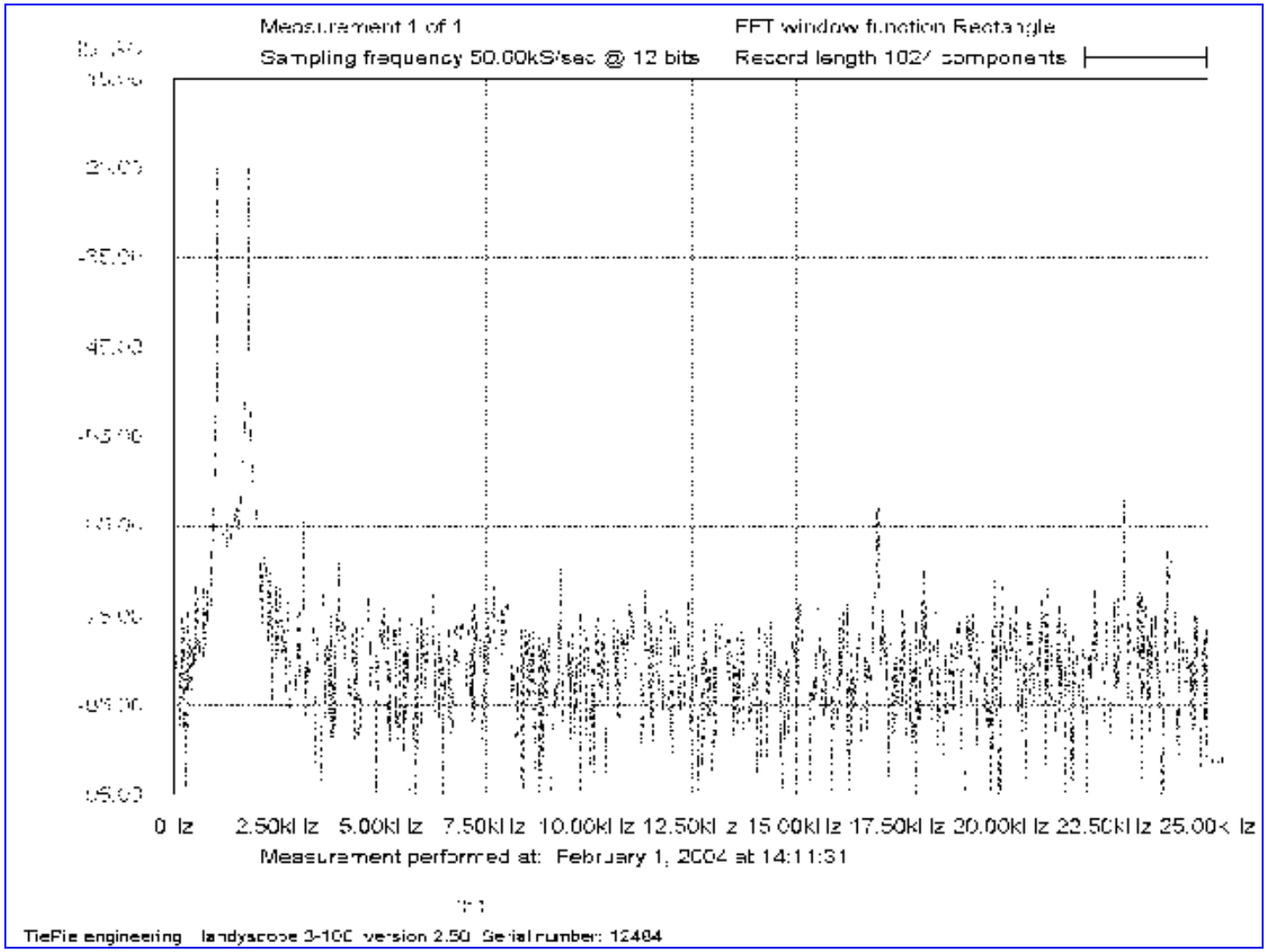
In particular, the distortion figures show that amp loading causes only very small variations, with any harmonics being predominantly from my audio oscillator. There are no visible or audible high order components to the distortion waveform. Output impedance was measured on a fully built amplifier, including the internal wiring. This entails around 200mm of wire in all (per channel), so the output impedance of the amplifier itself is obviously lower than quoted. For an 8 ohm load, the damping factor at 1kHz is around 800 (8 / 10 milliohms) - completely pointless of course, since any speaker lead will ruin that very quickly.

Noise was measured with inputs open-circuited, and at -54dBV may not look too wonderful, however this figure is very pessimistic. Remember that this is the unweighted measurement, with bandwidth extending to well in excess of 100kHz. Even so, signal to noise ratio (referred to full power) is 86dB unweighted, and the amp is completely silent into typical speakers. Indeed, even connecting a pair of headphones directly to the amp outputs revealed that no noise was audible. Naturally, your methods of construction will differ from mine, and you may not be able to get the same performance.

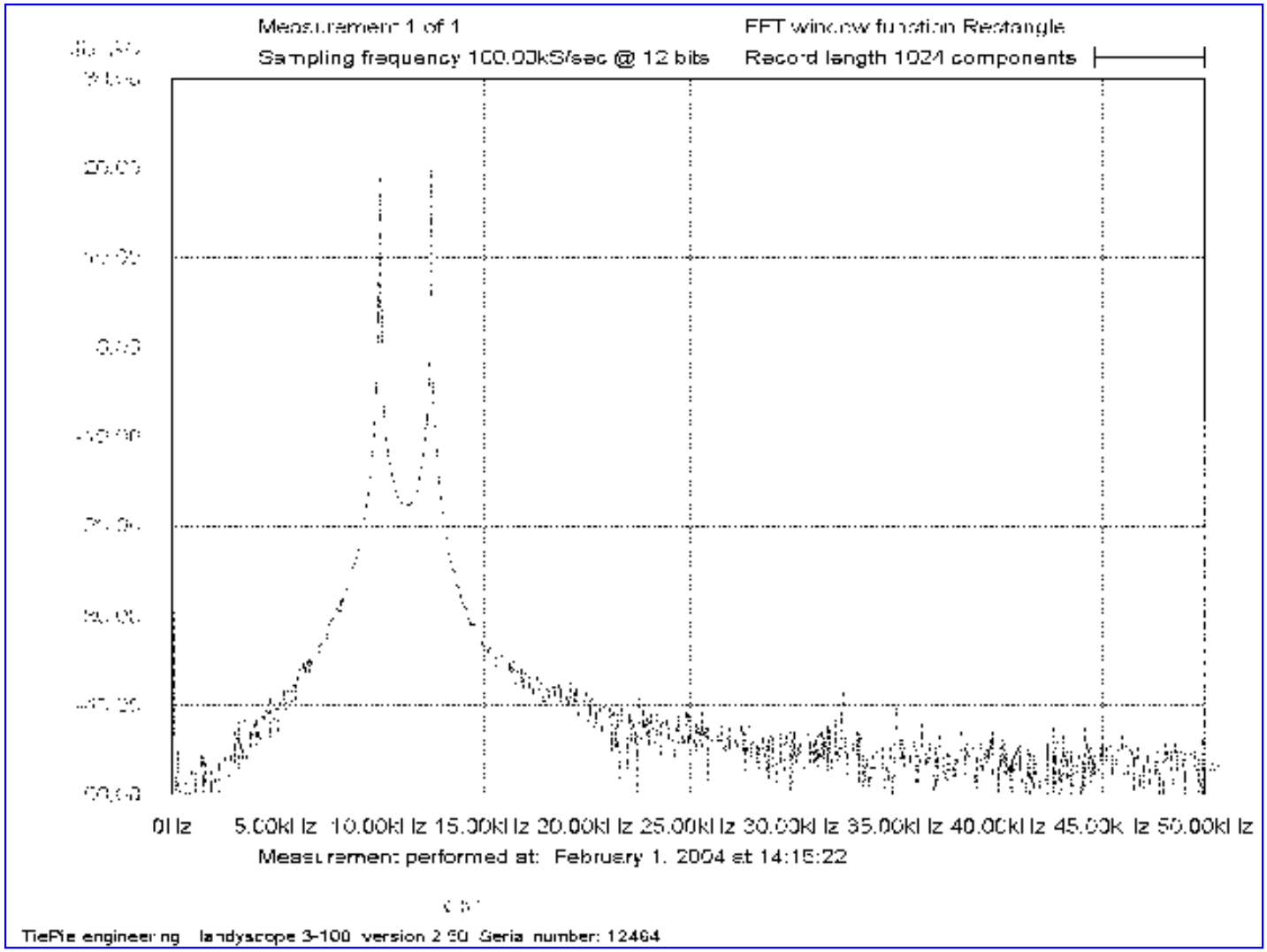
Intermodulation distortion cannot be measured with the equipment I have available, but I have included a screen capture of the three measurements I took. Most of the harmonic content visible (not that there is a great deal anyway) is present in the two generators I used, and the amplifier contributes virtually nothing.



1kHz + 2kHz at +30dbV Output (8Ω)



1kHz + 2kHz at -25dBV Output (8Ω)



10kHz + 12kHz at +20dBV Output (8Ω)

Click on any of the images above for a full resolution version.

Description

The very first thing you will notice is that I have broken with tradition with this amp, and there are no component values shown. Given the performance of the circuit, and the fact that I have already sold a couple as completed, finished amplifiers, I am not about to give away all my secrets for the design. If you want the component values, you must purchase the PCB. There are *no* exceptions, so don't ask.

The schematic of the amp is shown in Fig. 1, and it is about as simple as a high power MOSFET amplifier can get - it is considerably simpler than most, but lacks nothing in performance. The circuit diagram belies the ability of the amplifier though, so do not be tempted to think that it cannot perform as well as more complex designs - it does, and exceeds the performance of many (if not most) of them. It will be seen that I elected to use a bootstrap current source rather than an active version - there is negligible cost difference, but I was unwilling to make such a radical change after testing the prototype and being so impressed with the results. (If it ain't broke, don't fix it!)

The front end is a conventional long-tailed pair (LTP) using a current mirror load and an active current sink in the "tail". Interestingly, adding the current mirror made no difference to distortion, but reduced the DC offset to less than 25mV. The improvement was such that I elected to retain the mirror.

In tests thus far (both measurement and listening), I have been unable to detect even a hint of what is commonly referred to as the "MOSFET sound". The relatively high levels of low order distortion and suceptibility to crossover (or "notch" distortion that plague most MOSFET designs are completely missing - indeed, even with *zero* bias on the MOSFETs, crossover distortion below 10kHz is barely measurable, let alone audible!

Note Carefully:
The most critical aspect of the design is the PCB layout, and it is very doubtful that if you make your own board, that you will get performance even approaching mine. Power output is essentially unchanged, but distortion and stability are achieved by a compact and carefully designed layout for the front end and driver circuits, which minimises any adverse PCB track coupling that causes much higher distortion levels, and may cause oscillation.

This is not a ploy on my part to get people to purchase my PCBs - that has already been taken care of by leaving out the component values. The simple fact is that unless the PCB layout is done with the utmost care, any amplifier can be made to have far greater distortion levels and reduced stability margins than the published figures suggest.

Low Power Version
As shown in the schematics below (figures 1 and 2), the amplifier can be made in high or low power version, and although there is a bit of vacant PCB real estate in the low power design, it is significantly cheaper to make and will be more than sufficient for most constructors. If this version is built (using only 1 pair of MOSFETs), it is essential to limit the supply voltage to +/-56V so that it can drive both 4 and 8 ohm loads without excess dissipation. With this voltage, expect about 100W continuous into 8 ohms, and around 150W into 4 ohms. Naturally, dual MOSFET pairs may be used at this voltage as well, providing much better thermal performance (and therefore cooler operation), far greater peak current capability and slightly higher power. This version may be used at any voltage from +/-25V to +/-56V.

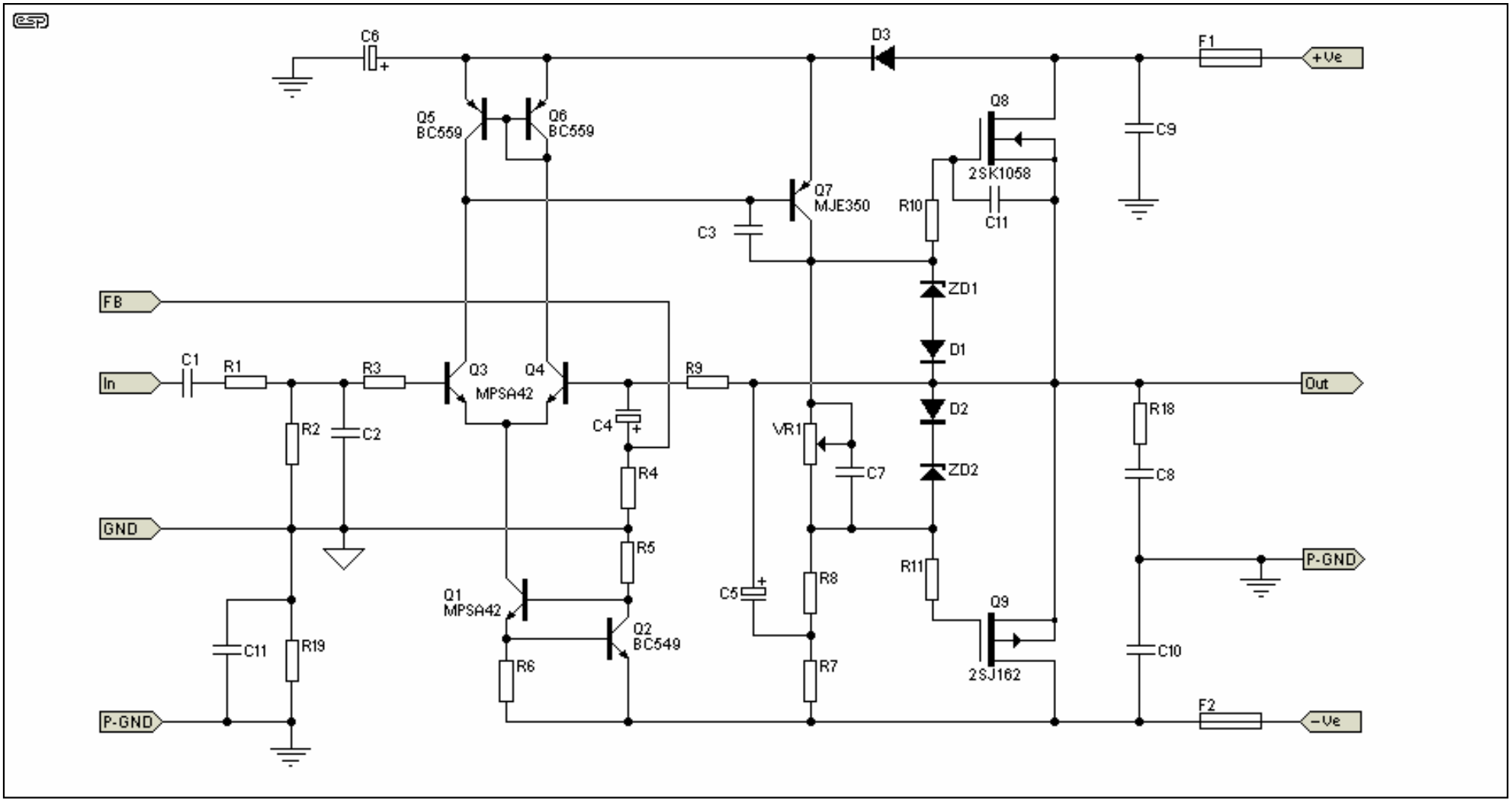


Figure 1 - Standard (Low Power) Version

The MOSFETs used are Hitachi lateral devices, 2SK1058 (N-Channel) and 2SJ162 (P-Channel). These are designed specifically for audio, and are far more linear than the (currently) more common switching devices that many MOSFET amps use. Unfortunately, they are not especially cheap, but their performance in an audio circuit is so much better than vertical MOSFETs, HEXFETs, etc., that there is no comparison. Note that using HEXFETs or any other vertical MOSFET type is not an option. They will fail in this circuit, as it was not designed to use them.

An alternative (and possibly marginally better than the 2S series) is the Exicon ECX10N16 and ECX10P16 (available from Profusion PLC in the UK). These have been used in most of the amps I have built, and they work very well. So potential constructors can verify that the semiconductors are available before purchasing a PCB, this information has now been included. All other parts are quite standard.

High Power Version

The same PCB is used, but has an extra pair of MOSFETs. Since the devices are running in parallel, source resistors are used to force current sharing. Although these may be replaced by wire links, I do not recommend this. This version may be operated at a maximum supply voltage of +/-70V, and will give up to 180W RMS into 8 ohms, and 250W into 4 ohms. Short term (peak) power is around 240W into 8 ohms and 380W into 4 ohms. These figures are very much dependent on your power supply regulation, determined by the VA rating of the transformer, size of filter caps, etc.

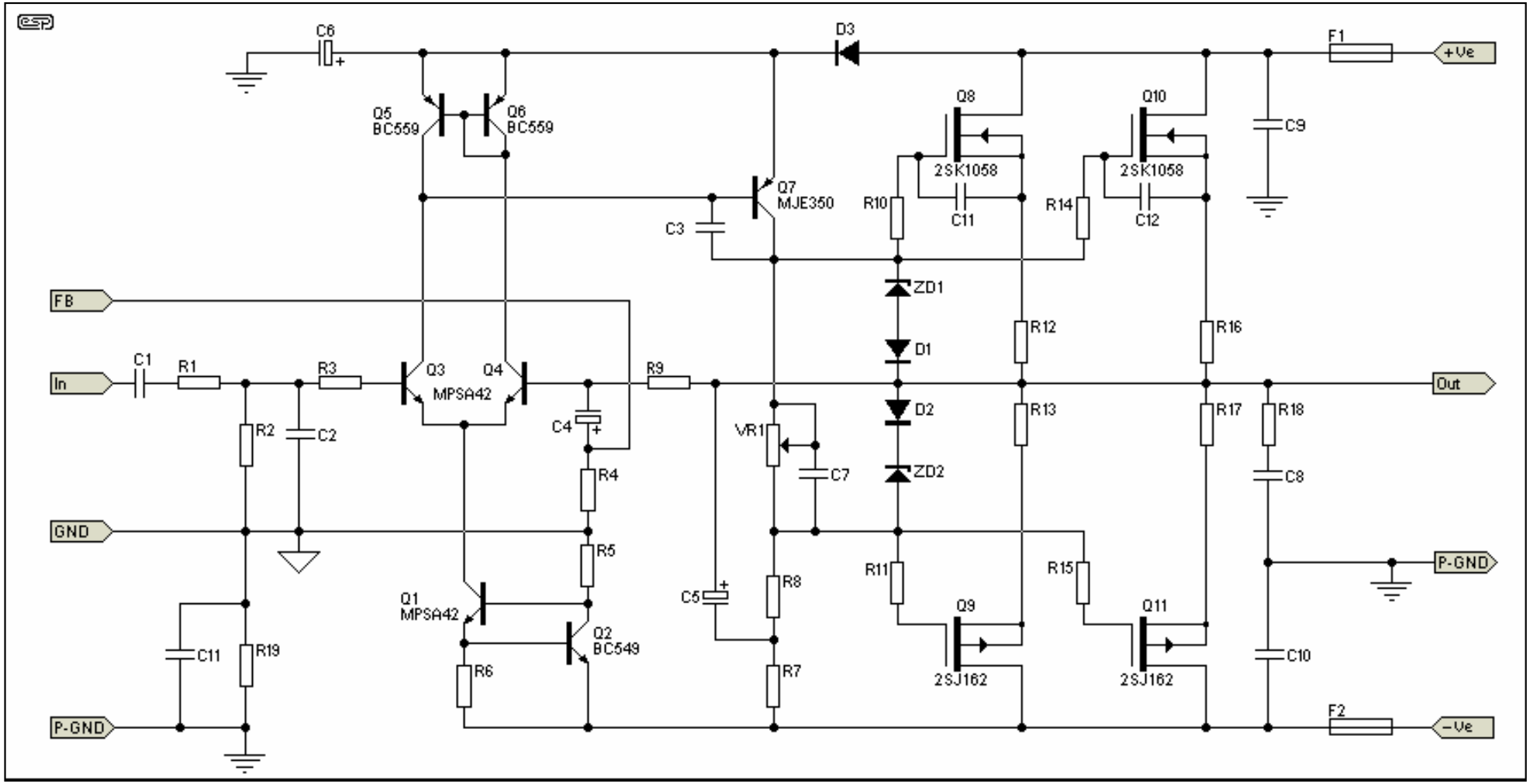


Figure 2 - High Power Version

Although not shown, the transistors and MOSFETs are the same in this version as for the low power variant.

As noted above, the PCB is the same for both versions, but for Fig. 2 it is fully populated with 2 pairs of power MOSFETs. The high

power version may also be used at lower supply voltages, with a slight increase in power, but considerably lower operating temperatures even at maximum output, and potentially greater reliability.

With both versions, the constructors' page gives additional information, and the schematics there include an enhanced zobel network at the output for greater stability even with the most difficult load. This is provided for on the PCB, and allows the amp to remain stable under almost any conditions.

The entire circuit has been optimised for minimum current in the Class-A driver, while still providing sufficient drive to ensure full power capability up to 25kHz. The slew rate is double that required for full power at 20kHz, at 15V/us, and while it is quite easy to increase it further, this amp already outperforms a great many other amps in this respect, and faster operation is neither required nor desirable.

> **Note** - There are actually two caps marked C5, and two marked C6. This is what is on the PCB overlay, and naturally was not found until it was too late. Since these caps cannot be mixed up, it will not cause a problem.

In both versions of the amp, R7 and R8 are selected to provide 5mA current through the voltage amplifier stage. You will need to change the value to use a different supply voltage ...

R7 = R8 = Vs / 10 (k) (Where Vs is one supply voltage only)

For example, to set the correct current for ±42V supplies ...

R7 = R8 = 42 / 10 = 4.2k (use the next *lower*standard value - 3.9k)

Construction

As suggested above, I strongly recommend that you purchase the PCB for this amplifier, or you will almost certainly get results that are nowhere near the amp's real ability. The PCB also makes construction a breeze, with everything except the power supply mounted on the board itself. Like many other ESP power amps, the MOSFETs are mounted underneath the board, requriing only two (or four) screws to attach the PCB and output devices. As always, full construction details will be available in the ESP secure site when you purchase the board(s).

The suggested power supply is completely conventional. Although a small amount of additional power can be obtained by using an auxiliary supply (to boost the rail voltage for the MOSFET drive stage), this is at the expense of greater complexity and more things to go wrong. The transformer for the supply should be matched to the expected power you wish to obtain from the amp. The following table shows the recommended transformer voltage and VA rating for a single channel - either use two transformers or a single unit with twice the VA rating shown for stereo.

AC Volts	DC Volts	VA	Power (8 ohms)
20-0-20	+/-28V	100	40
25-0-25	+/-35V	100	50
30-0-30	+/-42	160	80
40-0-40	+/-56V	200	150
50-0-50	+/-70V	300	240

Note that all powers shown are "short term" or peak - continuous power will always be less as the supply collapses under load. Peak power levels are usually achieved (or approached) with most music because its transients are generally between 6dB and 10dB greater than the average power output. Transformer VA ratings shown are a guide only - larger or smaller units may be used, with a marginal increase or reduction of peak power. Always use at least the size shown for subwoofer use!

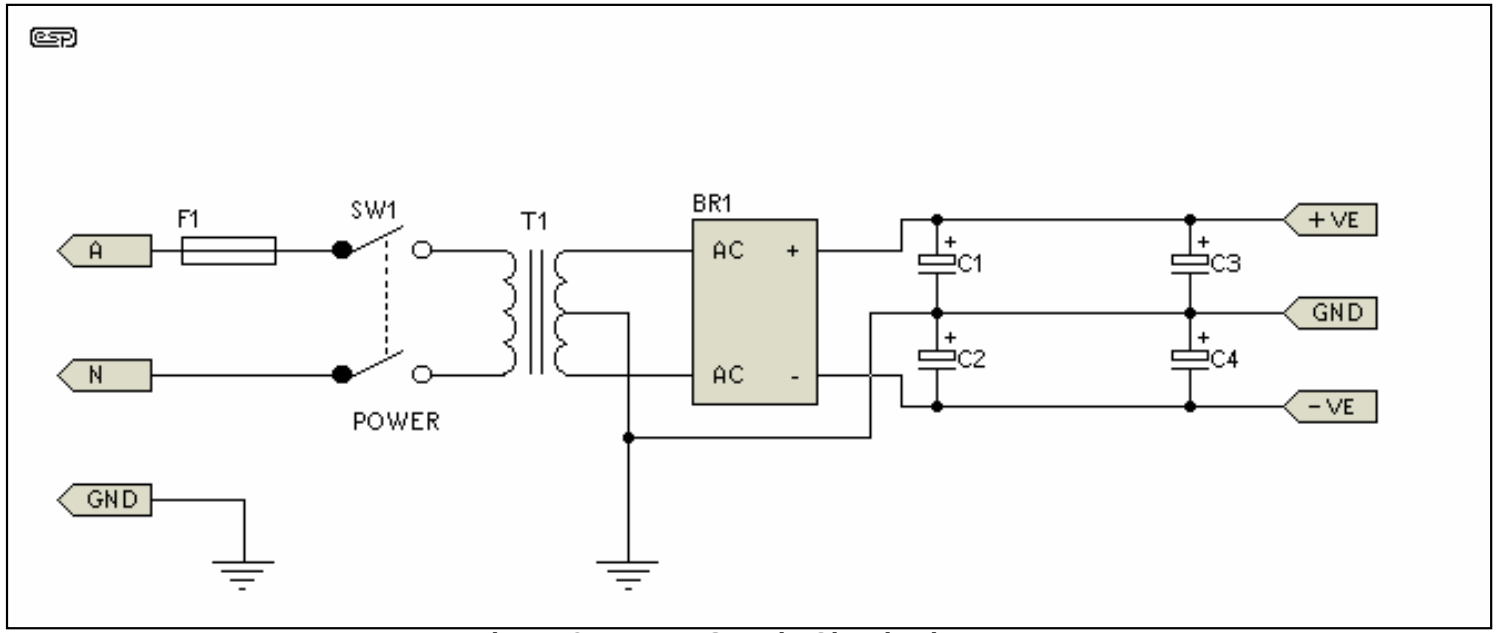


Figure 3 - Power Supply Circuit Diagram

Figure 3 shows the power supply circuit diagram for a ±56V supply, and there is nothing new about it. As I always recommend, the bridge rectifier should be a 400V/35A chassis mount type, and should be properly chassis mounted using heatsink compound.

Filter capacitors *must* be rated to at least the nominal supply voltage, and preferably higher. If possible, use 105°C rated caps, and join the earthed terminals very solidly to form the star earthing point.

Note - The fuse should be selected according to the size of the power transformer. For any toroidal transformer over 300VA, a soft start circuit is highly recommended. Use the transformer manufacturer's suggested fuse - if this information is not available, ask the supplier - *not me!*

The DC supply must be taken from the capacitor terminals - never from the bridge rectifier. Using several small capacitors will give better performance than a single large one, and is usually cheaper as well. For example, the performance of 10 x 1,000uF capacitors is a great deal better (in all respects) than a single 10,000uF cap, at between 50% to 70% of the cost of the large unit. This lunch is not free, but it is heavily discounted ;-)

When you purchase the PCB, you will not only get all component values, but will also have access to information for a power supply that is optimised for the best possible performance for a conventional supply. There is nothing especially innovative about the "advanced" supply schematic, but the overall results will surprise you.

Testing

Connect to a suitable power supply - remember that the supply earth (ground) must be connected! When powering up for the first time, use 10 ohm to 22 ohm "safety" resisors in series with each supply to limit the current if you have made a mistake in the wiring.

For a much more detailed description of the general test processes (as well as troubleshooting information if the amp does not work), please see Troubleshooting & Repair Guide. That article has much more detailed information than I can include in each project page.

Project Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright © Rod Elliott 07 Jan 2004./ Updated 01 Feb 2004 - added measured results

Elliott Sound Products **Project 36**

Death of Zen (DoZ) - A New Class-A Power Amp

Rod Elliott (ESP)
Updated 03 October 2005

Revision-A PCBs are available for this project. Click the image for details.

Project 70 is based on this design, but specifically for headphones. This is where you can really get the benefits of Class-A with none of the drawbacks.

Because the quiescent current can be quite unstable with variations in the supply voltage. Normal changes in the AC mains can cause Iq to shift above and below the preset value. A simple modification can be added that virtually eliminates the problem (or reduces it to the point where it is immaterial). This, plus another optional modification to help stabilise the bias current are included on the new Revision-A circuit board.

Introduction

The Zen - along with Zen improved, son of Zen, Bride of Zen, Second cousin of Zen (or did I imagine that one?) Class-A amp designed by Nelson Pass seems to have become popular. (See references.) I cannot imagine why, since the very concept is flawed in many ways. It has minimal feedback, but that is because it has minimal gain to start with, and appears very simple. Perhaps this is the attraction - but at what price? The capacitors needed for the power supply are massive to try to get rid of hum, and massive means expensive. The "improved" Zen is a little better, since it uses an inductor (or choke) in the supply - obviously the hum drove someone mad. Inductors are expensive too, and also hard to get, and the capacitance has been doubled in at least one version I have seen - ouch, this is seriously expensive!

Well, actually I can see why it is popular. It satisfies the requirement of many amplifier builders, in that it is simple, stable, and very tolerant of layout and component variations. The sonic characteristics will also appeal to many, due to the valve-like sound (or tube-like, if you prefer). Having looked at the original and many of the "improvements" currently on the web, I did a few tests of my own and frankly, found the amp lacking in the fidelity department. Hi-Fi this most certainly is *not*. But .... does it sound good? Apparently so, based on the number of people using (and praising) the Zen, but the feedback I have had on the DoZ so far (and my tests) is also very positive and encouraging. At the time of writing,

hundreds of DoZ amplifiers have been made, with comparatively few reported problems. The issues that have been encountered have been addressed in the Revision-A circuit boards which are now shipping.

Nelson Pass quotes Einstein as saying "Everything should be as simple as possible, but no simpler". I agree with this entirely, and quickly realised that the Zen is simpler than it should be for its intended purpose.

Therefore, I have done some serious work on "Death of Zen", a new Class-A power amplifier that will blow the Zen and all its kin into the weeds, without busting the budget or sacrificing sound quality. Minimal global feedback and lots of local feedback to ensure a very fast and linear amplifier, using the smallest number of components possible. This is the goal, and the remainder of this section explains why.

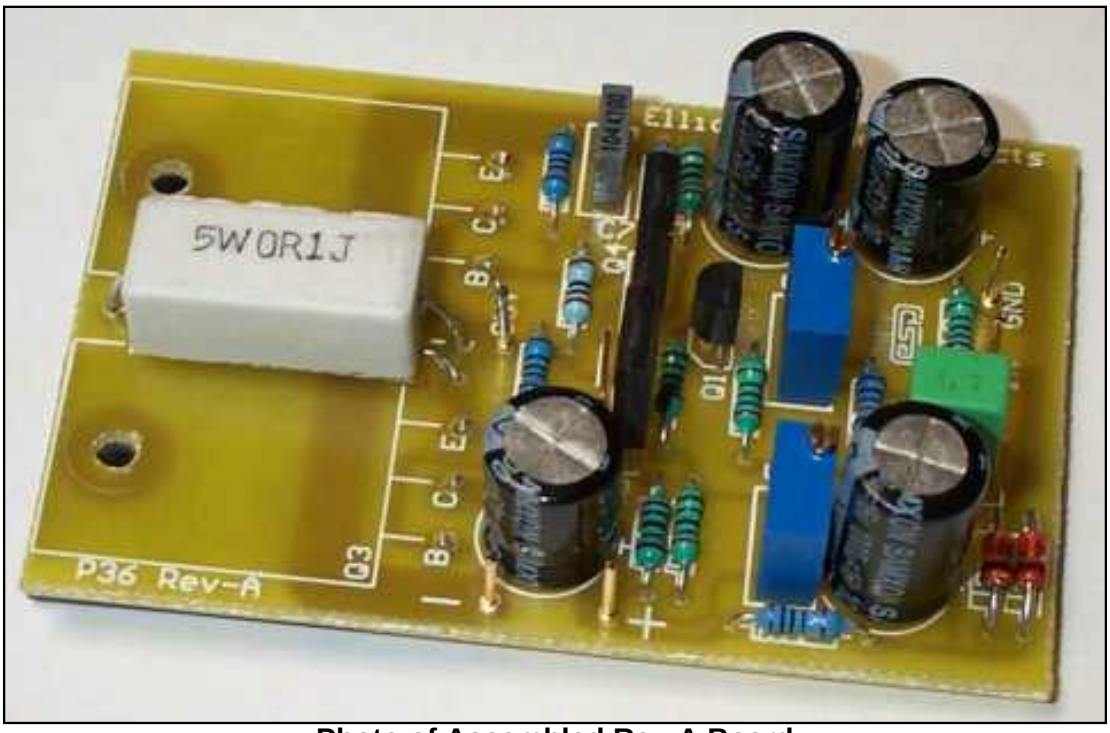


Photo of Assembled Rev-A Board

Lets look at the basic Zen concept, as shown in Figure 1. A power MOSFET is biased using a pot (needed to correct for different device characteristics) so that the voltage at the drain is about 1/2 the supply voltage. Current is limited using a constant current source, and this needs to be set to provide a current that is higher than the maximum peak current to the speaker. Since the amp is not DC coupled, an output capacitor is needed to keep the DC out of the loudspeakers. An input cap is also needed to stop the source (the preamp, or for my tests, an audio oscillator) from stealing the bias voltage.

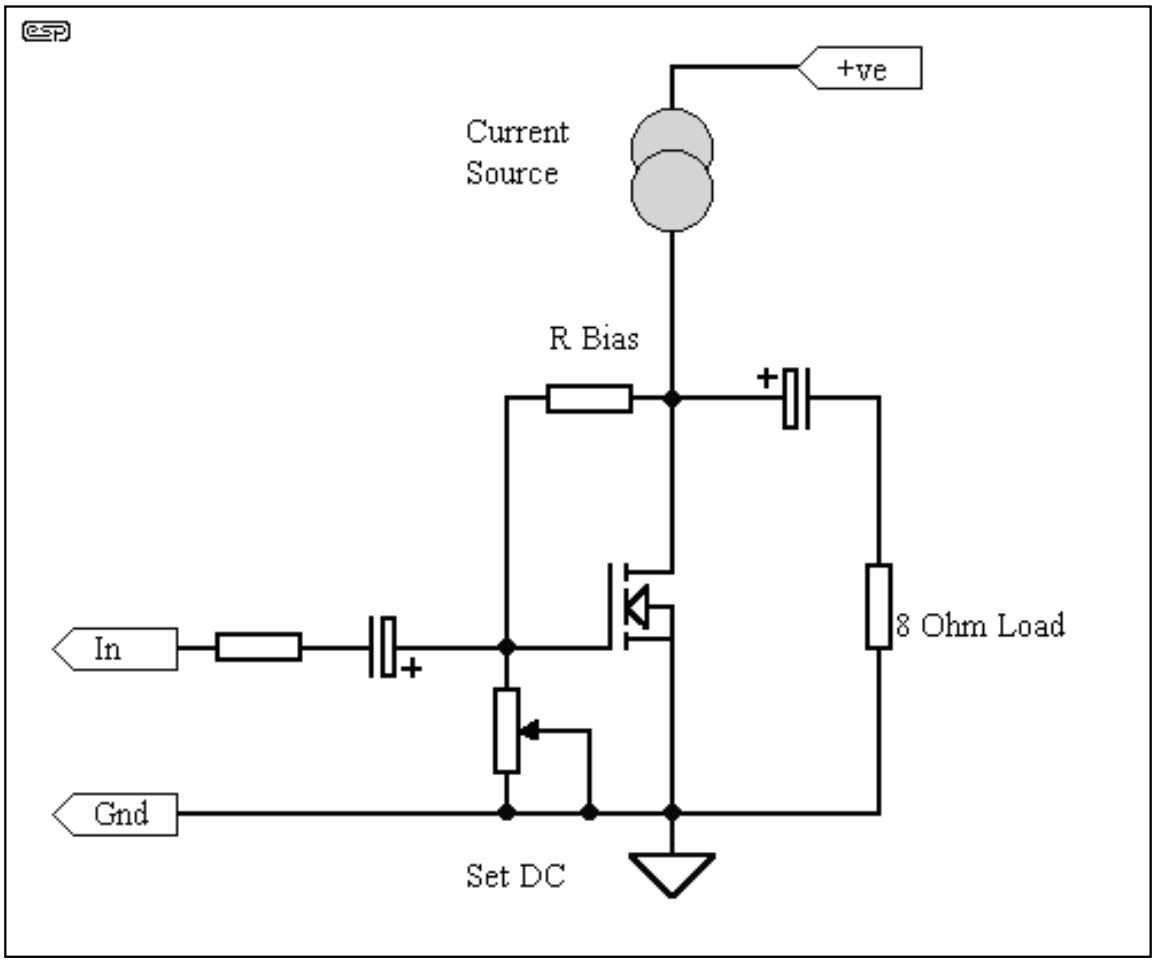


Figure 1 - Basic Zen Concept

Now at first sight the idea looks sound (pardon the pun). We do need to do some basic maths to determine the current needed, but this is easy. Using a 35V supply, the bias point will need to be about 1/2 supply (17.5V), and this means that for ideal devices the peak speaker current is +/-17.5 / 8 = 2.19A (say 2.2A). It is necessary to add a little more current to ensure that the active device current remains high enough to stay within the linear region, so lets say 2.5A

In theory, the +/- 17.5V should allow a peak power of 19W, but this is not possible due to the losses in the devices. As a result, the amp is rated at 10W, and this is reasonable. The output resistance (at DC - this is not the same as impedance) of this output stage is easily determined from Ohm's Law, so R = V / I = 17.5 / 2.5 = 7 Ohms. Although this is the resistance, the impedance will actually be fairly similar (generally lower though [small correction added - oops]). According to some, this is the first fault of the design, since damping factor will be at best 1.14 - this is a little shy of the 100 or more that most audiophiles strive for, but more on this later. With the addition of feedback (and yes, the Zen uses some feedback), the output impedance is quoted as about 1 Ohm.

Most readers of my pages will know by now that I am not a fan of MOSFETs, since they are far less linear than bipolar transistors. To me, this is the first failing, since I fully expected the distortion to be somewhat higher than I would consider acceptable for a ghetto blaster, let alone a hi-fi system.

The Big Test

I proceeded to set up a test, using a suitable MOSFET and comparing it with a transistor in the same circuit configuration. The test setup is shown in Figure 2, and I was able to directly substitute the transistor and MOSFET into the circuit, adjust the bias and run the test. Since I wanted to see distortion components alone, I simply used an 8 Ohm drain / collector load, as this is approximately equivalent to the circuit operating with a current source load and driving a speaker. I kept the operating level lower than normal to ensure that a suitable current reserve remained.

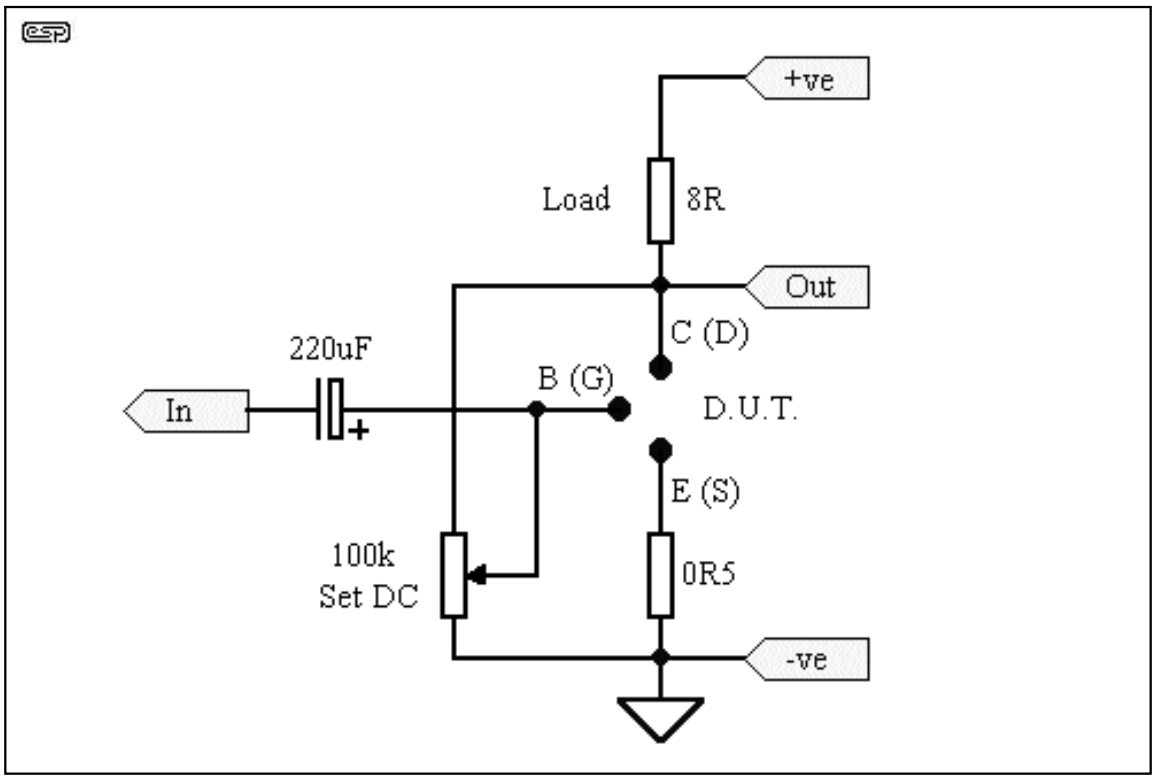


Figure 2 - The Test Setup

In the above, the D.U.T. is the device under test. Emitter, base and collector (or source, gate and drain) are connected as shown. For the power supply, I used my "monster" supply, which is variable only because I use a Variac (variable voltage transformer) to supply the incoming mains. I used a 22,000uF capacitor for added filtering (it was not enough!), and proceeded to take some measurements.

First step was to set the quiescent voltage with the pot, so I had 1/2 the supply voltage at the drain (I tested the MOSFET circuit first). With an applied DC of 30V, this meant a voltage of 15V, so the current was 1.875A or 28W dissipation (both in transistor and load - for a a total of 56W). The hum was higher than I would have liked, but I can make this disappear using the averaging capability on the digital oscilloscope.

Applying a 1kHz sine wave, I could see that the distortion was quite visible at close to clipping, and I was able to operate at a maximum of 6V RMS output before the distortion became too noticeable. I then hooked up my trusty distortion meter to see just how much there was. Remember from the test circuit that I have included a 0.5 Ohm resistor in the source to help linearise the circuit - not to too much avail it seems, since the distortion was measured at 1.58% (after hum removal), and it increased very rapidly if I increased the voltage. Hmmm. This verified my suspicions, but now I needed to test a bipolar transistor in exactly the same test setup to compare the two.

I used a darlington transistor (I dislike these too, but it was convenient and the extra gain is essential with bipolars), and was able to bias the transistor using the same circuit as before. Again, I applied a signal, and was not at all surprised to see that the maximum voltage before distortion was visible on the oscilloscope was considerably greater, and the output generally looked cleaner right up to the point of clipping. I would expect that a discrete complementary pair (the configuration I always use) will be better, but I was rushing to get this into print, so used the most convenient device to hand. At least this means that I can improve on these figures without too much trouble.

To be completely fair, I tested the distortion at 6V RMS again, and measured 1.03% - a worthwhile improvement I thought. Increasing the output to 8V, the distortion climbed to 1.18% - still less than the MOSFET, and with a much improved voltage swing. At this level, the MOSFET was delivering outrageous amounts of distortion, as it started to clip.

The measured distortions are not entirely fair, because of the distortion waveform. With the MOSFET, the distortion waveform was peaky, with quite sharp transitions (indicating high order harmonics). The RMS value is probably too low, and certainly does not indicate accurately the audible effect of the distortion. By comparison, the bipolar transistor had a very smooth and almost perfect 2nd harmonic, with very little evidence of any high order harmonics at all. Asymmetry in the residual distortion waveform showed that there was also 3rd harmonic distortion, but at a lower level - I think I will have to have a listen to the residual signal to determine the "musicality" or otherwise of the distortion I measured.

29 Oct - Further Tests

The following day, I decided to buy a MOSFET rather than use the one I had, and selected a MTP3055 as a budget device designed for audio and switching. I still don't know exactly what the other one is supposed to be for, but it shouldn't matter - bipolar transistors can be selected for linearity, but it isn't that big a deal. Not so with MOSFETs as I discovered - the new one was markedly better than the original, but still fails to touch the bipolar. Incidentally, the darlington bipolar I used was a TIP141, and is designed for switching (lest I be accused of fiddling my results by device selection). I did not retest the 60N06 at the lowest level, but given the other results I could see no point.

Since I now have 3 separate test results, I have tabulated them below.

Output Voltage (RMS)	TIP141 Darlington	MTP3055E MOSFET	60N06 MOSFET
2.5	0.05% (approx)	0.47%	N/T
6.0	1.03%	1.02 %	1.58%
8.0	1.18%	1.80%	> 2%

Table 1 - Device Comparisons In The Test Circuit

It would be useful to carry out these tests with a completely hum free supply so that the distortion is not affected by the supply ripple, but by averaging the measured result with the oscilloscope I believe the results are accurate enough for comparison, especially since the same configuration was used for all tests.

Quite obviously, the bipolar is a winner at low levels (where the distortion is most noticeable), and I am sure that using a nice linear transistor such as the complementary pair, these results would show the superiority of the bipolar transistor even more clearly. Again, as the supply limits are approached and the current through the devices varies the most, the bipolar is again well ahead.

UPDATE: The MTP3055 distortion figures actually are very close to those published by Pass Laboratories for the Zen, so this validates the test circuit I used, and that the figures are not exaggerated in any way. (27 Nov 99)

Full Test Of Death Of Zen (DoZ) Concept

The next step was to test something close to the final configuration, to see what things had an effect (profound or otherwise) on the performance. I still used the TIP141, knowing that I can improve on this greatly as I progress, although as the final circuit shows I eventually chose not to use a compound pair after all. Figure 3 shows the test circuit, still using the 8 Ohm resistor as a load, but I ran these tests using my bench supply to eliminate the hum problems. All tests were performed at an output of 6V RMS (equivalent to the 8V tests above, due to the lower supply voltage).

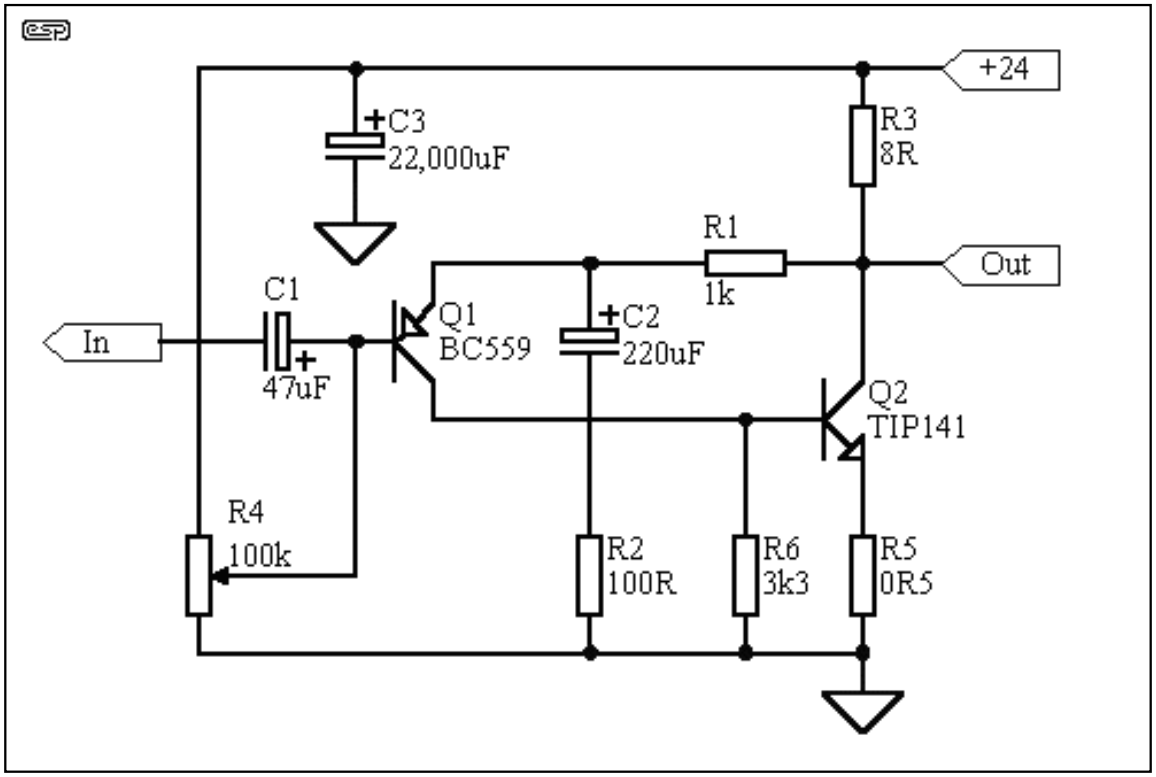
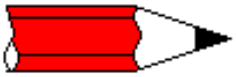


Figure 3 - DoZ Test Circuit

This is the basic configuration I will be using for the final design, although there will be some resistor value changes as I get closer to the final circuit, and the load resistor will be replaced by a constant current source. For those who want to try the circuit with a high output impedance, I will also include the modified feedback network.



STOP ! - Do not build this circuit as a real amplifier. This is a test circuit, designed to verify some basic parameters of the design. The final design is shown in Figure 4.

Some interesting things came to light during testing, especially when I included the resistor (R6) from base to earth on Q2. With no resistor, I measured a distortion of 0.15%, and this was almost completely 2nd harmonic. There was a very noticeable degradation of the positive going slope on a 10kHz square wave, and a fairly low slew rate resulted. Adding the resistor improved this dramatically, and reduced the distortion to 0.05% - but it was now almost completely 3rd harmonic.

This will create a conundrum for some - would you rather have very low levels of 3rd harmonic distortion, or considerably larger amounts of 2nd harmonics (bearing in mind that the 3rd harmonics are still there). I cannot see any good reason to tolerate any more distortion than is absolutely necessary, so considering the much better slew rate (and therefore high frequency performance), I will be including this in the final design. You might want to leave it out if you want the 2nd harmonics, but I don't think the end result will be very satisfactory.

This is due to the transistor's turn-on and turn-off characteristics becoming more symmetrical by providing a base discharge path, but I did not expect such a large difference. The frequency response extends to over 100kHz at full power (6V RMS for these tests), and square wave response shows that the amp is both fast and stable - and this with a very ordinary switching darlington. I saw no evidence of measurable distortion above the 3rd - there must be some, but I have no way of measuring it. The 3rd harmonic appears to be an almost perfect sine wave, with some very small variations.

Slew rate is better than 6V/us (positive going) and over 20V/us negative going - not as good as some, but I blame the TIP141 for this. I have checked the specs on it, and it is a fairly slow device (like most darlingtons) as confirmed by these tests.

None of this testing has been done with a circuit board. In all cases I simply bolted the device to a heatsink, and attached the other components as required. Power connections were all made using alligator clip leads. Since I have used exactly the same "rats-nest" wiring for all testing (including these last tests), and I have not been able to induce additional (or reduced) distortion by moving leads about, the amp looks as if it will be fairly tolerant of assembly methods (all known assembly methods will be superior to what I have done so far).

More Testing

So, there I was on Saturday (27 Nov), thinking suddenly - "I wonder how the amp would work with a MOSFET instead of the transistor?" Off to the workshop and I tried it. The answer is ... horrible. Apart from the greatly reduced voltage swing because of the topology, the distortion at 1V (125mW into 8 Ohms) was about 0.4%, and I was mightily disappointed.

You see, I really would like to use a MOSFET in an amp that I would like listening to, because the idea appeals to me. They are fast, need minimal drive current, and the overall concept is wonderful. But they still don't sound any good - what a shame.

Thus chastened, I thought I would just have a quick muck about with the current source (having re-installed the bipolar transistor) and listen to it on a speaker. Even with the slow TIP141 transistors (I used one for the current source too), the rats-nest wiring I was using instantly caused me problems (clip leads all over the place, and components supported by sky-hooks and each other).

The issues were easily sorted out (re-arrange the clip leads :-), and I ran some distortion tests with and without a load. Distortion was almost exactly the same as previous testing, and I was finally able to measure the actual output impedance, which measured at 0.22 Ohms. Not too shabby for such a simple amp.

As for the sound, well I must admit that it sounded just like an amplifier. I couldn't hear any nastiness, and even with the limited power it was quite loud enough. The main problem I have now is that I don't have any conventional hi-fi speakers (mine are biamped), so was forced to listen on one of my lab speakers. Even there I am limited, since the main lab system is tri-amped. SInce the unit is hardly portable in its test condition, I will have to rely mainly on the distortion meter and my oscilloscope until I have a complete unit built.

One thing I have realised in all of this testing, is that a Class-A amp is a far more irksome thing to design and test than a Class-AB amp. This is partly because of the high current that is always present (and the fact that massive heatsinks are needed just to do simple testing), and partly because in this design I am trying to achieve maximum performance from minimum components - this has turned out to be more difficult than expected, and some of the simplest changes can make a great difference to the performance.

I am now leaning more and more towards the concept suggested by John Linsley Hood [3], where the bias current is modulated to provide Class-A but with less quiescent dissipation. You would be surprised how hot a pair of 1 ° C/W heatsinks get with a quiescent current of 2.5A and a supply of 40V. Mind you, I did manage to get 18W into 8 Ohms at the onset of clipping, but with the sinks just lying on the bench I had to be careful that the whole amp didn't destroy itself. For further tests, I had to build it properly - what a pain!

---

12 December Tests

So, pretty much having made up my mind on the topology, I set about building the amp. I must say that the final result lived up to nearly all expectations, and works extremely well. One word of warning - I used TO-3 case transistors, and I strongly suggest that you do the same. No plastic case device has good enough case to heatsink thermal resistance, and with a final dissipation of 28W per device, even the TO-3s get hot.

For ease of working (and so it would stabilise quickly) I only used a small heatsink, and ran a 12V fan at 6V to keep it cool. At 1/2 voltage, the fan was very quiet, so you might want to consider this as a possibility for Class-A amps in general.

The final circuit (at least for the time being) for the DoZ is shown in Figure 4, and it can be seen that it is a bit different from the last attempt. This being the full and proper circuit, it is fully functional, and I have tabulated test results below.

After all the experiments I carried out before, it turns out that the current source is more critical than I would have hoped. Although this is a simple circuit, it is supposed to supply a constant current at any frequency, and this is harder than you might imagine. R7 and R8 were added in an attempt to speed up the current source, and were only partially successful. As it stands, the amp will provide full power up to about 16kHz, which is actually more than enough for any application.

For final testing you will need two multimeters, one to measure current and the other for voltage. If you only have one, use a 1 Ohm resistor in series with the power supply positive lead. When you measure 1 Volt, this means that the amp is drawing 1 Amp. The resistor can remain in circuit, providing a useful reduction in supply ripple. You will lose 1.7V at operating current, and a 5W resistor is sufficient - it will get hot though.

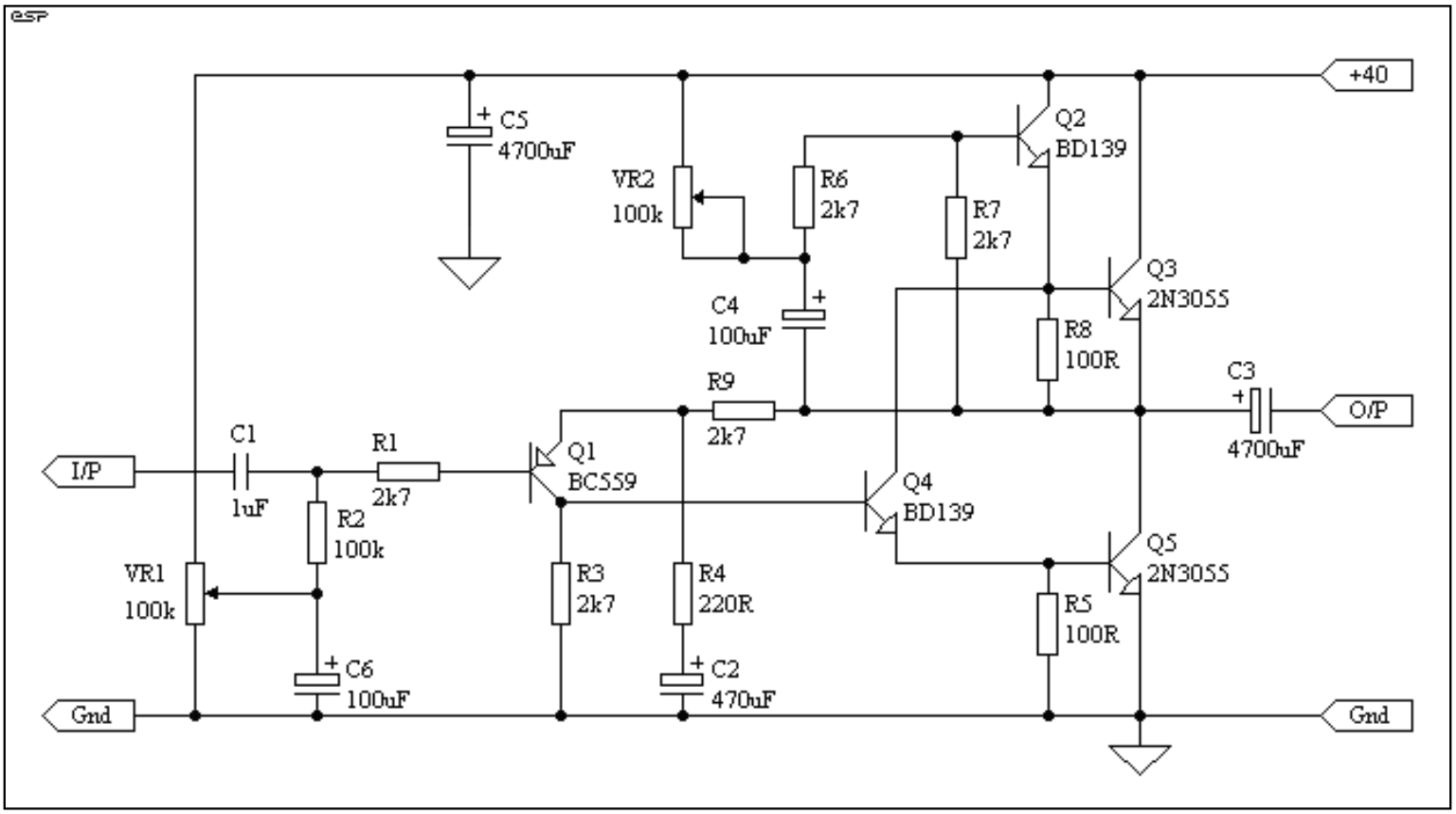
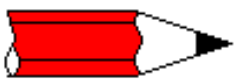


Figure 4 - Semi-Final DoZ Circuit



Q3 and Q5 (the output transistors) must be on a substantial heatsink (see below), and Q2 and Q4 also need heatsinks. These do not need to be especially large - TO-220 U-shape heatsinks will be fine (or make suitable sinks with scrap aluminium). The drivers get excessively hot with no heatsink.

A quick circuit description is in order. VR1 is used to set the DC voltage at the +ve of C3 to 1/2 the supply voltage (20V for a 40V supply), by setting the voltage at the base of Q1. The 100uF cap ensures that no supply ripple gets into the input. Q1 is the main amplifying device, and also sets the gain bt the ratio of R9 and R4. As shown, gain is 13, or 22dB, providing an input sensitivity of about 1V for full output.

Q4 is the buffer for the output transistor Q5, and modulates the current in Q2 and Q3. VR2 is used to set quiescent current, which I found needs to be about 1.7A for best overall performance. C4 and R6 are part of a bootstrap circuit, which ensures that the voltage across R6 remains constant. If the voltage is constant, then so is the current, and this part of the circuit ensures linearity as the output approaches the +ve supply.

After some more testing, I found that the optimum quiescent current was 0.75 times the peak speaker current. At lower currents, third harmonic distortion predominates, while at higher current distortion seems to remain stable (but device dissipation is increased).

You will notice that there is no zobel network on the output, and the amp is unconditionally stable without it.

Before applying power, set VR1 to the middle of its travel, and VR2 to maximum resistance (minimum current). Be very careful - if you accidentally set VR2 to minimum resistance the amp will probably self destruct - more or less immediately.

With an ammeter (or 1 Ohm resistor) in series with the power supply, apply power, and carefully adjust VR2 until you have about 1A. Set VR1 to get 20V at the +ve of C3, and re-check the current. As the amp warms up, the current will increase, and you need to monitor it until the heatsinks have reached a stable temperature. If necessary, re-adjust VR2 and VR1 once the amp has stabilised. If you use a heatsink of more than 0.5°C/W the amp will overheat and will be thermally unstable - *this is not desirable* (*note use of extreme understatement* :-

I used a 40V supply, and was able to obtain 20W at the onset of clipping. Clipping is a lot smoother than most solid state amps, and the amp has no bad habits as it clips. Using a 1uF capacitor directly across the output caused no problems, other than some mild overshoot with a square wave input.

---

14 Oct 2000 - Update

As the supply voltage changes with normal variations in AC mains voltage, the quiescent current also shifts. This is not desirable, and is easily solved with the addition of a resistor and a zener diode (or a series string for odd voltages). If you are using a regulated supply, this mod is not needed.

The process is very simple. First, measure the actual nominal supply voltage - the amp(s) must be connected. Subtract 5 to 7 volts from the measured voltage, to obtain a value that can be matched by standard zener diodes. For example, your supply voltage might be 38V, so a zener voltage between 31 and 33 volts is needed. Since 33V is a standard voltage, that will be fine.

The complete updated circuit is shown in Figure 5, and also shows the actual circuit used on the PCB (excluding the modification described here). The voltage for the quiescent current setting and output voltage is now reasonably well fixed, so mains voltage variations will have very little effect on the overall current of the output stage. Minor variations are also prevented from causing slow voltage shifts at the output. These were never audible, as the circuit is deliberately very slow, but eliminating them cannot be a bad thing. The modification requires that one track be cut, and a resistor and zener (or

zener string) attached to the underside of the board. Despite the sound of this, it is completely painless :-)

D1 and D2 are zener diodes - you may only need one of them, depending on your supply voltage. The selection process is described fully below. Within the useful range of zeners, the following values are standard and suitable for the purpose ...

10V 12V 15V 16V 18V 20V 24V 30V 33V

Values higher than 33V are uncommon in retail electronics outlets, and anything lower than 10V is not recommended in a series string for this application.

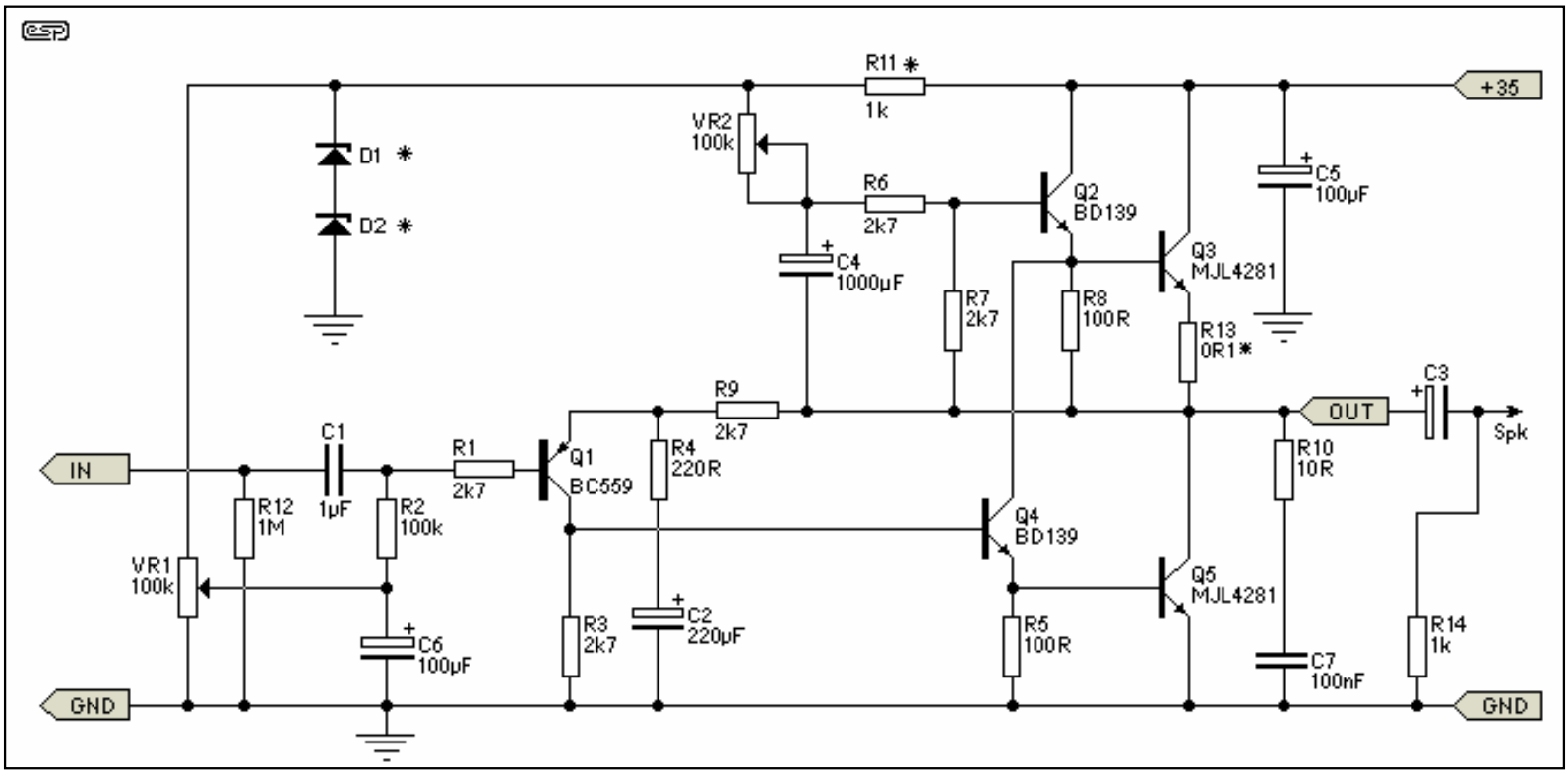


Figure 5 - The Complete DoZ Schematic (With Iq Stabilisation)

* These components are optional.

There are a few other small changes to the circuit, but these are simply to reflect the PCB design and are of no real consequence. A zobel network has been included - not because the amp needs it, but just in case a reactive load that may cause instability is connected. C5 has been reduced in value so it will fit on the board, and C3 (still very much needed !) is mounted off the board as it is too large for PCB mounting (it would almost double the board size).

A quick calculation example for the zener rating and resistance are in order, so it is properly understood. The maximum zener current for a given voltage is easily calculated ...

Iz = Pz / Vz where Iz is zener current, Pz is the power rating, and Vz is the zener voltage

Small zeners are typically rated at 400mW and 1W. A 33V 400mW zener therefore has a maximum current of 12mA. Allowing for a resistor voltage drop of 5 to 7 volts means that the zener current will be 5 to 7 mA (1V across 1k gives 1mA). Since it is recommended that zeners be operated at between 5% and 80% of the maximum rated current, this fits very nicely into our requirements.

Should you find that you need a zener voltage of (say) 36V, then two 18V zeners would be the most appropriate. Try to ensure that the zener current is at least 5% of the maximum, or the regulation will not be as good as it should be. This means that you will need to use a pair of higher voltage zeners in preference to one high voltage and one low voltage.

For example, 36V (near enough) can also be created with a 33V and a 3.3V zener in series. The 3.3V zener will have a maximum current of 120mA (assuming 400mW), and you will barely achieve the 5% minimum current needed (6mA in this case).

The modification described here does not change the measured performance of the amp, and creates no audible differences whatsoever. It is designed to stabilise the quiescent voltage and current, and it does that quite well. Some small variations will still be measured, but are so reduced in magnitude as to be considered negligible for all practical purposes.

Test Results

On the basis of the tests, I would rate this amp at 15W, although I did get more. Distortion rises with increasing level, and starts to get a bit high above 15W - at low power (such as a couple of watts) the distortion was about the same as the residual of my oscillator, which means that it must be below 0.04%, but I have no idea just how low it gets. All distortion components are predominantly second harmonic at all tested levels.

I simply used components as I found them, and did no matching or any selection. All test results are based on the prototype, which uses ordinary resistors, a couple of old salvaged computer caps for the high values, and standard electrolytics for the others. The input capacitor is an MKT polyester type.

Supply Voltage	40V
Quiescent Current	1.7A
Maximum power 8 Ohms	20W (15W)
Output Noise (unweighted)	<1 mV
Distortion @ 1kHz, 15W	< 0.2%
Output Impedance	0.378 Ohm
Frequency Response (-0.5dB @ 1W)	<10Hz to >50kHz

Table 2 - Measured Performance Of Figure 4

I could hear no noise at all, even with power supply ripple of 300mV peak-to-peak. The noise level I

measured was about 0.5mV, but it is not easy to measure accurately at such low levels. There appeared to be no residual hum that I could see on the oscilloscope, even with averaging turned on.

From this it appears that the amp is quite tolerant of supply ripple, and a simple supply will probably be fine. I tested with my normal "monster" supply, which has a fair bit of ripple, and still could not measure any supply hum at the output. A suitable power supply would be the capacitance multiplier circuit (Project 15), or a simpler one can be used. I shall work out the details of a suitable power supply and describe it within the next couple of weeks.

The amp is extremely tolerant of voltage, so for less power, use a lower voltage supply - no other changes are necessary. I found that the amp worked perfectly with supplies down to 15V, but less than 20 is unlikely to be useful (even this will only give about 4.5 Watts into 8 Ohms).

Do not try to increase the power by using a higher voltage, as the dissipation in the transistors will exceed their ratings. If you need more than 15W, I strongly suggest that you use a circuit such as the 60W amp (Project 3A), which actually has lower distortion than this design.

The amp will also tolerate a short circuit with no ill effects (I wouldn't keep it up for too long though), and even (blush) reverse polarity. I accidentally connected the supply up backwards while testing, and thought "Oh, no. Now I'll have to rebuild the blessed thing" (if the truth be known I thought something much shorter!). However, I connected the supply the right way 'round, and away it went, as if nothing had ever happened. This is *not* an experiment I suggest to others.

I found that the design is also unaffected by quite a few component variations. When I first started testing there were no emitter-base resistors in the current source, and when I added them, I simply readjusted the two pots to get everything back where it was. I retested distortion after making the changes, and could measure no difference.

Final Test Results + More Info

Before I finish with this project, I have tried some faster transistors (see below). The ones I used initially are the absolute base model 2N3055/1, so I tested with something faster (but with less power - I used plastic 130W devices). The 2N3055 comes in a whole bunch of different flavours - fast, slow, 60V, 100V etc. The ones I have are slow (FT is only 800kHz) 60V devices (I salvaged them from a bunch of old power supplies).

I do not suggest that you use plastic case transistors unless the supply voltage is kept below 35V, or if you must use plastic cased devices (such as TIP3055s) use two transistors in parallel. The latter method will make a nice simple amp quite complex. Otherwise, you can use MJL1381 transistors. These are very nice devices, and have a sufficiently high power rating so that parallel operation is unnecessary.

As it stands, this amp is so simple that I wired mine with everything just soldered directly to the output transistors (which I put in sockets), and a couple of insulated stand-offs. I might even put in a photo of it - it looks rough (only because it is), but who cares.

Also, something for the experimenter. I will be providing the info on how to increase output impedance. The calculations are too irksome to work through, so I suggest that you do it experimentally, and this allows you to tweak the impedance while you are listening. Be careful though, because it also changes the gain, so don't confuse a change in volume as a change in sound quality (our ears will do this very easily, and are easy to trick in this way).

16 Dec 99: I tried transistors with an fT of 2.5MHz, and it made only a slight difference to the speed of the amp (slight, as in not worth the effort). Distortion remained about the same, and the amp remained stable.

I also tested the amp into 4 Ohms while still set up for 8 Ohms, and it managed 10W at a distortion of about 0.4% (far too high for my liking). I had to reduce the voltage and increase current so it clipped symmetrically. I have prepared a table of voltage, current, impedance and expected power below for those who want to experiment further.

Z (ohms)	Volts	Iq (Amps)	Diss. (W)	Power
4	40	3.40	136.00	32.41
8	40	1.70	68.00	19.81
16	40	0.85	34.00	10.87
4	35	3.00	105.00	24.51
8	35	1.50	52.50	15.02
16	35	0.74	25.90	8.26
4	30	2.50	75.00	18.01
8	30	1.30	39.00	10.89
16	30	0.63	18.90	6.01

Table 3 - Operation At Different Voltage And Impedance

The table is not meant to be too accurate, but as a guide only (the figures were calculated in a spreadsheet, using a fairly basic empirical formula). Note that with a dissipation of 136W, 4 Ohm operation from a 40V supply is not recommended. Trying to get a continuous 68W of heat each out of transistors and into a heatsink is extremely difficult, even with TO-3 transistors. Even 34W each requires meticulous attention to detail with the transistor mounting if you want to keep the thermal resistance (and hence temperature) down.

If you can't keep your fingers on transistors, then they are hotter than I like to operate them - I know they will take much more, but it shortens their life.

In reality, the amp can be operated into any impedance you like, while still set up for 8 Ohm operation. Just remember that for lower impedances, the output will not clip symmetrically, and output power is reduced. At higher impedance, the power is reduced, but so is distortion.

One problem with such a simple design is that the quiescent current is supply voltage dependent. Mostly this will not cause major problems, particularly if a generous heatsink has been used (any why would you use anything else?). It does require that the current is set carefully, and should be monitored to make sure it cannot reach destructive levels.

Heatsink

As I have said before, this amp needs a really, *really* good heatsink, as do all Class-A amplifiers. If you have a few metalworking skills and tools, have a look the site below - its a neat idea, and you can save a lot of hard earned by making a monster heatsink yourself. Have a look at my article on heatsinks to see what sort of radiating surface is best. A thin coat of flat black enamel paint seems to be the most effective in my experience (other than black anodising, which is the overall winner).

There used to be a site that showed you how to make your own heatsinks, but sadly, it has departed this Web. Basically, these were the complete plans for a chassis and heatsink, so you can make it exactly the size you want. When I get an opportunity, I will generate the files to show you how it can be done, and post that as a separate project.

Conclusions

From everything that has gone on during development, I feel that my anti-MOSFET stance is justified for a Class-A design, at least, and that the concept of such an ultra simple amplifier is flawed if low distortion is the ultimate goal. Certainly, some degree of low order harmonic distortion is not unpleasant, and may even add some degree of musicality to an otherwise "clinical" sound, but the cost (i.e. loss of definition of complex passages, etc) is too high for my liking.

Even the bipolar transistor version is unacceptable in the most basic configuration, since although the distortion is lower, it is still too high - this does not qualify as hi-fi by any definition. The addition of global negative feedback (as shown in Figure 3) is the ONLY way to reduce the distortion to within acceptable limits.

In the introduction, I said that I would discuss the damping factor issue further, and so I shall. In my article on impedance (see references), I went into some detail about one of my amps that I modified to have a very high (about 200 Ohms) output impedance. The overwhelming majority of people who heard it said that it sounded just like a valve amp - but what is the sound of a valve amp when its at home?

Firstly, I have to disagree with the "overwhelming majority" to some extent. There are other subtle differences between my modified amp and a valve amp that cannot easily be eliminated. As long as the level is reasonably low (not even thinking about clipping), these differences are difficult to detect by ear, but include:

Valve Amp	Modified Transistor Amp
Increasing distortion with level (> 0.05% and rising)	Relatively constant distortion (< 0.01%)
Distortion is almost all low order	Some higher order components
Medium output impedance (about 2 to 6 Ohms, typically)	Very high output impedance (> 200 Ohms)
Relatively low bandwidth	Wide bandwidth

Table 4 - Valve Vs Transistor Amplifiers

The output impedance is by far the most noticeable effect, and results in apparent better bass (but not always - it is sometimes peaky and with some speaker systems becomes "flabby"), and also an improvement in high frequency performance. As the impedance of a speaker rises - either at low frequencies as it approaches resonance, or at high frequencies due to the inductance of the voice coil - a normal amp simply delivers less power. This is because if the voltage is fixed (by the gain of the amp), then less power is delivered to a higher impedance than a lower impedance.

Likewise, where there are impedance dips - often caused by crossover networks - the "normal" amp will deliver more power, often to the detriment of the acoustic balance. An amp with a high(ish) impedance can be thought of as a constant power device - for any given input voltage, the power will remain constant regardless of variations of load impedance. This is not strictly true, of course. No amp can be tailored so the output impedance is just exactly right to give equal power regardless of impedance, but you can get passably close.

What this means is that the speakers will provide more output at the frequencies you are unused to hearing from them, and less where they might have sounded too prominent. The overall sound is not always more accurate (often less so), but it is different, and will generally *seem* to sound better. If my tests are any guide (I hope so!), then most speakers actually benefit from a small amount of induced output impedance, and I am using this technique on my own system for the bass drivers.

UPDATE: Indeed, Nelson Pass mentions this very point, and advises that some speakers will not like the output impedance of the Zen, and will not sound the way they should. (27 Nov 99)

The Zen (and all derivatives thereof) will produce exactly this phenomenon, and I suspect that this is one of the main attractions. If this is the case, the DoZ amplifier should be just what you are looking for. Although more complex than the Zen (but only marginally), it will provide pure Class-A at passably low distortion levels.

I aimed for an output power of at least 12 Watts (Wow - that much!), obtained 19W, and even this will require some fairly large heatsinks to keep temperatures down to reasonable levels. The power supply is much cheaper than that of the Zen, and can use a basic supply or a capacitance multiplier to eliminate ripple, without sending the builder broke with the cost of high value capacitors. The output is capacitor coupled, allowing a simple single supply and no DC servos to try to minimise the DC offset.

I have also designed a simple, high performance preamp circuit (all discrete Class-A), which looks pretty good so far (it has been built and tested - see Project 37). Simulations of this indicated that I will

be completely unable to measure the distortion (less than 0.0001%) because it is so far below the limits of my equipment. The reality is a little different, but the distortion is still very low, and frequency response is very good indeed.

Both the power and preamp are completely new designs, but due to the limited number of sensible amplifier topologies, will almost certainly look like something you have seen already. I have yet to include (as an option) the ability to increase the output impedance up to about 8 Ohms, so you can have the one benefit of the Zen approach without the shortcomings.

References

Zen Amplifier - Designed by Nelson Pass
Impedance - ESP Article, Rod Elliott
10W Class-A Amplifier - An edited version of John Linsley Hood's original

NOTE: I have not published (and do not intend to) the circuit for the Zen amp. It is not my intellectual property and to reproduce it here is not my policy. A web search for "zen" will find it (along with 10,000+ other sites), or you can just go to the Nelson Pass site above. This will not show all the variations, but the idea is much the same for all of them.

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright (c) 28 Oct 1999
Updates: 12 Feb 2000 - corrected a few mistakes, and removed references to what was to be done (since it has been)./ 16 Dec - a few more test results and table 3./ 15 Dec - Added final circuit, and various commentary, fixed a couple of errors./ 29 Oct - Added test info for MTP3055 FET./ 30 Oct - Figure 3 and text./ 27 Nov - Various notes added 27 Nov 99./ 1 Jul 2000 - added some additional commentary and removed redundant (old) comments../ 03 Oct 05 - Rev-A PCB details and photo added.

Elliott Sound Products **Project 15**

A Simple Capacitance Multiplier Power Supply For Class-A Amplifiers

Rod Elliott - ESP (Original Design / Basic Principles)

Introduction

Since I have provided the schematic for John L Linsley-Hood's Class-A amplifier, I felt that some readers may wish to experiment with the concept. Unfortunately, a very low ripple power supply is needed for all Class-A amps, and the most common solution is to use a regulated supply. A basic circuit is provided in the article, but it is assumed that the builder knows all the pitfalls. The supplied schematic is in fact for a capacitance multiplier filter (not a regulator), but is somewhat lacking (I feel) and can be improved dramatically.

While the performance of a true regulated supply will be excellent (if properly designed and built), there are a number of problems for a high-current, low ripple design if regulators are used. Two of the main ones are:

- The regulated output voltage must be lower than the lowest possible voltage from the rectifier/filter combination - including mains ripple. This depends on transformer regulation and mains variations.
- The circuit must be capable of dissipating all excess voltage from the rectifier/filter at the highest possible mains voltage.

For the sake of the exercise, assume that we want the following specifications:

- Output Voltage - 20 Volts (+ve and -ve)
- Output Current - 2.5 Amps max. (1.25 Amps average)
- Mains Voltage - 220 VAC nominal
  - 260 VAC Maximum
  - 200 VAC Minimum

These specifications are typical, since Europe uses 220V mains, Australia uses 240V - the voltages can easily be scaled for the US 115V mains supply - and all are subject to variations, both long and short term.

We are not actually all that interested in the mains input voltage, only the possible variations of the output of the transformer / rectifier / filter combination.

For a regulated output of 20 Volts, we need a minimum input voltage of about 23 Volts, since most regulator circuits have a "dropout" voltage, below which they cannot regulate. This voltage is the absolute minimum, including the mains ripple which will be superimposed on the DC (See Fig 1). Note that for all calculations I am assuming 50Hz mains supply. The results will be slightly different for 60Hz (as used in the US), but are not significant.

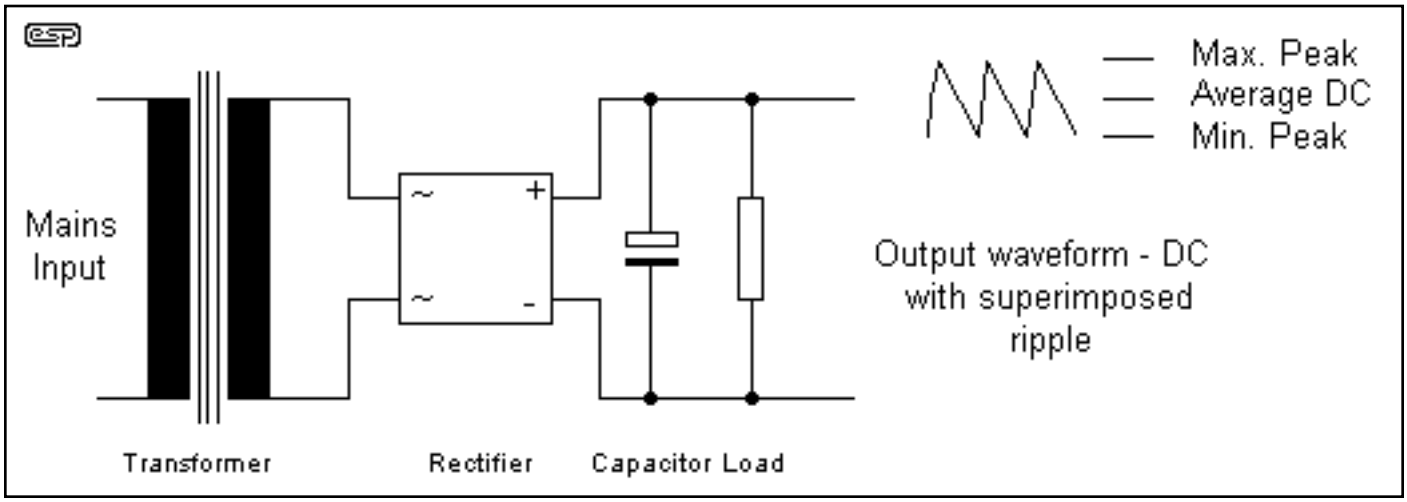


Figure 1 - Basic Rectifier

Once the regulator's input voltage drops below the dropout voltage, regulation will naturally fail, and ripple will appear at the output. This will eventually find its way to our ears, causing much muttering and complaining, and rude words surely cannot be far behind !

Design Considerations - Regulator

We must assume that the transformer / rectifier / filter will have a regulation in the order of 10% (this is fairly typical for a full-wave bridge rectifier). Using the normal 1.414 RMS to peak conversion (the square root of 2), plus a few assumptions based on experience, we therefore have our minimum requirement:-

- Transformer output (at no load) - 16.3 V RMS (each supply) plus diode losses (0.65V) = 17.6V (approx)
- Assume 0.3 Ohm equivalent series resistance in transformer and rectifier diodes
- Assume a filter capacitance of 4700uF as an initial value

This will provide a no load voltage of about 23.5 Volts as expected. When loaded to about 2.5 Amps, this will change:

- Output voltage falls to 19.5 V DC (average)
- Ripple voltage is just over 1 Volt RMS (1.5 Volts peak, triangular wave)
- Minimum output voltage is now 19.5 - 1.5V ripple = 18 Volts

(These figures were simulated, but reality will be suspiciously close!)

It can be readily seen that far more voltage is needed to ensure that the minimum voltage of 23 Volts is maintained. It turns out (again from my trusty simulator) that a transformer voltage of 22 V RMS is needed, which provides an average DC of 25.4 Volts, less about 2 Volts peak ripple. Close enough.

Now comes the really nasty bit! All of the above must be the case at the lowest possible mains voltage. For the sake of (my) sanity, this is assumed to be 200 V AC, so at the above worst case maximum of 260V, the 22V output of the transformer is now 28.6V. At full load (2.5A), this yields a DC voltage of nearly 35V average.

So the regulator will have a minimum input voltage of 25.4 Volts, and a maximum of 35V, so power dissipation will be:

- 6.75 Watts average at minimum input voltage and 1.25A average current
- 18.75 Watts average at maximum input voltage (also at 1.25A)
- Somewhere in between for nominal mains supply voltages and output current variations.

Note that the above figures are for the 1.25A average, but peak dissipation (at 2.5A) will be double, at about 37W for the worst case. This is a lot of heat to dispose of, and must be catered for. I should also mention that a minimum of 200V AC for a nominal 220V is probably optimistic (10% of 220V is 22V) and in reality we may need to cater for even lower voltages. This makes the equation even worse !

To accommodate the worst case, the heatsink for the power supply must be capable of ensuring the maximum device temperature is not exceeded at the highest mains voltage anticipated. At no "normal" mains voltage may the regulator come out of regulation, or severe ripple will appear at the output, degrading the sound quality, and causing audible hum (at double the mains frequency, and with a triangular waveshape, which sounds horrid).

Capacitance Multiplier - Design Considerations

The only real thing to worry about is the degree of filtering needed! We must assume that at least 3 Volts will be lost across the capacitance-multiplier filter, to ensure that the DC input (including ripple component) always exceeds the output voltage.

Because there is no regulation, the power amplifier must be capable of accepting the voltage variations from the mains - every standard power amplifier in existence does this quite happily now, so it is obviously not a problem. Note that the output power is affected, but this happens with all amps, and cannot be avoided without a regulator.

We can now design for nominal mains voltage (say 220V AC), and with very simple circuitry, provide a filter which will dissipate no more than about 4 Watts in normal use - regardless of mains voltage. Figure 2 shows the basic configuration of a capacitance-multiplier filter, where the capacitance appearing at the base of the output device is effectively multiplied by the gain of the device - thus a 1000uF capacitor appears (electrically) to be a 1 Farad (yes, 1,000,000uF) cap, assuming a gain of 1000 in the output device.

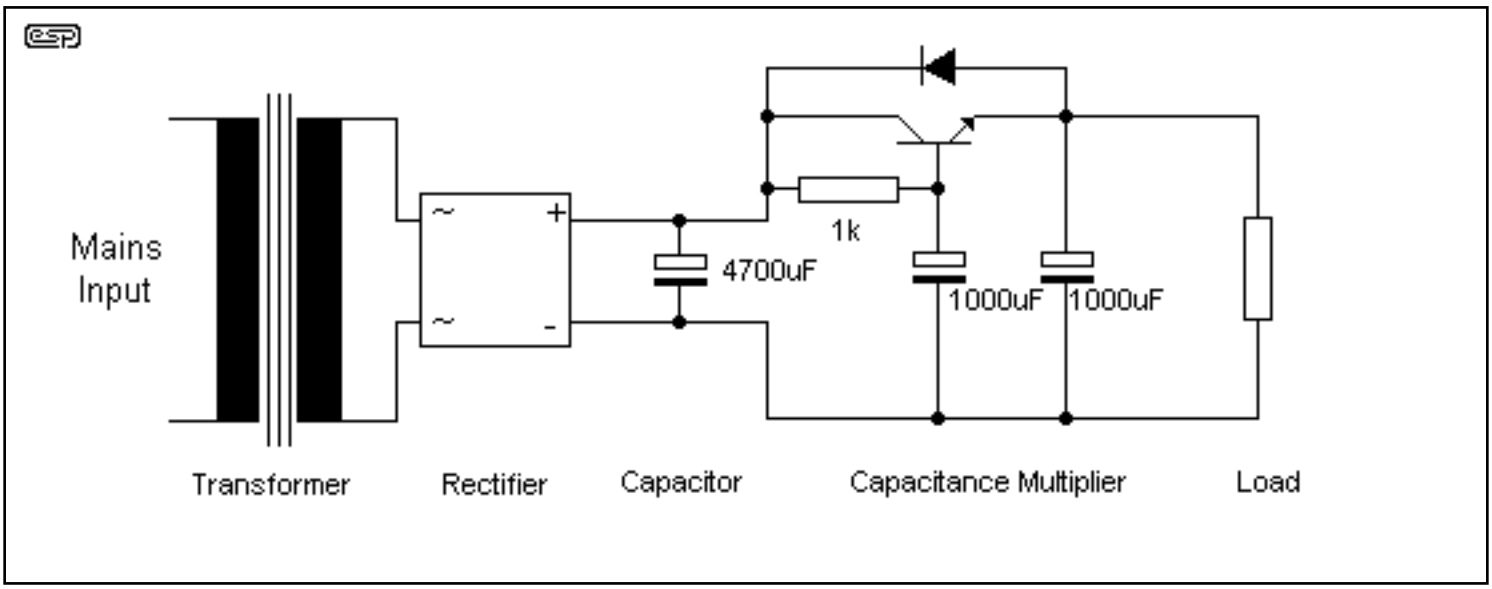


Figure 2 - Single (Basic) Capacitance Multiplier

One could simply use a pair of 1F caps for a dual supply, but I have noticed a dearth of such devices (other than the 5V "Supercaps" used for memory backup in computers). Since they will need to be rated at about 35V, and be capable of considerable ripple current, I cannot help but feel that this is not a viable option.

Both methods will provide a ripple of well under 5mV RMS, but the multiplier has the advantage of removing the triangular waveform - it is not a sinewave, but has a much lower harmonic content than would be the case even with a 1F capacitor.

To obtain a gain of 1000 for a transistor, we need to use a darlington - either an encapsulated darlington device, or a pair of "ordinary" transistors connected in a darlington pair (See Figure

3). The latter method is my preferred option, since it allows greater flexibility in obtaining devices, and will often have better performance.

The Final Design

The simple capacitance multiplier filter described above is quite satisfactory as a starting point, but its final operating characteristics are too dependent on the gain of the output transistors. What is needed is a circuit whose performance is determined by resistors and capacitors, and which is relatively independent of active devices (although these will still have an impact on the degree of filtering provided).

We can also improve the ripple rejection, and the final circuit for a dual supply is shown in Figure 3. This circuit reduces ripple to less than 200uV with typical devices (160uV as simulated), and dissipates less than 4 Watts per output transistor at 1.25A continuous operating current. It is unlikely that you will achieve this low hum level in practice, since real wire has resistance. However, with careful layout you should easily be able to keep the output hum and noise to less than 10mV, and this level is more than acceptable for any power amp application.

By splitting the capacitance with an additional resistor, we create a second order filter (12dB/ octave rolloff), which reduces the hum more effectively, and also removes more of the higher order harmonics (which tend to make a "hum" into a "buzz" - much more audible and objectionable). The resistor to ground stabilises the circuit against variations in transistor gain, but increases dissipation slightly. This is done deliberately to ensure that there is sufficient voltage across the multiplier to allow for short term variations.

The 12k resistor shown may need to be adjusted to suit your transistors and supply voltage. Reducing the value increases dissipation in the output devices and lowers output voltage. It is unlikely that any benefit will be obtained by increasing this resistor, other than increased hum (hardly a benefit).

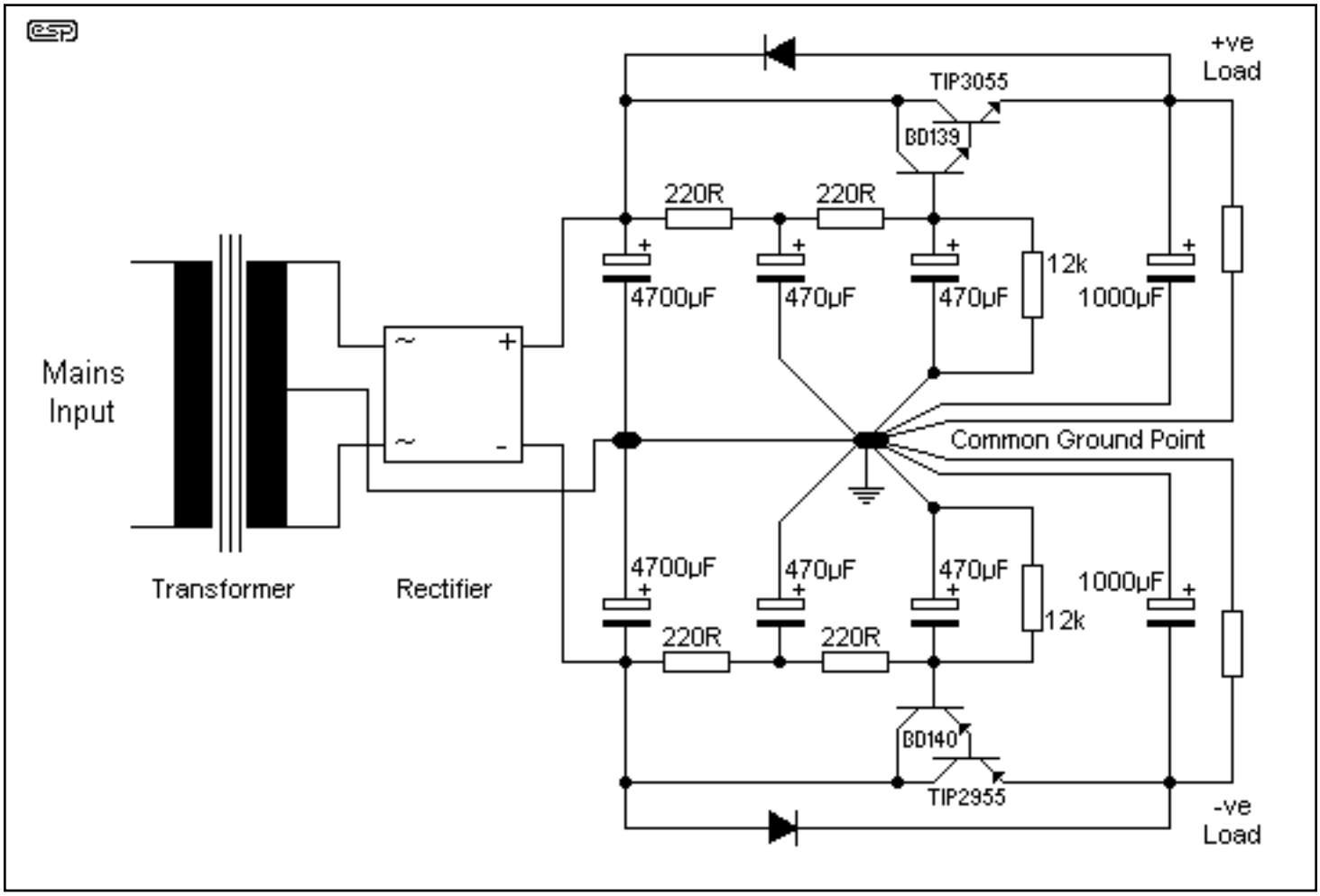


Figure 3 - Complete Dual Capacitance Multiplier

This is an easy design to build, but requires great care to ensure that ripple currents are not superimposed on the output because of bad grounding or power wiring practices. The schematic is drawn to show how the grounds of the various components should be interconnected, using a "star" topology. If this is not followed, then excessive hum will be the result.

The previous drawing (which you may have downloaded or printed) showed the bridge output not directly connected to the main filter capacitors - this was not my original intention, and the correct method is shown above. Normally, in a schematic diagram the idea is to show the connections, rather than the physical circuit layout. This diagram is an exception, and the physical layout should match the schematic (in as much as that is possible, at least).

Note that the transformer is centre-tapped, and requires equal voltage on each side - in this case, somewhere between 18 and 22V AC. It is most important that the centre-tap is connected to the common of the two input filter capacitors (4700uF), and that this common connection is as short as possible. Use of a solid copper bar to join the caps is recommended. Likewise, a solid copper disk (or square) is suggested for the common ground, tied as closely as possible to the capacitor centre tap. The resistance of the main earth connection is critical to ensure minimum hum at the output, and it cannot be too low.

Because the circuit is so simple, a printed circuit board is not needed, and all components can be connected with simple point-to-point wiring. Keep all leads as short as possible, without compromising the star grounding. For convenience, the driver transistors may be mounted on the heatsink, which does not need to be massive - a sink with a thermal resistance of about 5 degrees C per Watt (or better) should be quite adequate (one for each output device). Remember that the lower the thermal resistance, the cooler everything will run - this improves reliability.

Increasing the capacitance (especially at the input) is recommended, and I would suggest 10,000uF as the ideal instead of the 4700uF shown. This will reduce hum even further, and provide greater stability against short term mains voltage changes. Increased output capacitance will help when powering Class-AB amplifiers, to account for their sudden current demands. I do not recommend more than 4,700uF, as the charging current will be very high.

Although generic transistor types (such as the 2N3055) can be used, it is better if devices with somewhat more stable characteristics (from one device to the next) are used. Plastic (T0-220) devices are fine for the output as shown, but if higher voltage or current is needed you might have to use TO-3 types.

For the components, I would suggest the following as a starting point (or equivalents):

Output Transistors	TIP3055 (TIP2955 for the -ve supply)
Drivers	BD139 (BD140 for the -ve supply)
Resistors	1/4W metal film for all resistors
Diodes	1N4001 or similar
Electros	No suggestions, but make sure that their operating voltage will not be exceeded, and observe polarity. (Bypassing with polyester is not really necessary, but if it makes you feel better, do it)
Bridge rectifier	20 to 35A Amp bridge is recommended. This is overkill, but peak currents are high, especially with large value capacitors. Also ensures minimum diode losses at normal currents.
Transformer	Use a toroidal. Power (VA) rating for supply as shown should be as required for the amplifier. A dual 20W Class-A amp will ideally have a minimum transformer rating of 200VA - 5 times the amplifier power. (Note that VA is sometimes incorrectly quoted in watts). Primary voltage is naturally dependent upon where you live.

Matching the output and driver transistors is not be necessary, but will result in marginally

improved performance if done. Use devices with the highest gain possible for best results.

To use the above circuit in single-ended mode, the transformer will need only a single winding (or paralleled sindings if this is possible). Simply wire the transformer and bridge as shown in Figure 2, and leave off the negative multiplier circuit (i.e. everything below the common ground point).

Using A Capacitance Multiplier Filter With Class-AB Amps

Note that this circuit is quite suitable for Class-AB amplifiers, but since their current requirements vary so widely adding a (much?) larger capacitance to the output, is a must. The diode is recommended as shown to prevent the possibility of reverse biasing (and destroying) the transistor(s) when power is removed.

The benefits of such filtering are subtle, but worth the effort. Many power amps are now built with truly massive capacitance after the rectifier. This reduces hum which is introduced into the signal during loud passages. In theory, this is inaudible - but if so, why do amps with very large capacitor banks always seem to sound better? (Or so the reviewers keep telling us.)

If you are desirous of trying this circuit with a Class-AB amp, I would strongly recommend that the input to output voltage differential be increased (reduce the 12k resistor to do this), or you are unlikely to be pleased with the result. For optimum performance (depending on output voltage, the current variation, etc), I would suggest that a differential of 6V to 10V should be Ok, depending upon the power of the amp. Dissipation will need to be calculated (or measured), and remember that Class-AB amps can (and do) create peak currents which can be very high indeed. For the 20V (nominal) supply shown, peak current into an 8 ohm load is 2.5A (which was the design goal in the first place), but if the voltage is increased, peak currents increase in proportion.

As an example, consider a 100W amp (8 ohms). Peak current into a resistive load is about 3.6A, but when driven into a typical speaker load (whose impedance dips to (say) 3 ohms), the peak current will be 9.6A. This is not mere speculation, but harsh reality - such peak currents are quite common - one of the reasons many manufacturers quote the peak output current of their amps. These specs. can be as high as 40 Amps (for a 100W unit), which is probably overkill, but should be remembered.

It must be remembered that this circuit acts in a manner very similar to a regulator - just without the regulation. If the output current is highly transient in nature, the circuit will allow hum to pass if the input voltage suddenly drops due to increased load (in the same way a regulator will).

Also note that the supply voltage to the power amp(s) will be modulated by the instantaneous current drain of the amp (which happens with "conventional" supplies too). Maintaining a voltage differential sufficient to accommodate these variations is imperative.

WARNING: Because this power supply is mains operated, there is the risk of electrocution if extreme care is not exercised while constructing or testing the unit. If you are not confident in your abilities with mains powered equipment, ***do not attempt construction*** under any circumstances .... please!

Projects Index Page

Home Page

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Updated 21 Apr 2001 - Changed drawings, and amended start earth and transformer rating info

talkabout network® we've got all the answers

Register & Login
Nick Password Login >>
> Create new account > Recover lost password

PA versus Powered Mixer

by "Frank D" <fdo.nospam@> > Sep 24, 2004 at 07:06 PM

We are a four piece band with 2 guitars, bass, and acoustic drums. I'm looking for amplification that as well as voice(3 Mics), will also mic my kick drum and maybe accept a keyboard. We'll be playing smaller venues and basement/garage jams. My local dealer suggested a SoundTech PMXR200 power mixer. Any suggestions out there RE: PA versus Powered Mixer
Frank

79 Posts in Topic:

PA versus Powered Mixer	"Frank D" <f	2004-09-24 19:06:22
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-09-24 23:25:42
Re: PA versus Powered Mixer	"Sugarite" <	2004-09-24 20:18:39
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-09-25 00:57:28
Re: PA versus Powered Mixer	"liquidator" &l	2004-09-24 21:01:29
Re: PA versus Powered Mixer	foodsteaks@ [EMAIL PROTECT	2004-09-30 15:33:54
Re: PA versus Powered Mixer	"Dale C" <so	2004-10-01 06:43:43
Re: PA versus Powered Mixer	"liquidator" &l	2004-10-01 15:06:51
Re: PA versus Powered Mixer	"Dale C" <so	2004-10-01 14:42:00
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-10-01 20:10:22
Re: PA versus Powered Mixer	Ralph Staub <ralph@[EM	2004-10-01 12:17:29
Re: PA versus Powered Mixer	"liquidator" &l	2004-10-01 15:11:20
Re: PA versus Powered Mixer	Ralph Staub <ralph@[EM	2004-10-02 01:34:02
Re: PA versus Powered Mixer	walkinay@[EMAIL PROTECTED	2004-10-02 02:49:13
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-10-02 06:03:48
Re: PA versus Powered Mixer	Ralph Staub <ralph@[EM	2004-10-02 14:27:27
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-10-02 14:33:08
Re: PA versus Powered Mixer	"liquidator" &l	2004-10-02 09:01:58
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-10-02 13:59:34
AC,DC, and clipping: WAS: PA versus Powered Mixer	Ralph Staub <ralph@[EM	2004-10-02 16:11:48
Re: AC,DC, and clipping: WAS: PA versus Powered Mixer	George <g.p. gleason@[E	2004-10-02 16:29:07

Re: AC,DC, and clipping: WAS: PA versus Powered Mixer	"Steve White" &	2004-10-02 17:44:11
Re: AC,DC, and clipping: WAS: PA versus Powered Mixer	foodsteaks@ [EMAIL PROTECT	2004-10-02 19:24:58
Re: AC,DC, and clipping: WAS: PA versus Powered Mixer	Ralph Staub <ralph@[EM	2004-10-03 02:57:01
Re: AC,DC, and clipping: WAS: PA versus Powered Mixer	"Sugarite" <	2004-10-04 00:06:39
Re: AC,DC, and clipping: WAS: PA versus Powered Mixer	"Phil Allison"	2004-10-04 15:09:27
Re: AC,DC, and clipping: Is this horse dead yet?	"djk" <dklei	2004-11-08 02:37:01
Re: AC,DC, and clipping: Is this horse dead yet?	"Phil Allison"	2004-11-08 20:10:56
Re: AC,DC, and clipping: Is this horse dead yet?	Mike T. <miket@ [EMAIL	2004-11-08 16:04:20
Re: AC,DC, and clipping: Is this horse dead yet?	"Phildo" <Ph	2004-11-08 21:41:33
Re: AC,DC, and clipping: Is this horse dead yet?	"Boa" <bo@[E	2004-11-10 13:08:44
Re: PA versus Powered Mixer	"Arny Krueger"	2004-10-02 20:08:21
Re: PA versus Powered Mixer	"liquidator" &l	2004-10-02 22:43:12
Re: PA versus Powered Mixer	"Phil Allison"	2004-10-03 00:00:38
Re: PA versus Powered Mixer	Ralph Staub <ralph@[EM	2004-10-02 16:29:14
Re: PA versus Powered Mixer	"liquidator" &l	2004-10-02 13:49:25
Re: PA versus Powered Mixer	Ralph Staub <ralph@[EM	2004-10-02 22:11:44
Re: PA versus Powered Mixer	"Phil Allison"	2004-10-03 13:54:08
Re: PA versus Powered Mixer	"Arny Krueger"	2004-10-02 20:06:38
Re: PA versus Powered Mixer	"Phil Allison"	2004-10-03 12:57:28
Re: PA versus Powered Mixer	"Arny Krueger"	2004-10-03 15:22:55
Re: PA versus Powered Mixer	"Phil Allison"	2004-10-04 12:03:16
Re: PA versus Powered Mixer	"Arny Krueger"	2004-10-04 08:44:57
Krueger blunders Again	"Phil Allison"	2004-10-04 23:28:15
Re: Krueger blunders Again	"Arny Krueger"	2004-10-04 12:43:01
Re: Krueger blunders Again	"Phil Allison"	2004-10-05 11:32:12
Re: PA versus Powered Mixer	"Frank D" <f	2004-09-26 23:07:33
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-09-27 11:19:08
Re: PA versus Powered Mixer	"Sugarite" <	2004-09-27 15:39:04
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-09-27 20:06:02
Re: PA versus Powered Mixer	"Sugarite" <	2004-09-28 13:45:40
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-09-28 19:32:30
Re: PA versus Powered Mixer	"liquidator" &l	2004-09-28 17:51:31

Re: PA versus Powered Mixer	"Sugarite" <	2004-09-29 02:31:50
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-09-29 11:02:57
Re: PA versus Powered Mixer	"Sugarite" <	2004-09-29 13:18:09
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-09-29 17:45:06
Re: PA versus Powered Mixer	"Sugarite" <	2004-09-29 15:35:39
Re: PA versus Powered Mixer	"liquidator" &l	2004-09-29 16:12:06
Re: PA versus Powered Mixer	"Sugarite" <	2004-09-30 04:10:40
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-09-30 10:17:54
Re: PA versus Powered Mixer	"Sugarite" <	2004-09-30 07:48:00
Re: PA versus Powered Mixer	Steve Hawkins <res0pf0	2004-09-29 15:14:12
Re: PA versus Powered Mixer	bmiawmb@	2004-09-24 18:48:07
Re: PA versus Powered Mixer	"Tim Padrick" &	2004-09-24 19:11:33
Re: PA versus Powered Mixer	"Dan Fraser" &l	2004-09-26 03:57:45
Re: PA versus Powered Mixer	"Sugarite" <	2004-09-27 01:59:17
Re: PA versus Powered Mixer	"Dale C" <so	2004-09-28 08:14:09
Re: PA versus Powered Mixer	"liquidator" &l	2004-09-28 17:49:37
Re: PA versus Powered Mixer	"Dale C" <so	2004-09-29 09:06:02
Re: PA versus Powered Mixer	"Sugarite" <	2004-09-29 13:29:25
Re: PA versus Powered Mixer	"liquidator" &l	2004-09-29 14:58:03
Re: PA versus Powered Mixer	"Sugarite" <	2004-09-30 07:55:19
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-09-30 18:06:59
Re: PA versus Powered Mixer	"Sugarite" <	2004-10-01 12:40:20
Re: PA versus Powered Mixer	"liquidator" &l	2004-10-01 14:55:12
Re: PA versus Powered Mixer	George <g.p. gleason@[E	2004-10-01 20:07:30
Re: PA versus Powered Mixer	"liquidator" &l	2004-09-29 15:17:49
Re: PA versus Powered Mixer	"Dan Fraser" &l	2004-09-30 04:14:49

Post A Reply:

Sign Up Now and receive benefits of the talkabout network.
Post replies, new topics, bookmark post, email post to friends and much more. Click here to Sign Up.

All trademarks and copyrights on this page are owned by their respective owners. The posts express the views of the author of the post and are not necessarily the view of the Talk About Network. If you feel a posted message is objectionable, you are encouraged to contact us immediately by email.

Developed and Powered by MediaGlitch.com. Copyright © 2003-2005 Talk About Network, Inc. All right reserved. Talk About Network® is a registered trademark.

tan3V112 Mon May 29 12:21:55 EDT 2006.

News Desk

Negative feedback doesn't always decrease amplifier distortion!

By John Atkinson

December 21, 1997 — Martin Colloms argues persuasively in the January 1998 *Stereophile* that negative feedback is not the panacea that amplifier designers believe it to be. His experience of an amplifier (the Cary CAD-805C) and a preamplifier (the Conrad-Johnson ART) that use no negative feedback other than local degeneration, yet have sound quality better than he has previously experienced, convinces him that even when a design's closed-loop distortion appears to be acceptably low, the listener is still aware of an amplifier's very distorted open-loop behavior.

In the Six/1997 issue of *Audio Electronics*, an article by Jan Didden, "Feedback---Curse or Blessing?," provides supporting evidence for Martin's conjecture. Didden summarizes research performed by the late British engineer Peter Baxandall (though no reference is given). Baxandall found that, as loop negative feedback is introduced to a solid-state (FET) amplifier circuit, the amplitude of the low harmonics is reduced by the feedback ratio, as expected. But what was not expected was that the amplitudes of the higher harmonics *increased* as the feedback was increased in level. These high harmonics started to decrease in level only when the amount of loop negative feedback crossed a threshold. With 40dB of negative feedback, the level of the fifth harmonic was still higher than it was without any feedback!

So what, you may be asking yourself? The point is that the higher harmonics are considerably more audible than the low harmonics, and are subjectively more unpleasant. You can test this yourself with *Stereophile Test CD 2*, which includes a track allowing listeners to determine for themselves how much second- and seventh-

Subscribe
Renew
Give a Gift
Sub Services
Special Web Offer: 12 ISSUES for $12.97!

Search Site
eNewsletter
Your E-ma
Dealer Locator
Zip Code

2 days until showtime!
• Seminar Schedule Now Online
• Rooms Only $129!
• Analog Clinic with Mike Fremer
• Record Pressing Plant Tour
• Near the Los Angeles Airport
• More . . .

Equipment Reviews From Stereophile Magazine

harmonic distortion they can hear with a pure tone. While it's hard to hear 1% of second-harmonic content, 0.1% of seventh-harmonic is still audible to many listeners.

So before you dismiss audiophiles in love with the sound of single-ended tube amplifiers as anachrophilic tweako-cultists, you should note that the open-loop behavior of such designs is extremely linear compared with that of typical push-pull solid-state designs. This means that they can get away with a very small amount of loop negative feedback, and that what harmonic distortion they *do* have is almost all low-order.

Audio Electronics, by the way, is the new title of *Audio Amateur* magazine, published six times a year by Audio Amateur Corporation, P.O. Box 576, Peterborough, NH 03458. Tel: (603) 924-9464. Fax: (603) 924-9467. E-mail: audiotech@top.monad.net.

More News

Sponsored Links

- **Digital Audio Chips**
  2W-80W of Pure Class-D Pure Sound, from Silicon Valley
  www.MonolithicPower.com
- **Great High-End Audio Site**
  Over 2,000 High End Audio Products Amplifiers, Cables, Speakers, Etc.
  www.audioadvisor.com
- **Speakers & Components**
  Audio equipment for professional, home theater, automotive and more!
  www.mcminone.com
- **Fix Any Computer Easy**
  Partition, Format, Fix Hard Disk. Rescue & Install Windows Easy!
  PCbeginner.com
- **amplifier**
  All the Design Resources You Need for Your Next Op Amp
  amplifiers.national.com

> Recent Additions
> Budget Components

Amplification
> Tube Preamps
> Tube Power Amps
> Solid-State Preamps
> Solid-State Power Amps
> Integrated Amps/Receivers
> Phono Preamps

Accessories
> Cables
> Phono Accessories
> Powerline Accessories
> Room Acoustics Treatments
> Stands/Racks
> Headphones/Headphone Accessories
> Miscellaneous

Loudspeakers
> Floorstanding Loudspeakers
> Standmounted/Bookshelf Loudspeakers
> Subwoofers

Digital Sources
> CD Players/Transports
> Hi-Rez Disc Players/ Transports
> Digital Processors
> Computer Audio
> Media Servers/MP3 Players

Analog Sources
> Turntables
> Tonearms
> Phono Cartridges
> FM Tuners

More Articles From *Stereophile* Magazine

Columns
> As We See It
> Listening / Art Dudley
> The Fifth Element / John Marks
> Music in the Round / Kal Rubinson
> Fine Tunes / Jonathan Scull

Other Features
> Special Features
> Reference
> Interviews
> Think Pieces
> Historical

Music Features
> Recording of the Month
> Records 2 Die 4
> Music/Recordings

Site Features

Blogs
> Wes Phillips
> Stephen Mejias
> Robert Baird

Community
> Forums
> Galleries
> Vote
> Previous Votes

Latest News
> Audio News
> Past eNewsletters
> CES 2006 Reports
> FSI 2006 Reports

The Music Room
> Record Reviews
> Groove Migrations

Spotlights
> Vikuiti See Clear Now
> XM Experience

Resources
> Dealer Locator
> AV Links
> Current Issue
> Master Article Index
> Audiophile Societies
> Contact Us
> Customer Service

TRIAL ISSUE!

Here's the fastest way to bring home the only magazine that entertains, informs and educates its readers about the latest technology--*Stereophile Magazine*--at no risk! During this special online offer, you can get a **TRIAL ISSUE** and receive 11 more (a total of 12 issues) for ONLY $12.97-**you save over $58 off the cover price!**

If you choose not to subscribe, just write "cancel" on your invoice, send it back and owe nothing. Either way the trial issue is yours to keep--without obligation. Just complete the information below and click submit.

GIVE A GIFT

Email:
First Name:
Last Name:
Address Line 1:
Address Line 2:
City:
State: Zip:

Select a payment option:
- (•) Charge my credit card
- ( ) Bill me later

Do you have a promotional coupon code?

Enter Code:

☑ Please send me special offers and exclusive promotions from Stereophile's premiere partners.

Submit Order

> HE 2006 Reports

AUDIOPHILE ECOMMERCE

STEREOPHILE PRODUCTS
> Recordings
> Backissues
> Buyer's Guides

AV EQUIPMENT DEALS
> AV Marketplace

STEREO OFFERS
> Phono Preamp
> Hi-Fi
> Phono Cartridge
> Amplifiers
> Stereo Speakers

OTHER HOME TECH SITES OF INTEREST

AUDIO VIDEO SITES
> Ultimate AV
> Home Theater
> Audio Video Interiors

PHOTOGRAPHY SITES
> Shutterbug

AV SHOWS
> Home Entertainment Show

Home/News | **Subscribe** | **Give a Gift** | **Sub Services** |
Recordings | **Back Issues** | **Buyer's Guide** | **AV Marketplace**
Print & Web Media Kit | **Privacy** | **Terms of Use** | **Become An Affiliate** | **Contact Us**

Copyright © Primedia Magazines, Inc. All rights reserved.

Transtronics®

Search Transtronics

Boolean: AND Case Insensitive

Capacitor ESR Ratings

ESR Equivalent Series Resistance

See the Cap wizard
Surplus Capacitors
Secret Capacitor codes

The ESR rating of a capacitor is a rating of quality. A theoretically perfect capacitor would be loss less and have an ESR of zero. It would have no in-phase AC resistance. We live in the real world and all capacitors have some amount of ESR. To understand why let us review what a capacitor is and what they are made of and how we rate them.

What is a Capacitor?

A capacitor consists of two conductive metal plates separated by an insulating dielectric. The dielectric can be made of glass, ceramic, Tantalum oxide, or plastics such as polyethylene or polycarbonate. Even air can be used as the dielectric. When the capacitor holds some energy in the form of extra electrons on one plate and electron holes on the other we say that the capacitor is charged.

Farads

Capacitance (C) is the amount of charge per volt of potential that a capacitor holds. (C =Q/V where Q = coulombs (the unit of charge) and V = Volts)

Capacitance is measured in farads, but most often a small fraction of a farad thus:

- micro-farads uF millionths (10^{-6}) farads
- pico-farads pF (10^{-12}) farads (sometimes called 'puffs' in engineering slang)

The energy stored in a capacitor is $E = CV^2/2$ E is in joules.

Thus, the average power in watts is $P_{av} = CV^2/2t$ where t = time in seconds.

The maximum voltage rating and its capacitance determine the amount of energy a capacitor holds. The voltage rating increases with increasing dielectric strength and the thickness of the dielectric. The capacitance increases with the area of the plates and decreases with the thickness of the dielectric.

Thus, the capacitance of a capacitor (C) is related to the plate area (A), plate separation distance (d) and permittivity (ε) of the dielectric by the following equation:

$C = \varepsilon A/d$ Here A and d are based on meters as the unit and ε is in coulombs squared per Newton-meters squared notice the force unit involved - it explains why capacitor microphonics (remember the good old condenser microphone?) and a mechanical failure mode of capacitors).

Dielectric Constants

Dielectric constant (k) gets it's value by comparison of the charge holding ability of a vacuum where k = 1. Thus, k is the ratio of the capacitance with a volume of dielectric compared to that of a vacuum dielectric.

$K = \varepsilon_d/\varepsilon_0$ Where ε_d is the permittivity of the dielectric and ε_0 is the permittivity of free space

Air has nearly the same dielectric value as a vacuum with k = 1.0001. Teflon, a very good insulator, has a value of k =2 while the plastics range in the low 2s to low 3s. Mica gets us a k =6. Aluminum oxide is 7, Tantalum's k is 11 and the Ceramics range from 35 to over 6,000.

Dielectric constants vary with temperature, voltage, and frequency making capacitors messy devices to characterize. Whole books have been written about choosing the correct dielectric for an application, balancing the desires of temperature range, Temperature stability, size, cost, reliability, dielectric absorption, voltage coefficients, current handling capacity (ESR). (Ivan Sinclair wrote a nice book on passives; unfortunately, it is out of print. This points to the fact that our universities are no longer teaching this material). See the Secrets of capacitor codes page for more information

Dielectric strength

Dielectric strength is a property of the dielectric that is usually expressed in volts per mil (V/.001") or volts per centimeter (V/cm). If we exceed the dielectric strength, an electric arc will 'flash over and often weld the plates of a capacitor together.

Q or Quality Factor

The *Q* of a capacitor is important in tuned circuits because they are more damped and have a broader tuning point as the *Q* goes down.

$Q = 1/RX_C$ where X_C is the capacitive reactance *where* $X_C = 1/(2\pi FC)$ and *R* is the soon to be defined term of ESR.

Q is proportional to the inverse of the amount of energy dissipated in the capacitor. Thus, ESR rating of a capacitor is inversely related to its quality.

Dissipation Factor

The inverse of Q is the dissipation factor (δ). Thus, $\delta = ESR/X_C$ and the higher the ESR the more losses in the capacitor and the more power we dissipate. If too much energy is dissipated in the capacitor, it heats up to the point that values change (causing drift in operation) or failure of the capacitor.

Ripple Current Rating

The ripple current is sometimes rated for a capacitor in RMS current. Remembering that $P = I^2R$ where R in this case is ESR it is plain to see that this is a power dispassion rating.

Dielectric Absorption

This is the phenomenon where after a capacitor has been charged for some time, and then discharged, some stored charge will migrate out of the dielectric over time, thus changing the voltage value of the capacitor. This is extremely important in sample and hold circuit applications. The typical method of observing Dielectric Absorption is to charge up a cap to some known DC voltage for a given time, then discharge the capacitor through a 2 ohm resistor for one second, then watch the voltage on a high-input-impedance voltmeter. The ratio of recovered voltage (expressed in percent) is the usual term for Dielectric absorption.

The charge absorption effect is caused by a trapped space charge in the dielectric and is dependent on the geometry and leakage of the dielectric material.

ESL

ESL (Equivalent Series Inductance) is pretty much caused by the inductance of the electrodes and leads. The ESL of a capacitor sets the limiting factor of how well (or fast) a capacitor can de-couple noise off a power buss.

The ESL of a capacitor also sets the resonate-point of a capacitor. Because the inductance appears in series with the capacitor, they form a tank circuit.

ESR Defined

ESR is the sum of in-phase AC resistance. It includes resistance of the dielectric, plate material, electrolytic solution, and terminal leads at a *particular frequency*. ESR acts like a resistor in series with a capacitor (thus the name Equivalent Series Resistance). This resister can cause circuits to fail that look just fine on paper and is often the failure mode of capacitors.

To charge the dielectric material current needs to flow down the leads, through the lead plate junction, through the plates themselves - and even through the dielectric material. The dielectric losses can be thought of as friction of aligning dipoles and thus appear as an increase (or a reduction of the rate of decrease -- this increase is what makes the resistance vs freq line to go flat.) of measured ESR as frequency increases.

As the dielectric thickness increases so does the ESR. As the plate area increases, the ESR will go down if the plate thickness remains the same.

To test a Capacitors ESR (see the Cap Wizard) requires something other than a standard capacitor meter. While a capacitor value meter is a handy device, it will not detect capacitor failure modes that raise the ESR. As the years go by, more and more designs rely on low ESR capacitors to function properly. ESR failed caps can present circuit symptoms that are difficult to diagnose.

Formulas at a glance

(if you can't see the formulas it is time to get a browser update)

$$\delta = \frac{ESR}{X_c}$$

$$X_c = \frac{1}{2\ \pi CF}$$

$$\mathrm{X_L} = 2\ \pi \mathrm{LF}$$

$$\omega = 2\ \pi \mathrm{F}$$

$$\mathrm{I} = \frac{\mathrm{E}}{\sqrt{R^2 + \left(\omega\ \mathrm{L} - \frac{1}{\omega\ \mathrm{C}}\right)^2}}$$

$$\mathrm{F_r} = \frac{1}{2\ \pi\sqrt{LC}}$$

Where *k* = dielectric constant, *A* = area, *t* = thickness of the dielectric, *Q* = coulombs the unit of charge, and *V* = Volts

$$C = \frac{(8.85 \times 10^{-12})kA}{t}$$

$$C = \frac{Q}{V}$$

Where *A* (area) and *d* (thickness) use meters as the unit and ε is in coulombs (squared per Newton-meters squared), ε_d is the permittivity of the dielectric, and ε_0 is the permittivity of free space

$$C = \frac{\in A}{d}$$

$$k = \frac{\in_d}{\in_0}$$

Where energy E (in joules) stored in a capacitor is given by

$$E = \frac{CV^2}{2}$$

Thus, the average power in watts where t = time in seconds.

$$P_{av} = \frac{CV^2}{2t}$$

$$Z_0 = \sqrt{ERS^2 + X_c^{\,2}}$$

- **T**

  ime **D**omain **R**efletometry (**TDR**) formulas
- Characteristic Impedance of cable formulas
- Discontinuance of transmission characteristic impedance

Z_a = characteristic impedance through which the incident wave travels first and Z_b is the characteristic impedance through which the incident wave travels next. V_r is the reflected wave amplitude, V_i is the incident wave amplitude, and V_t is the transmitted wave amplitude.

$$\frac{\mathrm{Vr}}{\mathrm{Vi}} = \frac{\mathrm{Zb} - \mathrm{Za}}{\mathrm{Zb} + \mathrm{Za}}$$

$$\frac{\mathrm{Vr}}{\mathrm{Vt}} = \frac{2\,\mathrm{Zb}}{\mathrm{Zb} + \mathrm{Za}}$$

$$\mathrm{Zb} = \mathrm{Za}\,\frac{\mathrm{Vi} + \mathrm{Vr}}{\mathrm{Vi} - \mathrm{Vr}}$$

$$\tau = RC = \frac{Z_0}{2} C$$ eristic impedance:

Was this Information Useful?

If you found this information useful - all I ask is to look at our home page and see if we have any products that might be of use to you or a colleague. Link to us if you have a web page. If you have some thing to add to this page please send it to the e-mail below.

Transtronics®	3209 W.9th street Lawrence, KS 66049 USA	Ph FAX Email WEB	(785) 841 3089 (785) 841 0434 inform@xtronics.com http://xtronics.com	Bookmark this page

Search Transtronics

Boolean: AND *Case* Insensitive

Transtronics Home Page See our line of industrial control electronics	PLC's Index	PC test equipment and EPROM programmer	Process Control Panel meters	Current sensors	Resource library handbooks, primers and spec sheets

Corporate information and privacy statement

(C) Copyright 1994-2006, Transtronics, Inc. All rights reserved

Transtronics® is a registered trademark of Transtronics, Inc.

ELECTRONICS WORLD

To view the website please go to the new address.

All enquiries should be directed to Michael Allen, Nexus Media Communications.

Phone: +44 (0) 1322 611319
Fax: +44 (0) 1322 616317
Email: m.allen@nexusmedia.com

Website: http://www.ewmag.co.uk

NEXUS
MEDIA COMMUNICATIONS

INDUCTANCE OF A STRAIGHT WIRE

Tim Healy, Santa Clara University

This note concerns the inductance of a straight wire. Surprisingly, it is not easy to find an expression for the inductance of a straight piece of wire, and it is far more difficult to find out how that expression was obtained. And yet this result is very useful as a building block for more complex structures, and also because the inductance of a wire is important in high frequency or high speed electrical circuits.

The first really convenient derivation of the inductance of a wire which I have been able to find is due to Rosa(1). The first dilemma which Rosa faced is that there is no unambiguous definition of how to define the inductance of a straight wire. If we consider the wire in isolation we ignore the question of how the current gets to the wire. But that current, however it is delivered, will affect the flux which is developed in the vicinity of the wire. But this flux is a part of the definition. Rosa resolved the dilemma arbitrarily (there is no other way) by defining the inductance as the ratio of the flux developed in the region bounded by lines perpendicular to the beginning and end of the wire. Let's look at Rosa's own words in which he recognizes the dilemma in the definition.

"I have derived the formulae in the simplest possible manner, using the law of Biot and Savart in the differential form instead of Neumann's equation, as it gives a better physical view of the various problems considered. This law has not, of course, been experimentally verified for unclosed circuits; but the self inductance of an unclosed circuit means simply its self-inductance as a part of a closed circuit, the total inductance of which cannot be determined until the entire circuit is specified. In this sense the use of the law of Biot and Savart to obtain the self-inductance of an unclosed circuit is perfectly legitimate."

One might question here what the word 'legitimate'. Rosa has made an assumption in order to obtain a result. That is fine, but I don't think you should say it is legitimate, in the sense of being according to law. There is no law to guide this assumption. In any event let us proceed to the solution. The approach is the following.

1. Set up the geometry of a point at some distance from a straight line.

2. Write dH from Biot Savart at that point. Let the current equal one for simplicity.

3. Find the total H due to all of the current in the line by integrating over the line.

4. Set B = uH.

5. Find the magnetic flux phi in a differential area which is parallel to the wire by integrating B over this area which is a fixed distance from the wire.

6. Find the total flux phi over all of the area from the edge of the wire to infinity by integrating over the distance from the wire to infinity.

7. Since the current was set equal to one, the total flux equals the inductance.

I am not going to put in the full derivation until I find a more math-friendly language than HTML. You can get the full derivation by contacting me (thealy@scu.edu), or by looking up the original Rosa reference below.

The total low frequency inductance (internal plus external) of a straight wire is:

$$Ldc = 2L[\ln(2L/r) - 0.75]nH$$

where Ldc is the "low-frequency" or DC inductance in nanohenries (nH), L is the length of the wire in cm, and r is the radius of the wire in cm.

This result is based on the assumption that the radius r is much less than the length L, which is commonly true.

For sufficiently high frequencies skin effects cause the internal inductance to go to zero and the inductance becomes:

$$Lac = 2L[\ln(2L/r) - 1.00]$$

E.B. Rosa, "The Self and Mutual Inductances of Linear Conductors", Bulletin of the Bureau of Standards, Vol.4, No.2, 1908, Page 301ff.

Self resonant frequency of a capacitor

Self-resonant frequency of a capacitor

Nonsense about so-called "self-resonant frequency" of a capacitor.

If you cut off a capacitor's legs at the knees, you will double its self-resonant frequency – Ivor Catt

Martin Eccles takes the biscuit

My 1994 book *Electromagnetism 1* is at http://www.ivorcatt.com/em.htm

Nigel Cook on Ivor Catt's ideas, (London) Electronics World (was Wireless World), aug02, pp46-49

More nonsense is at http://www.ivorcatt.com/2605.htm

Yet more nonsense

Yet more nonsense

Scandals in Electromagnetic Theory http://www.ivorcatt.com/28scan.htm

RIPOSTE Ivor Catt's view of Capacitors, by Leslie Green CEng MIEE

Discussion of Riposte

http://www.atceramics.com/pdf/technotes/effective_capacitance_vs_fr.pdf talks about self-resonant frequency. The "parasitic inductance" does not exist inside the capacitor. However, the article is very primitive, because the graphs go in the wrong direction, with C increasing with frequency, anyway. The alleged L reduces the C, not increasing it as the author seems to think.

http://www.capacitors.com/multicap/phase-esr/phase-esr.html

"the capacitor can be used up to the natural self-resonant frequency or"

http://www.educatorscorner.com/experiments/pdfs/exp79.pdf
Unfortunate students are made to measure the “self-resonant frequency” of a capacitor. Tell them to cut off the poor capacitor’s legs!

http://www.capacitors.com/consider/consider.htm
This guy is stumbling in the right direction, but he gets his number wrong. Inductance caused by the legs should be proportional to the length of the legs.

“Lead length alters a capacitor's range of operating frequency. Here a 2 uf capacitor's self-resonance decreases from 490 kHz to 290 kHz when its leads are lengthened from 3/8 inch to 3 inches. In other words, the capacitor's usable operating range is reduced by almost half.”

More nonsense;

http://www.qsl.net/kf4trd/varactor.htm
your poor capacitor behaves just like an inductor! The frequency at which both impedances
are equal is known as the self-resonant frequency. This frequency is set by the materials used and the construction of the capacitor.

http://www.emf-emi.com/dosanddonts.shtm
Ensure that the SRF (self-resonant frequency) of capacitors is above the highest frequency to be bypassed.
…. 10.) Select and mount decoupling capacitors having self-resonance (SRF) above logic band-width (1/)

http://www.cypress.com/pub/appnotes/decouple.pdf
This writer is in a bad way. Pure fantasy. – IC
What a pity this poor fellow has been pulled from the www by his paymasters. The hyperlink now jumps straight to another page, bypassing the nonsense I saw on

30jan02. Perhaps I have influence! Ivor Catt. 5may02

http://www.benchmarkmedia.com/appnotes-a/caig/caig06.asp
However, all capacitors have their own self-resonant frequency

To avoid series inductance, a well behaved capacitor keeps its legs together – Ivor Catt

@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@

The amount of nonsense drifting around the world, of which the above are examples, is vast.
See my 1978 article at http://www.electromagnetism.demon.co.uk/z001.htm ; "Series inductance does not exist. *Pace* the many documented values for series inductance in a capacitor, this confirms experience that when the so-called series inductance of a capacitor is measured it turns out to be no more than the series inductance of the wires connected to the capacitor. No mechanism has ever been proposed for an internal series inductance in a capacitor."
The key point in my article is that "No mechanism has ever been proposed for an internal series inductance in a capacitor."
The IEE and IEEE have helped to cause the confusion to escalate by suppressing my 1978 article http://www.electromagnetism.demon.co.uk/z001.htm , which puts an end to a capacitor's series inductance. Also, competent experimentation will show that a capacitor has no internal series inductance. http://www.ivorcatt.com/em_test04.htm

– Ivor Catt, 30jan02

@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@

Ivor Catt 22apr02

In 1963 I bought the EH-125 pulse generator. This delivered a –10v step with a 100picosecond fall time into a 50 ohm load (e.g. 50 ohm

coax.).

The pulse generator could also deliver a –ve 10v spike with a width of 150psec. I decided to try to create a positive 10v spike. I cut into the 50 ohm coax, and joined the incoming inner to the outgoing outer via a red 1uF tantalum capacitor. I also joined the incoming outer to the outgoing inner via another 1uF tantalum capacitor. Further downstream I found that I had a positive 150psec spike with no discernable degradation (in rise time or pulse width) compared with the initial –ve spike. That is, I had a +ve 10v spike with a width of 150psec.

It is interesting to calculate the physical width of a 150 psec wide spike travelling down normal coax, which has a dielectric with a dielectric constant of 2. Whereas light travels one foot in vacuo in one nsec, it would travel 8 inches in material with a dielectric constant of 2. Thus, a 150psec spike in the coax has a width of about one inch. So I sent a TEM spike with a width of 1 inch through these 1uF capacitors. [Note 1] Obviously, I kept their legs short. It is sad that during the ensuing 40 years the New York IEEE and the London IEE prevented me from informing electronic engineers that they did not have to add “high frequency” decoupling capacitors to their logic boards, that the 1uF would do perfectly well on its own. This obstruction has cost the industry many millions of pounds. However, a bolshie IEEE and a bolshie IEE cost us a lot more than that in other ways. Ivor Catt 22apr02

Note 1.
Anyone who wants to play with frequencies can be told that the fundamental of the 150psec spike will be around 3GHz. Put that in your “self-resonant” pipe and smoke it! IC

Note 2.
As the spike passes the capacitors placed to each side, the situation is as in http://www.ivorcatt.com/2_1.htm Figure 14. The characteristic impedance of each capacitor is very small, less than 1% of 50 ohms.

Thus, the mismatch is less than 2%, causing a minimal reflection of less than 1%.
At the same time, if the legs of the capacitors are kept down to a total of one quarter of an inch in length, and the two parallel legs represent a quarter inch transmission line of characteristic impedance 150 ohms, then the mismatch will cause a reflection of 50%, see http://www.ivorcatt.com/1_4.htm Figure 11 and the reflection formula. This will be reduced by the fact that the 150psec spike covers a distance of one inch and a half, so that the reflections on entering the 150 ohms region tends to be masked by the opposite mismatch on re-entering the 50 ohm impedance of the next section of coax. This reduces the reflection to one sixth, i.e. 8%.

@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@

18may02 More drivel. Fig. 2 at http://www.ultracad.com/seminar_caps.htm

Google Hit no. 7 for "self resonant frequency" + capacitor

This article high on the Google hit list has row of capacitors, and each one decouples (digital electronic equipment for) its particular frequency range.
This farce is obvious if one realises that a 2uF capacitor is made by glueing together two 1uF capacitors. Thus, a supposedly "high frequency capacitor" is merely the front little bit of a 1uF capacitor. Of course, you can ruin the performance of either by leaving it with long legs, making a series one-turn inductor to stifle its performance. However, the idea that a 10pF capacitor has shorter legs than a 1uF capacitor is based on nothing at all.

What is so tragic is that the formula these clods use for self resonant frequency, 1/ sqrt LC , means that if C is big, then the resulting calculated "self resonant frequency" is low. This is a sensible idea if a resonant circuit is being designed out of a discrete C and a discrete L,

where L can be varied. But if, as in our case, we (have legs of fixed length and) can only vary the value of C, then the calculation deludes. If we start with a pair of legs of fixed length, that is, with a fixed external L, then the bigger the C, the lower the resonant frequency according to the formula w = 1/ sqrt LC. These buffoons are buying capacitors for the very reason that they have less capacitance, not that they have less L. They buy these “high frequency capacitors” for the very reason that, lacking much C but helping the w = 1/ sqrt LC formula, they are inferior at doing the decoupling job that they have been bought to do. All their nonsense is counter-productive. This was pointed out in my book more than 20 years ago. Since digital electronics took over from radio as the majority of electronic engineering 40 years ago, it is high time the radio men gave us at least some access to the IEEE, the London IEE and to Cambridge and MIT. Even a single digital electronics course by someone who understood the subject at either MIT or Cambridge University would help a lot. I would love to give it. However, I am sure the radio men will continue to shut me and my colleagues out, as they have done for a number of decades, hoping that their antique radio theory will continue to appear to address the needs of digital electronic equipment. http://www.ivorcatt.com/em_test04.htm Ivor Catt 18may02

Since a capacitor is a two-conductor transmission line with very low characteristic impedance, the transient impedance that it presents to a step is resistive, not reactive. This is the way it behaves when decoupling digital circuits; as a local energy store for the 5v supply with a very low resistive source impedance, not a reactive source impedance. Calculation of the impedance is made by using the normal formula for the characteristic impedance of a transmission line made up of two parallel plates with width a and separation b. See p73 of my book “Electromagnetics 1”, pub. Westfields 1975. The (resistive) impedance is very low because the dielectric constant is very high indeed, and the separation b is tiny. Ivor Catt 18may02

In the surreal world created with inappropriate mathematical stunts by physically ignorant operators, a capacitor is looked on with disdain, not because it has more L, but because it has more C.

http://www.ivorcatt.com/em_test04.htm Ivor Catt 18may02

@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@

Recap. Take the formula for the resonant frequency for an inductor-capacitor tank circuit.

The frequency (in radians per sec.) squared equals (1/ inductance x capacitance)

Thus, either increase in inductance or increase in capacitance reduces the resonant frequency. This has led physically ignorant mathematical mugwumps to think, not that the best capacitor has the least capacitance, which even they might realise is ridiculous, but that the best capacitor has the least inductance, making it able to perform to a much higher frequency up to its higher resonant frequency. They have failed to realise that they would realise their dream, of a high self resonant frequency, by reducing the capacitance just as well as by reducing the inductance. They think that it is an accident that lo value capacitors have the highest self resonant frequency. They think it is because of the difference in inductance, which it is not.

However, all this is nonsense when decoupling digital logic. What matters with digital logic is the transient performance of a decoupling capacitor, when some switching logic wants to grab as much charge as possible to launch down a transmission line towards the next logic gate. The true model, which should have replaced the series L C R model for a capacitor, was already published in 1978, http://www.electromagnetism.demon.co.uk/z001.htm , and has been ignored for 24 years by radio men who continue to teach and publish the old model which is inappropriate and damaging in digital electronics. Note that today, most capacitors are used in DC voltage decoupling.

The only way out of this impasse is for students to create problems during the lecture when lecturers continue to pump out the old, wrong drivel. Otherwise these lecturers and text book writers will continue to copy and repeat each other from a bygone age when electronics was about radio, and such a misconception about the physical nature of a capacitor was not so damaging. http://www.ivorcatt.com/em_test04.htm Students have much to gain by disrupting their lectures. It is probably more difficult to learn and be examined in material which is false. Ivor Catt. 18may02.

@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@

In 1965, living in the USA, I telephoned the design engineers in Sprague, who manufactured capacitors. They told me that they tested for the high frequency performance of a capacitor by testing at 5kHz and 50kHz, and deduced its performance at 1MHz and above using the series L C R model. Thus, the published self-resonant frequency of a capacitor is the result of lo frequency testing extrapolated using the L C R model.

By making this error, engineers in the capacitor manufacturers might have doubled their companies' sales, ensuring that a second "high frequency" capacitor would be added to every 1uF decoupling capacitor

in every digital system. Ivor Catt 18may02.

@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@

Academic apes

xx

Riposte
I make the commitment that anyone wishing to counter any assertion made on this site will be guaranteed a hyperlink to a website of their choosing at the point where the disputed assertion is made. ivor@ivorcatt.com
Ivor Catt. 18june02

Scandals in electromagnetic theory http://www.ivorcatt.com/28scan.htm

x

(Possibly we need a standard word for this. I suggest "Riposte", or the symbol [R] .) Ivor Catt, 30june02. ivor@ivorcatt.com

-

National Semiconductor
The Sight & Sound of Information
EDN Innovation 2005 WINNER
Golden Mousetrap Winner
analogZONE Product of the Year
Sign-On | 日本語 | 简体中文 | 繁體中文 | 한국어

Search | Order Parts | Cross-Ref

Home | About Us | Investors | Press Room | Private Sites

THE BEST OF BOB PEASE

Understand Capacitor Soakage to Optimize Analog Systems

Dielectric absorption can cause subtle errors in analog applications such as those employing S/H circuits, integrating ADCs and active filters. But knowing how to measure this soakage and compensate for it helps you minimize its effects.

Veteran circuit designers often got a shocking introduction to dielectric absorption when supposedly discharged high-voltage oil-filled paper capacitors reached out and bit them. Indeed, the old oil-filled paper capacitors were notorious for what was once called soakage -- a capacitor's propensity to regain some charge after removal of a momentary short. Today, you won't find very many of these capacitors in use, but you will still encounter soakage. Do you know how to deal with it?

Nowadays, you're more likely to notice the effects of dielectric absorption in some more subtle way, perhaps in the performance of an integrator that can't be reset to zero or a sample/hold that refuses to work correctly. But whether you literally feel its effects or merely observe them in a circuit's behavior, dielectric absorption is an undesirable characteristic that every capacitor possesses to some degree. This characteristic is inherent in the dielectric material itself, although a poor manufacturing procedure or inferior foil electrodes can contribute to the problem.

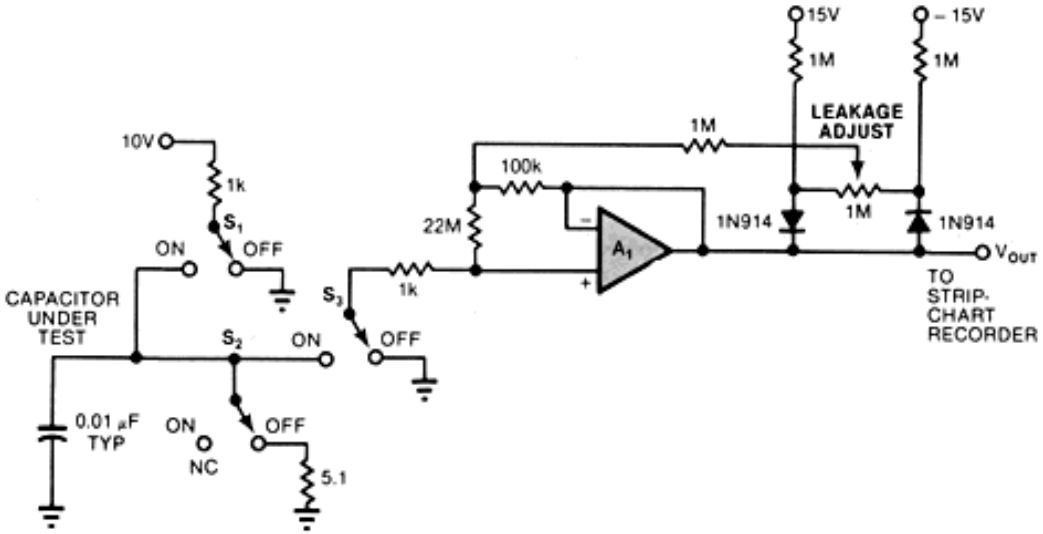


Fig 1 - A simple test fixture lets you evaluate dielectric absorption at low speeds. To use the one shown here, start with all switches off and throw S_1 and S_2 on for 1 min; throw S_1 and S_2 off and wait 6 sec, throwing S_3 on during the wait period. Next, turn S_2 on and watch V_{OUT} for 1 min. To compensate for leakage, leave all switches off for 1 min and then throw S_2 and S_3 on. Monitor V_{OUT} for 1 min and subtract this value from the V_{OUT} value obtained earlier. (View a larger version of the image.)

Indeed, soakage seems an apt term for dielectric absorption when you note what the capacitor seems to be doing. Consider a typical example: A capacitor charges to 10V for a long time T and then discharges through a small-value resistor for a short time t. If you remove the short circuit and monitor the capacitor terminals with a high-impedance voltmeter, you see the capacitor charge back to 0.1%, 1% or as much as 10% of the original voltage. For example, a 1-µF Mylar capacitor charged to 10V for 60 sec (T_{CHARGE}) and discharged for 6 sec ($T_{DISCHARGE}$) charges to 20 or 30 mV after 1 min (T_{HOLD}). Fig 1 shows a simple evaluation circuit for measuring this characteristic.

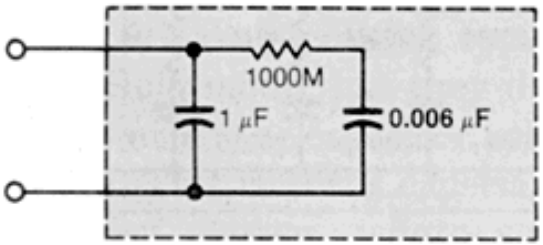


Fig 2 - To model the soakage characteristic of a 1-µF Mylar capacitor, consider a circuit that incorporates a 0.006-µF capacitor to represent the dielectric's charge-storage characteristics.

A capacitor exhibiting dielectric absorption acts as if during its long precharge time the dielectric material has soaked up some charge that remains in the dielectric during the brief discharge period. This charge then bleeds back out of the dielectric during the relaxation period and causes a voltage to appear at the capacitor terminals. Fig 2 depicts a simple model of this capacitor: When 10V is applied for 1 min, the 0.006-µF capacitor gets almost completely charged, but during a 6-sec discharge period it only partially discharges. Then, over the next minute, the charge flows back out of the 0.006-µF and charges the 1-µF capacitor to a couple of dozen millivolts. This example indicates that a longer discharging time reduces soakage error but that discharging for only a small fraction of that time results in a larger error. Illustrating this point, Fig 3 shows the results of conducting Fig 1's basic test sequence for 1-, 6- and 12-sec discharge times. Note that the capacitor tries to remember its old voltage, but the longer you hold it at its new voltage, the better it forgets -- in the Fig 3 case, soakage errors equal 31 mV at $t_{DISCHARGE}$=l sec, 20 mV at $t_{DISCHARGE}$=6 see and 14 mV at $t_{DISCHARGE}$=12 sec.

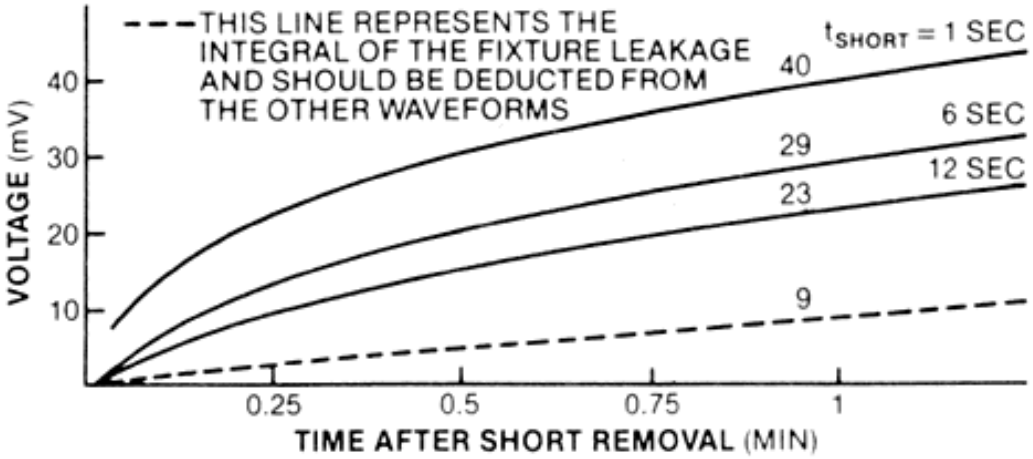


Fig 3 - Obtained using Fig 1's test circuit, these dielectric-absorption-measurement results for a 1-µF capacitor shown that longer $t_{DISCHARGE}$ times reduce soakage-caused errors.

High-speed tests predict S/H performance

You might now ask whether these low-speed tests have any bearing on a capacitor's suitability in fast millisecond or microsecond sample/hold applications. If you repeat the Fig 1 experiment for T_{CHARGE} = T_{HOLD} = 1000 µsec and $t_{DISCHARGE}$ = 100 µsec, you see very similar capacitor-voltage waveforms but with about 10-times-smaller amplitudes. In fact, for a constant T:t ratio, the resulting soakage error decreases only slightly in tests ranging in length from minutes to microseconds.

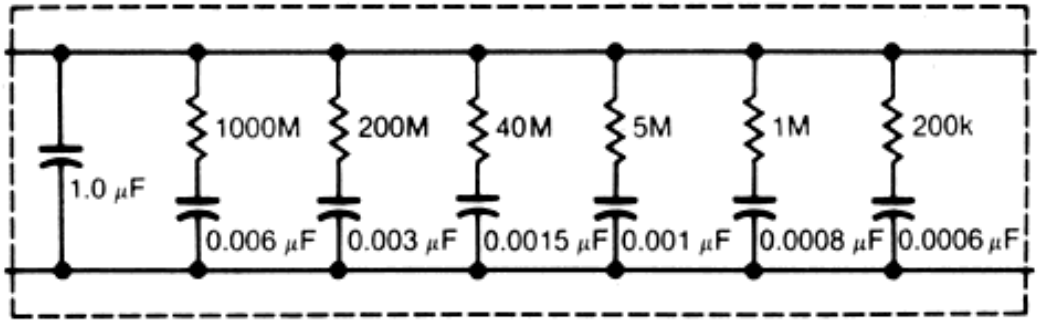


Fig 4 - More precise than Fig 2's equivalent circuit, a capacitor model employing several time constants proves valid for a wide range of charge and discharge times. This model approximates a Mylar capacitor.

Fig 4's circuit approximates this capacitor characteristic, which you can observe on actual capacitors by using Fig 5's test setup. Here, a sample/hold IC exercises the capacitor under test at various speeds and duty cycles, and a limiter amplifier facilitates close study of the small residual waveforms, without over-driving the oscilloscope when the capacitor is charged to full voltage.

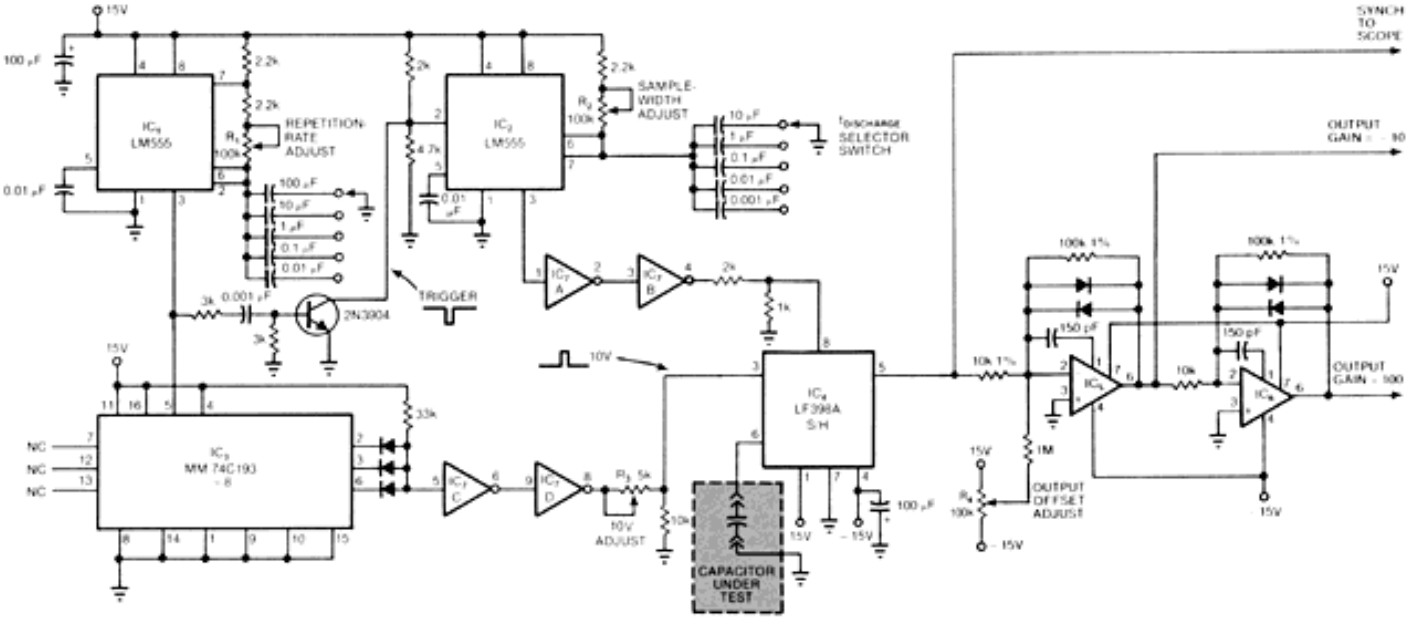


Fig 5 - Capable of automatically sequencing the dielectric-absorption tests, a circuit employing timers, a sample/hold and limiting stages allows you to make measurements for a wide range of T_{CHARGE}, T_{HOLD}, and $t_{DISCHARGE}$ values. Fig 7 shows the results obtained using the circuit shown here. (View a larger version of the image.)

Notes:

1. ALL DIODES = 1N914
2. IC5, IC6 = LM301A
3. IC7 = MM74C04
4. USE R4 OR -10 GAIN TO KEEP SCOPE WAVEFORM BELOW 200mV SO AS TO AVOID DISTORTION OR FALSE ATTENUATIONS

Such experiments illustrate that if you put a certain amount of charge into a less-than-ideal capacitor, you will get out a different amount of charge, depending on how long you wait. Thus, using low-soakage capacitors proves important in applications such as those involving high-resolution dual-slope integrating ADCs. And sure enough, many top-of-the-line digital voltmeters do use polypropylene (a low-soakage dielectric) devices for their main integrating capacitors.

But dielectric-absorption characteristics are most obviously detrimental in applications involving sample/holds. Manufacturers guarantee how fast these devices can charge a capacitor in their Sample mode and how much their circuits' leakage causes capacitor-voltage droop during the Hold mode, but they don't give any warning about how much the capacitor voltage changes because of soakage. This factor is especially important in a data-acquisition system, where some channels might handle small voltages while others operate near full scale. Even with a good dielectric, a sample/hold can hurt your accuracy, especially if the sample time is a small fraction of T_{HOLD}. For example, although a good polypropylene device can have only 1-mV hysteresis per 10V step if T/t=100 msec/10 msec, this figure increases to 6 mV if the T/t ratio equals 100 msec/0.5 msec. Because most sample/hold data sheets don't warn you of such factors, you should evaluate capacitors in a circuit such as Fig 5's, using time scaling suited to your application.

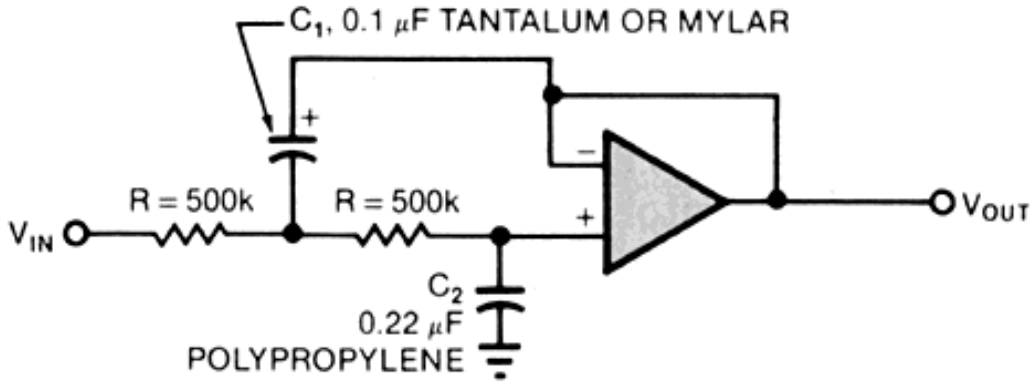


Fig 6 - Soakage can present problems when you're designing a fast-settling amplifier or filter. In the circuit shown here, for example, C_1 can be a Mylar or tantalum unit, but making C_2 a polypropylene device improves performance.

Other applications in which soakage can degrade performance are those involving fast-settling ac active filters or ac-coupled amplifiers. In Fig 6's circuit, C_1 can be a Mylar or tantalum unit because it always has 0V dc on it, but making C_2 polypropylene instead of Mylar noticeably improves settling. For example, settling to within -0.2 mV for a 10V step improves from 10 to 1.6 sec with the elimination of Mylar's dielectric absorption. Similarly, voltage-to-frequency converters benefit from low-soakage timing capacitors, which improve V/F linearity.

Some dielectrics are excellent at all speeds

Fortunately, good capacitors such as those employing polystyrene, polypropylene, NPO ceramic and Teflon dielectrics perform well at all speeds. Fig 7 shows the characteristics of capacitors using these dielectrics and others such as silver mica and Mylar. In general, polystyrene, polypropylene or NPO-ceramic capacitors furnish good performance, although polystyrene can't be used at temperatures greater than 80°C. And although NPO ceramic capacitors are expensive and hard to find in values much larger than 0.01 µF, they do achieve a low temperature coefficient (a spec not usually significant for a S/H but one that might prove advantageous for precision integrators or voltage-to-frequency converters). Teflon is rather expensive but definitely the best material to use when high performance is important. Furthermore, only Teflon and NPO ceramic capacitors suit use at 125°C.

Fig 7 - Soakage-measurement results for a variety of capacitors illustrate the effects of $t_{DISCHARGE}$ values on dielectric-absorption-caused errors. Note the curves for two different samples of NP0 ceramic capacitors intersect.

If you look at Fig 7's dielectric-absorption values, you can see wide differences in performance

for a given dielectric material. For example, polypropylene sample A is about as good as B at t=6 see, but B is four times better at high speeds. Similarly, NP0-ceramic sample A is slightly worse than NP0-ceramic sample B at low speeds, but A is definitely better at high speeds. And some Mylar capacitors (sample A) get better as speed increases from 1000 to 100 µsec, but others (sample B) get worse. So if you want consistently good performance from your capacitors, evaluate and specify them for the speed at which they'll be used in your application. Keep in mind that because most sample/holds are used at much faster speeds than those corresponding to the 1- or 5-min ratings usually given in data sheets, a published specification for dielectric absorption has limited value.

In addition, other dielectrics furnish various levels of performance:

- Because any long word that starts with poly seems to have good dielectric properties, how about polycarbonate or polysulfone? No -- they are about as bad as Mylar.
- Does an air or vacuum capacitor have low soakage? Well, it might, but many standard capacitors of this type are old designs with ceramic spacers, and they might give poor results because of the ceramic's hysteresis.
- If a ceramic capacitor is not an NP0 device, is it any good? Most of the conventional high-K ceramics are just terrible -- 20 to 1000 times worse than NP0 and even worse than tantalum.
- Is silicon dioxide suitable for small capacitances? Although Fig 5's test setup, used in preparing Fig 7's chart, only measures moderate capacitances (500 to 200,000 pF), silicon dioxide appears suitable for the small capacitors needed for fast S/Hs or deglitchers.

Cancellation circuit improves accuracy

A practical method of getting good performance with less-than-perfect capacitors is to use a soakage-cancellation circuit such as one of the form shown in Fig 8, in which a capacitor of the type modeled in Fig 4 serves as an integrator. (Only the first two soakage elements are shown.) The integrator's output is inverted with a scale factor of -0.1, and this voltage is then fed through one or more experimentally chosen RC networks to cancel the equivalent network inherent in the capacitor's dielectric material.

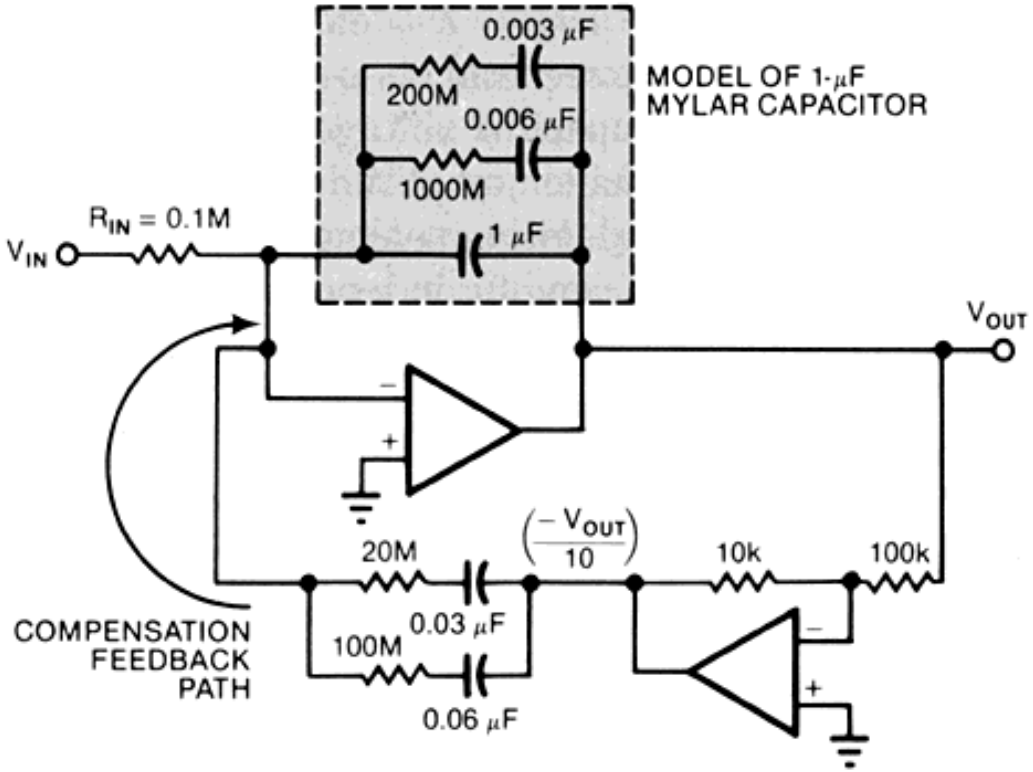


Fig 8 - You can compensate an integrator for dielectric absorption by feeding its inverted output back to the input through one or more experimentally chosen RC networks, which cancel the equivalent network inherent in the capacitor's dielectric material.

Fig 9 shows a practical sample/hold circuit with an easily trimmed compensator. This network provides about a 10-fold improvement for sample times in the 50- to 2000-µsec range (Fig 10). Although this compensation is subject to limitations at very fast or slow speeds, the number of RC sections and trimming pots employed can be extended.

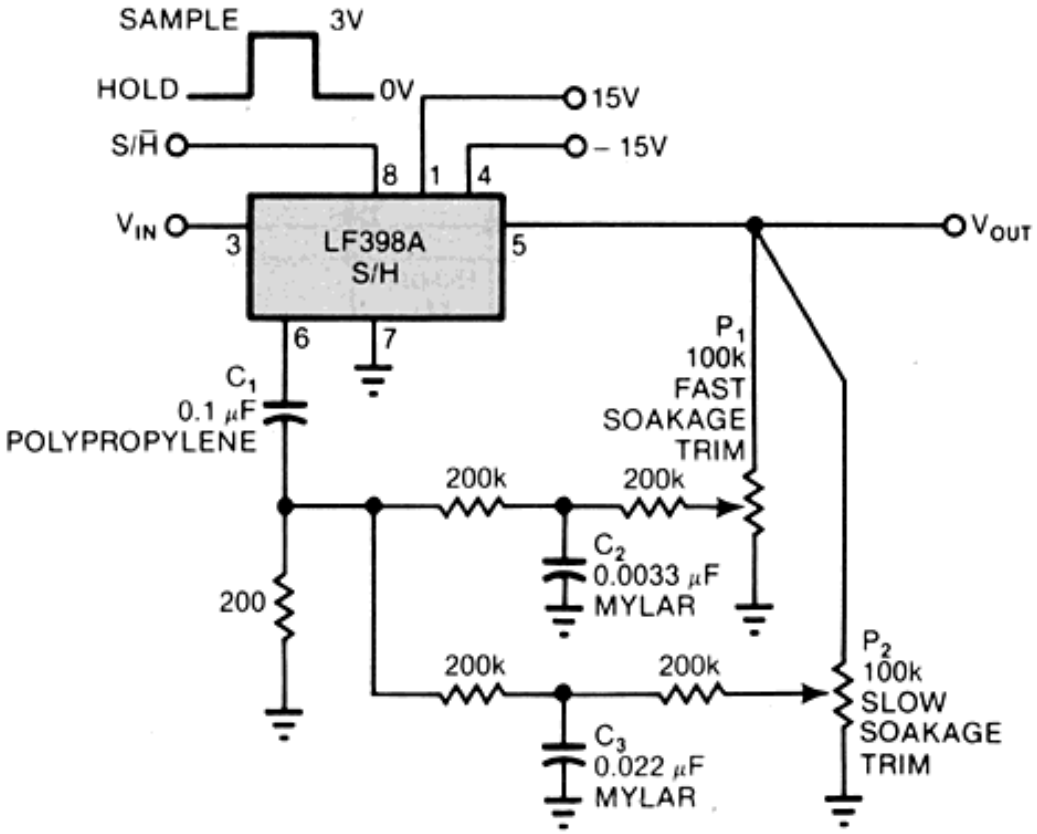


Fig 9 - Adding compensation circuitry to a sample/hold yields better-than-Teflon performance with a polypropylene capacitor. Using Teflon capacitors in such circuits can yield a 15- to 17-bit dynamic range.

Simple circuits similar to Fig 9's or Fig 8's have been used in production to let inexpensive polypropylene capacitors provide better-than-Teflon performance. In turn, using these compensator circuits with a good Teflon capacitor furnishes a dynamic range of 15 to 17 bits.

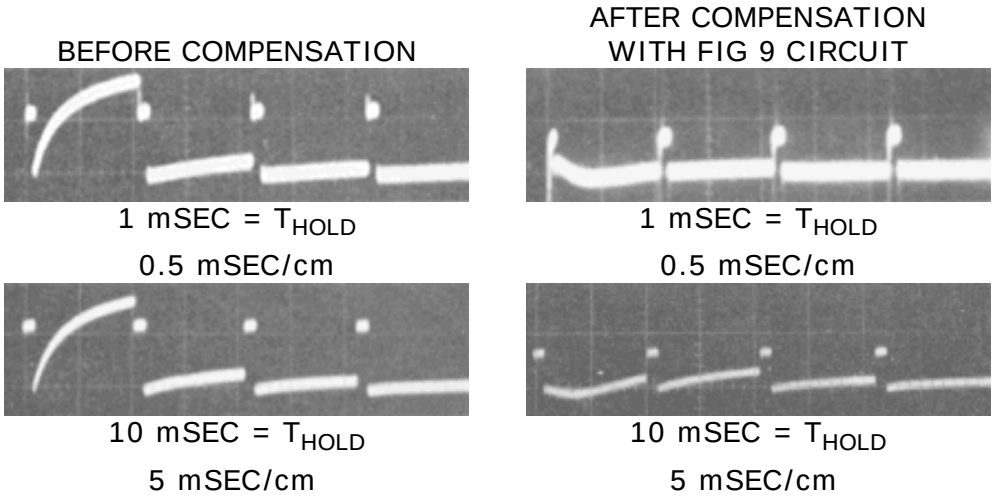


Fig 10 - Adding Fig 9's compensation network to a sample/hold circuit yields a 10-fold performance improvement for sample times of 50 to 2000 µsec; additional RC networks and trimming pots can extend the time range. The short pulses represent normal S/H jumps and occur during the sample time. The exponentially rising waveform during the hold time results from soakage. Note that soakage effects are still visible during the second hold period.

Notes:

1. Dielectric absorption errors with good polypropylene capacitor
2. All waveforms at 1mV/cm; T_{SAMPLE}/T_{HOLD} = 1/10

Originally published in EDN,
October 13, 1982.

RAP Update: *FIFTEEN years ago, back in 1982, when I wrote this, I had never seen any study of capacitors and their "soakage" -- nor of the kind of circuits you could use to shrug off the effects of soakage. Nobody ever talked about this, at high speed.*

To this day, I have not seen any other articles that covered either subject. So this is still about the prime source of info on how to evaluate capacitors for soakage, AND how to build good Sample-and-Hold circuits, so as to NOT get hurt by that soakage.

Your comments are invited! / RAP

Site Map • About "Cookies" • National is ISO/TS16949 Certified • Privacy/Security Statement • Contact Us/Feedback
Site Terms & Conditions of Use • Copyright 2006© National Semiconductor Corporation

CapSite 2006
Aging
Film
Ceramic
Electrolytic
Misc. Dielectrics
Construction
ESR
Markings
Reliability
Line-Filter
Insulation Resis.
More Information
Obscure Caps
Dielectric Const.
Cap Model
Purchasing

CapSite 2006

Introduction To Capacitors

release 3.7

"...has many omissions and contains much that is apocryphal, or at least wildly inaccurate..."
Douglas Adams
The Hitchhiker's Guide to the Galaxy

This site started out as a simple exercise in web page making that sort of got out of control. Once I started writing, I found it hard to stop.

A few disclaimers: Some of the information in this "FAQ" should be considered to be generalities. Material source, processing and capacitor construction can greatly affect a capacitor's electrical properties. Different manufacturers tend to have different opinions on the characteristics of their dielectric materials. For example, most show polystyrene as having a slight negative temperature slope while at least two show it as being slightly positive. When in doubt, contact the manufacturer to guarantee that the parameter you are interested in is good enough for your needs. This is especially true when using the various graphs. These should be used more for comparing different capacitor types than for hard design data. That is best obtained from individual manufacturers. In some cases, construction variations so distort the basic capabilities of the material that it seemed best not to include graphical data. When using charts of dielectic characteristics, it is best to look at a number of sources. One chart for example, shows the dielectric absorption of mica to be around 0.003%, another shows it to be more like 0.5%. They can´t both be right (.5% is probably closer).

Another disclaimer: If the graphs sometimes look a little odd (less than optimal colors for example) blame Microsoft. Translating Excel 2000 graphs to GIFs does not seem to work as well as with earlier versions of Excel. I´ll look into this someday. Drawings made in TurboCAD were another problem. No program I have can both import them correctly, and correctly export them to GIF format. The next version may correct this.

Much of the available quantitative information of capacitors is not very complete. Dissipation factor vs. temperature is almost always shown for 1 kHz. DF vs. frequency is normally shown for 20C or 25C. If you want to know the DF of polyester at 95C @ 20 kHz, you will probably be out of luck. Certainly the manufacturers have some of this sort of information (the good ones anyhow), but it is rarely published. It´s the rare hobbyist who needs to know that sort of thing however. If you are looking for information on just the dielectric, as opposed to a finished capacitor, the material manufacturers often have very complete information.

On the other hand, as consumer electronics moves into GHz frequencies, designers need information on dielectric behavior that may not yet be available.

Some capacitor parameters are difficult to measure with good absolute accuracy. An instrument for measuring dissipation factor for example, may give good relative results for comparing different capacitors, but not agree with another brand of instrument on the actual numbers.

The information given here is as good as I can make it, which probably means it has its share of errors. There are also many omissions; my main interest is in analog electronic applications so that is how this FAQ is oriented. There is some mention of high-power capacitors, but sort of thing is too specialized for good coverage here. This is meant as an introduction to capacitors, not a book. This FAQ is aimed more at students, hobbyists, scientists, and new engineers than at electronics "insiders". It does not (yet) contain a dictionary of terms, that has been done by other people. Not covered are some specialized types like feed-through capacitors and trimmers. In any case, there are references to other sources for some of the things I omitted.

This is also a good place to note that some limitations mentioned, like size, voltage ratings, packaging, temperature

ratings, etc., will usually only apply to high-volume capacitor manufacturers. There are specialty manufacturers who can make most anything you can think of.

What is a Capacitor?

A capacitor is a energy storing device made up of two parallel conductive plates separated by an insulating (dielectric) material. When a voltage is applied across the plates, the electric field in the dielectric displaces electric charges, and thus stores energy. It is assumed that there are no free charges in the dielectric (at least in the ideal case), and that while they are displaced, they are not free to move around as in a conductor. The closest analogy in the mechanical world is probably energy stored by a spring. Dielectrics come in two types, "polar" and "nonpolar". Molecules where the "center-of-gravity" (as it were) of the negative and positive charges are at the same point are nonpolar. If this isn´t the case, the molecule is polar. H_2O, for example is polar, but H_2 is nonpolar.

The plates may be actual metal plates of various shapes but are most often in the form of metal foil or a metal film deposited on the insulating material. Since the first capacitor was invented, the Leyden jar, almost every conceivable dielectric material and form has been tried by someone. Wax, paper, plastics, ceramics, glass, oils, minerals, electrochemical films, air, etc., either alone or in combination such as paper/wax, paper/epoxy, plastic/plastic, paper/oil, plastic/oil have been used, to name just a few. Some people seperate capacitors types into "electrostatic" and electrolytic". This is really an artificial distinction, more of a construction detail than a fundamental difference.

Characteristics

The reason for the many types of capacitors is cost and performance. Different capacitor types are better at different things. No capacitor is just pure capacitance and there is no perfect capacitor at the perfect price. Every capacitor type has a number of parasitic elements associated with it such as leakage, ESL, and ESR, plus special quirks all its own. This makes it tricky to select just the right capacitor for every application.

Factors of concern to the user include (but are not limited to):

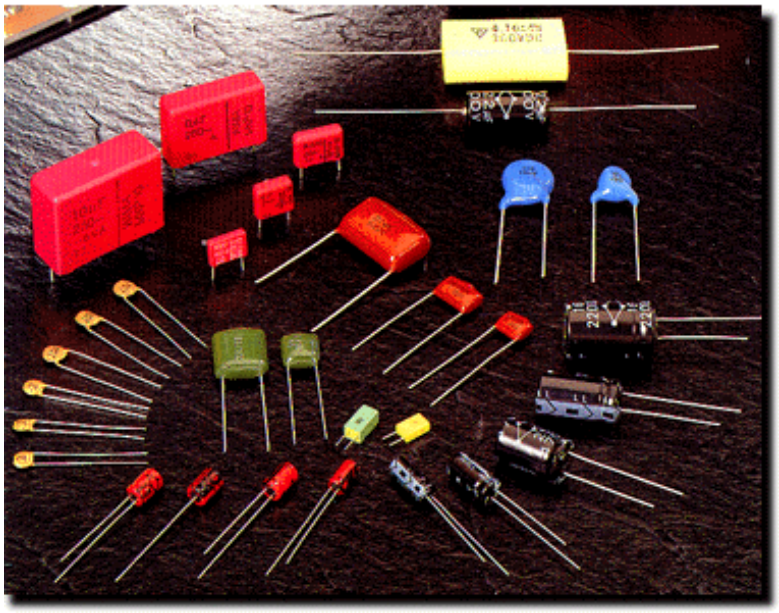
- Aging rate, mostly of concern for ceramic capacitors although all capacitors age, however little. Film capacitors typically age less than 1% over their life, while some ceramics may age >10%. With films at least, manufacturers tend to disagree over exact numbers. This is probably due to differences in test methods.
- **Capacitance vs. frequency**, determined by both the dielectric and construction details.
- **Capacitance vs. size** (or "volumetric efficiency") . Mainly determined by the dielectric constant (K) (sometimes called "permittivity") of the dielectric and its voltage breakdown vs. thickness. For a given dielectric, size tends to vary as the square of the voltage rating. This is a crude approximation at best. It assumes that the dielectric makes up most of the volume of the capacitor. In some capacitors, the packaging and associated materials actually makes up most of the volume. It also assumes that the breakdown voltage increases linearly with the thickness of the dielectric, also a dubious assumption.
- Capacitance Drift vs. Temperature, normally given as ppm/deg C. Every dielectric has a different temperature slope. This can be as low as +- 30 ppm/C for C0G ceramics (even +-15 ppm/C for special porcelain RF capacitors) to thousands of ppm/C for electrolytics and some plastic films. An old trick for getting reduced drift is to parallel two capacitors which have opposite temperature drifts, such as polyester and polypropylene, or polystyrene and polycarbonate. The temperature drift for a given dielectric can vary from one manufacturer to another however, making this technique uncertain. Polycarbonate for example, is shown by some manufacturers to have a slightly positive slope, and slightly negative by others. Caps with mixed dielectrics exist, but are not common.
- **Capacitance vs. voltage**, mostly of interest for Class 2-4 ceramic capacitors, and for some glass types.
- **Cost and availability.**
- **Current-carrying capacity**, of concern when the capacitor must handle high alternating current which would cause self heating. This a problem for electrolytics, and for film and ceramic capacitors used in high frequency and/or high current applications. Manufacturors often have guidelines.
- **Dielectric absorption** (or DA), sometimes called "soakage", "voltage retention", and other things. In high-voltage power cables it´s called "return voltage". A capacitor, once charged, stubbornly retains part of the charge, even after being discharged (shorted for some number of seconds), as if it had "soaked" into the dielectric. DA is modeled as an infinite series of RC networks in parallel with the primary capacitance, where the Rs are mostly very large and the Cs are smaller than the nominal capacitance. DA is a source of error in precision integrators and sample-and-hold circuits. Large high-voltage oil-filled capacitors (exceptionally high DA) can be shorted, yet retain enough charge to be dangerous. They are normally shipped with a short across the terminals that should not be removed until the part has been installed. High-voltage capacitors in TV sets are also still dangerous after being discharged, as is the parasitic capacitance in large oil-filled, high-voltage transformers and some high-voltage cables. There is a "standard" test for

DA (actually, everyone seems to do it a little differently). The capacitor is charged to some voltage for one minute (often 100 volts, but others can be used), and is then shorted for two seconds. After one minute, the recovered voltage on the capacitor is read using a very high impedance meter. The DA is expressed as the ratio between the recovered voltage and the charging voltage, in percent. MIL-C-19978D calls for a 5 minute/5 second/1 minute sequence. These tests are useful for comparing dielectrics, but don't really tell you just how a capacitor will perform in your application. In particular, the difference in DA of various dielectrics as seen one minute after the short is removed does not necessarily precisely correlate with the difference in DA as seen in the first few 10s of ms. In general, Teflon, polystyrene, and polypropylene are the best (as low as 0.02%), while the electrolytics, high-k ceramics, and oil-filled are the worst (1% on up). Impregnated film caps tend to have a DA that reflects the impregnant more than the film. Given the RC model for DA, it would make sense that the dielectrics with the highest insulation resistance (which also tend to have the lowest dielectric constant) would have the lowest dielectric absorption. There is also a significant difference from part to part. However, while the RC model is usefull for predicting how DA will behave, it does not reflect the underlying physics. See Space Charge. Also see: http://www.national.com/rap/Application/0,1570,28,00.html

- **Dissipation factor** (DF), of concern for AC power applications. DF is usually expressed in % or in ratio to unity (1% = .01). DF is equal to ESR/X_c . Actually, this is an approximation which ignores X_l, but it is usually close enough.

  Like most everything else about capacitors, it changes with time, frequency, and temperature. DF is the result of three loss factors. Metal losses (the resistance of the leads, end terminations, and the metal foil or film), the insulation resistance (mostly very small), and the dielectric losses. For film capacitors, film-foil types will have a lower dissipation factor than metallized film, especially at high frequency, because the metal losses will be lower.
- **Dissipation factor vs. applied voltage**, mostly of concern for Class 2 (and higher) ceramic capacitors.
- Dissipation factor vs. frequency.
- Dissipation factor vs. temperature.
- Equivalent series resistance (ESR), mostly of concern in high-frequency power applications and for power-supply filtering in high-speed digital systems. ESR is measured at different frequencies, depending on how a capacitor will be used. For cheap electrolytics this will be 120 Hz, for high-quality tantalums, it will usually be 100 kHz. Although measured in ohms, it is not independent of frequency, and in electrolytics, is somewhat higher at low frequency than at high frequency.
- **Equivalent series inductance (ESL)**, seen as an inductance in series. Mostly of concern for power-supply filtering in high-speed digital systems, and in RF applications.
- **Frequency range**. Some capacitors, like mica and C0G ceramic are usable to the GHz range. Others, like tantalum and polyester, become unusable somewhere in the 100 kHz to 1 MHz area (depending on construction details) due to loss of capacitance and increase in dissipation factor.
- Line filter capacitors. As a matter of safety, capacitors used to filter line (mains) power for electronic equipment, must meet strict performance requirements. The Europeans have lead the way in this, while American and Canadian requirements have been more lax.
- Insulation resistance (IR), seen as a resistance in parallel. This is referred to as "insulation resistance" when the leakage is very low (film capacitors), and as "leakage current" when the leakage is high (electrolytics). Electrolytic capacitors may have the most leakage, but it is also of concern for film capacitors in some analog applications (integrators, sample-and-hold circuits). For films, the lower the dielectric constant, the higher the insulation resistance tends to be. Data sheets normally state it in megohm-microfarads. To determine the IR for a given capacitor, divide the M-uF value by the actual capacitance. Although film capacitors and C0G ceramics have the best IR, all things being equal, low-quality ceramics can still work in some leakage-sensitive applications if the value is low, the voltage rating is high, and you can live with their other shortcomings.
- Markings. It can sometimes be a challenge to determine the characteristics of some capacitors from their markings. This partly due to size limitations, partly to a lack of universally followed standards, and partly to the failure to consistently follow the standards that do exist.
- **Mounting options**.
- **Noise.** Mostly of concern for electrolytic capacitors, due to leakage.
- Packaging and construction. Construction details and sealing methods can have a significant impact on performance and reliability. Most capacitor types require that air and moisture are sealed out to prevent degradation or contamination of the dielectric and corrosion of the metal film. The best Military-grade capacitors may be hermetically sealed in metal and glass. Lesser types may be molded in plastic, dipped in epoxy, or inserted in a plastic case and sealed with epoxy or urethane. The moisture absorption of the dielectric dictates the encapsulation methods allowed. As surface mounting of components replaces through-hole construction, a new problem faces the designer: can you get the capacitor you need in a surface mount package? Some capacitor types, like ceramics, are well suited to surface mounting. Others, like polystyrene and polypropylene, do not stand up to the heat of soldering and are not available for surface mounting.
- Reliability. Reliability and expected lifespan are of special concern in high-temperature, high-current, and high-voltage applications. However, capacitors used in more ordinary applications also have reliability issues.

- **Surge current capability**. Of concern for high-current pulse applications. Manufacturers will specify a maximum current or a maximum dV/dT (in volts/usec.) for capacitors rated for high-current applications.
- **Temperature limits**. Some dielectrics are limited to as low as 85C operating temperature while others can go to the 200-400C region. I have seen references to aerospace capacitors that can work to at least 1000C.
- **Voltage limits**, both operating and surge.



Capacitor Types

You can roughly divide all capacitors into four groups: film, ceramic, electrolytic, and miscellaneous.

FILM CAPACITORS

Film includes a variety of polymers, such as polyester, polycarbonate, Teflon, polypropylene, and polystyrene. Traditional film capacitors were only available in modest sizes, <10 uF. In recent years, film capacitors have sought to leverage their superior longevity compared to electrolytics, to move into some applications that call for much larger parts, even to thousands of uF. Film capacitors come in two broad categories, film-foil, and metallized film. Film-foil capacitors are made of alternating layers of plastic film and metal foil, while metallized film capacitors have the metal vacuum deposited directly on the film. In general, film-foil is better at handling high current, while metallized film caps are much better at self-healing. Various hybrid types can also be found.

Pros: The film capacitors mostly have reasonably well behaved electrical properties and offer many tradeoffs of performance and cost for people with precise requirements. The main parameters of interest include capacitance vs. temperature, dissipation factor, and dielectric absorption. Their main virtues include low leakage and low aging.

Cons: The main drawback of film capacitors is their low dielectric constants (K), which is only partly offset by their relatively good breakdown voltages. That means that film capacitors are physically large for their capacitance. Their Ks vary from a low of about 2.2 for Teflon to about 8 for PVDF (rarely used). Unfortunately, the rule-of-thumb is that the higher the K (and therefore the smaller the size), the worse the electrical properties tend to be. Film capacitors have not made an entirely graceful transition into the age of surface mounting. While some film dielectrics are suitable for surface mounting, most can't withstand the heat of soldering. Even polyester, the toughest of the traditional films, is barely good enough. However, capacitor makers have responded by developing several new dielectrics. SMD film capacitors are not as widely second-sourced as other capacitors however.

CERAMIC CAPACITORS

Ceramic capacitors offer a broad range of size vs. performance tradeoffs and are easily the most popular in numbers sold. Ceramic capacitors are available from < 1 pF to 1000s of uF.

Pros: The main virtue of ceramic capacitors are their relatively high dielectric constants. This can vary from C0G with a K of up to 60, which has excellent electrical properties but is relatively large and expensive, to ceramics with Ks in the tens of thousands but with very poor electrical properties. Large- value ceramics can compete with electrolytic capacitors in high-frequency applications like switch-mode power supplies because of their lower ESR. Ceramic capacitors are especially suitable for surface mounting due to their heat resistance, mechanical integrity, and the ability to make them in very small packages at low cost, for portable equipment. This has greatly added to their usage. To some extent ceramics are slowly displacing other types of capacitors.

Cons: Low breakdown voltage means that the low-K ceramics (Class 1), the ones with the good electrical properties, have poor volumetric efficiency, and are usually found only in small values. High-K ceramics (Class 2 and higher) have poor electrical properties, which are highly dependent on temperature, voltage, and frequency, plus a significant aging rate. Unlike many other capacitors, ceramics have no self-healing mechanism. This means that manufacturers must maintain a high level of quality control over the dielectric. Ceramics are most cost affective in small sizes at present. Very large ceramics are a bit of a challenge, especially in SMD.

ELECTROLYTIC CAPACITORS

"Electrolytic" means any capacitor that requires a conductive layer between the dielectric and one electrode. In the original electrolytic capacitor, the layer was an actual electrolyte, a conductive salt in a solvent. Some electrolytic capacitors today don´t actually use an electrolyte, but the word is still commonly used, to the annoyance of some. Electrolytic capacitors are made by growing a oxide film, the dielectric, on a metal, the anode, by electrochemical means. The films are very thin with fairly high Ks (roughly 10-25) which make for a lot of capacitance in a small package. The resulting devices pass current much better in one direction than the other, making a rectifier of a sort. Because of this, the metals are sometimes called "valve" metals. The metals presently used are aluminum, tantalum, and niobium.

Pros: Electrolytic capacitors are best used when you need a lot of capacitance in a small space and at a reasonable price, such as power supply filtering, or energy storage. They are available in sizes far beyond that of other capacitors. Aluminum electrolytics are presently available from 0.1 uF to several F. I have no idea why someone would use a 0.1 uF electrolytic capacitor however. Tantalum electrolytics are available from 0.1 uF to a few thousand uF.

Cons: Marginal electrical properties means that these capacitors must be applied with care. The parameters to be watched include leakage, service life vs. temperature, ESR, ESL, and low-temperature performance. Unlike other capacitors, electrolytic capacitors are not inherently non-polar, but non-polar types are available. Electrolytics are widely available in SMD packages, at least in moderate sizes, but users complain of more reliability problems than with through-hole styles.

MISCELLANEOUS DIELECTRICS

Miscellaneous capacitors include materials like glass, mica, porcelain, and even gas and vacuum. A few exotic dielectrics like silicon dioxide and sapphire are used in niche applications like microwave capacitors and trimmers. Some are available in surface-mount packages.

Pros: The electrical properties of the miscellaneous capacitors are generally most similar to film capacitors. However, they all have electrical properties that make them useful in some special applications.

Cons: These materials also have Ks similar to plastic films so they have no advantage in size. Except for mica, these capacitors are commonly available only in small sizes, <1 uF. They also tend to be more expensive than other capacitors of similar size.

Anyone who has corrections or additional information is more than welcome to contribute. If you have an idea for additional topics, expansion of existing topics, if something unclear, if you see an outright error, whatever, let me know. The site is still in its infancy, and I don´t really have a clear vision of what it will grow into. I would especially like to hear from the capacitor engineers of the world, they are the people who should have written this FAQ, but didn't. So far, my attempts to get these people involved has had limited success. It would be nice if an expert with specific knowledge would offer additional or corrected information not available to me.

What is good or bad about the site?
What would you like to see added?
Is your interest mainly for personal or business reasons?
Were you looking for something in specific, and did you find it?

endlr@execpc.com *: Research. Presently being treated for a long-standing capacitor fetish .
Dave D.: This was all his idea.
Schmuddy: editing (real name withheld by request).
John Woodgate: For considerable help with the Line-Filter Capacitor section.
ITW Paktron: For some background material and graphics on stacked-film capacitors.
Oren Elliott Products: For background material and graphics on air-variable capacitors.
Jennings Technology: For background material and a graphic on vacuum and gas capacitors.
Eichhoff Electronics: For graphics of line-filter capacitors and the line-filter modules they are used in.
Copernic2001: A great tool for searching the internet.
Google: Another great search tool.

Need a killer website designed?, see http://www.execpc.com/maevia/
No, Maevia didn´t do this site, they do much better work than this.

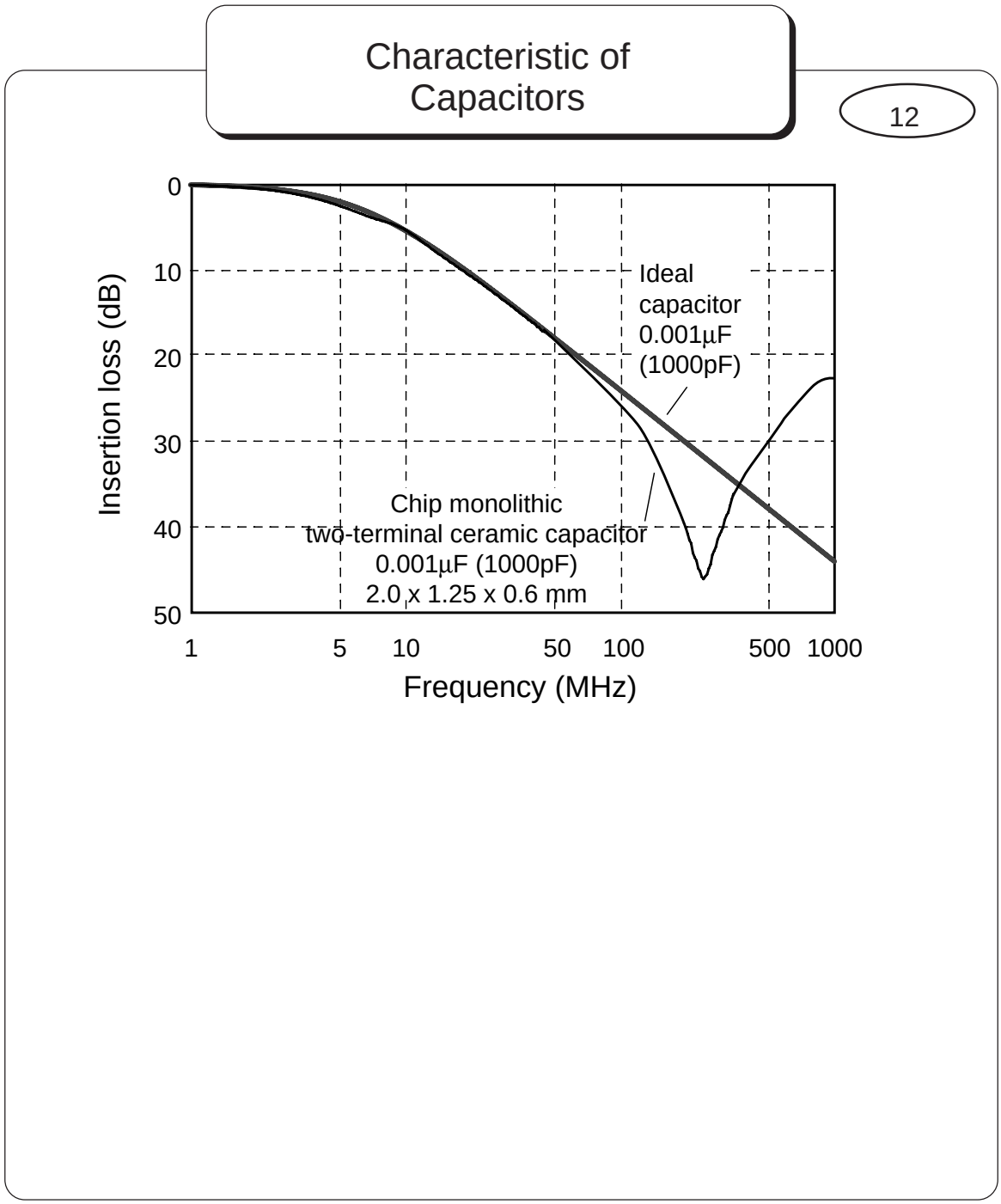
This site Copyright© 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006. All rights reserved.
Individuals may copy portions for personal use only.
All trademarks are property of their respective owners.
Links to other sites are not an endorsement of their information or products.

Date of last change: 12/24/05

* In a bit of bad news, my company was purchased by a conglomerate and at the end of July of `02 I became unemployed. However, I have made my plans for getting my career back on track.

3. Noise Suppression by Low-pass Filters

3.5. The Effect of Non ideal Capacitors



This section and the following sections describe the necessity and performance of capacitor-type EMI filters.

With the ideal capacitor, the insertion loss increases as the frequency becomes higher. However, with actual capacitors, the insertion loss increases until the frequency reaches a certain level (self-resonance frequency) and then insertion loss decreases.

[Notes]

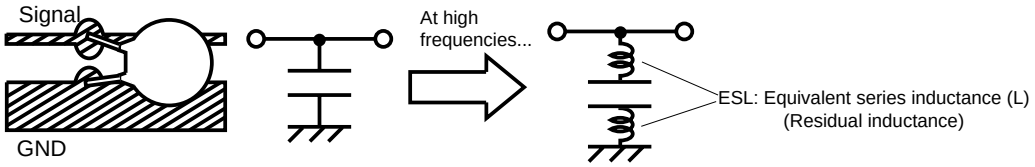
3. Noise Suppression by Low-pass Filters

3.5. The Effect of Non ideal Capacitors

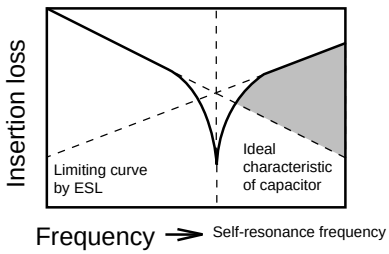
The Effect of Non ideal Capacitors

13

(a) Equivalent circuit of capacitor



(b) Effect by residual inductance



Self-resonance frequency

The frequency at which resonance occur due to the capacitor's own capacitance, and residual inductance. It is the frequency at which the impedance of the capacitor becomes zero.

From $j2\pi fL + 1/j2\pi fC = 0$,

$f = 1/2\pi\sqrt{LC}$

f: Self-resonance frequency

C: Capacitance

L: Residual inductance

[Notes]

The insertion loss of capacitors increase until the frequency reaches the self-resonance frequency and then decrease due to residual inductance of the lead wires and the capacitor's electrode pattern existing in series with the capacitance.Since noise is prevent from going through the bypass capacitors to the GND,the insertion loss decrease.The frequency at wich the insertion loss begins to decrease is called self-resonance frequency.

3. Noise Suppression by Low-pass Filters
3.5. The Effect of Non ideal Capacitors

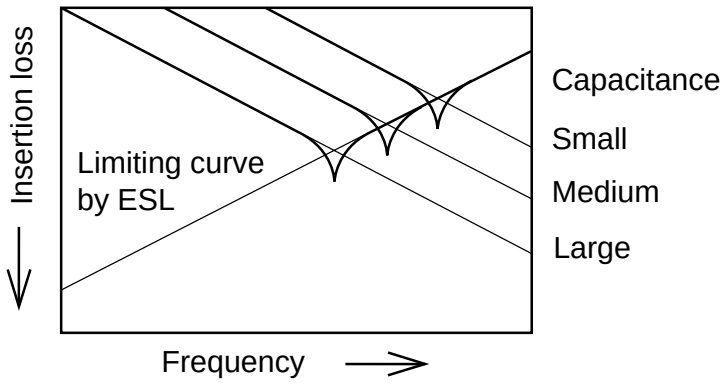
The Effect of ESL

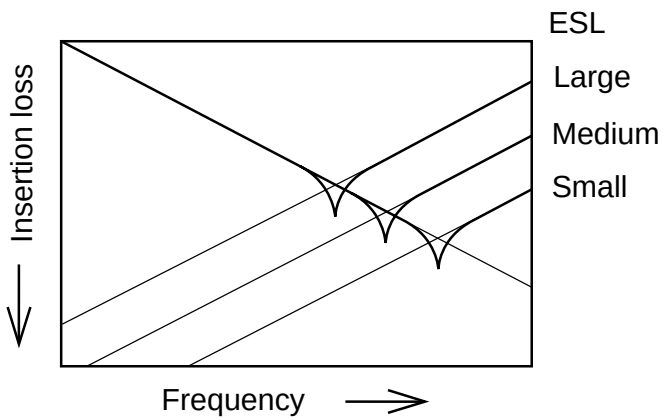
14

At frequencies higher than the self-resonance frequency, the insertion loss does not change regardless of whether the capacitance value is increased or decreased.

⇩

For use in a high-frequency range, a capacitor with a high self-resonance frequency, i.e. small residual inductance (ESL), must be selected.





[Notes]

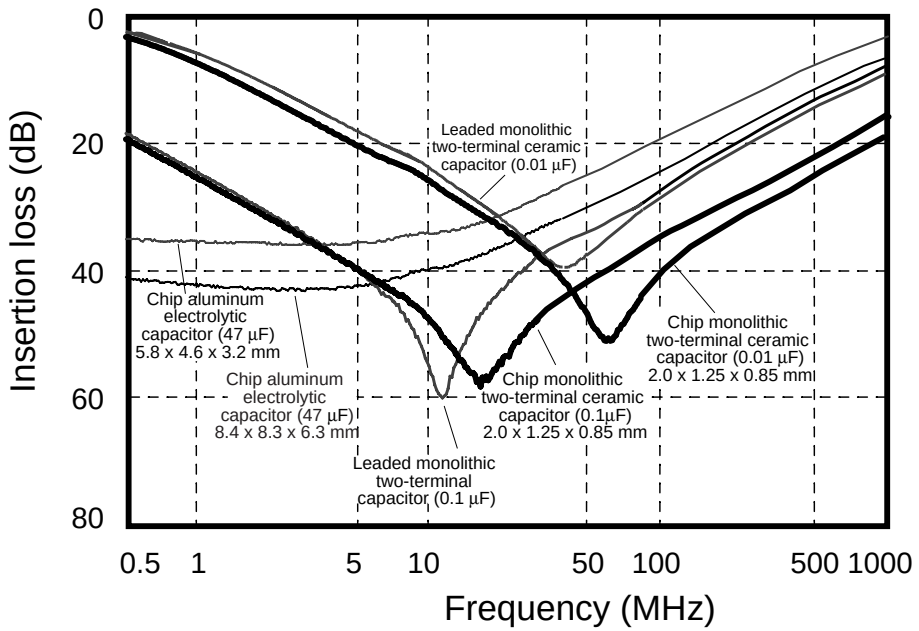
When the residual inductance is the same, the insertion loss does not change at frequencies above the self-resonance frequency, regardless of whether the capacitance value of the capacitor is increased or decreased. Therefore for greater noise suppression at frequencies higher than the self-resonance frequency, you must select a capacitor with a higher self-resonance frequency, i.e. small residual inductance.

3. Noise Suppression by Low-pass Filters

3.6. Characteristic of Typical Capacitors

Insertion Loss Characteristics of Typical Two-terminal Capacitors

15



[Notes]

The above drawing shows examples of insertion loss measurements of typical capacitors. For leaded capacitors, the insertion loss is measured with the lead wires cut to 1 mm.

3. Noise Suppression by Low-pass Filters
3.5. Characteristic of Typical Capacitors

Typical ESL Values for Capacitors

16

Type of Capacitor	Residual inductance (ESL)
Leaded disc ceramic capacitor (0.01 μF)	3.0 nH
Leaded disc ceramic capacitor (0.1 μF)	2.6 nH
Leaded monolithic ceramic capacitor (0.01 μF)	1.6 nH
Leaded monolithic ceramic capacitor (0.1 μF)	1.9 nH
Chip monolithic ceramic capacitor (0.01 μF, Size: 2.0 x 1.25 x 0.6 mm)	0.7 nH
Chip monolithic ceramic capacitor (0.1 μF, Size: 2.0 x 1.25 x 0.85 mm)	0.9 nH
Chip aluminum electrolytic capacitor (47 μF, Size: 8.4 x 8.3 x 6.3 mm)	6.8 nH
Chip tantalum electrolytic capacitor (47 μF, Size: 5.8 x 4.6 x 3.2 mm)	3.4 nH

[Notes]

The above table shows typical residual inductances (ESL) values for capacitors, which are calculated from the impedance curves shown on the previous page.
The residual inductance varies depending on the type of capacitor. It can also vary in the same type of capacitor, depending on the dielectric material and the structure of the electrode pattern.

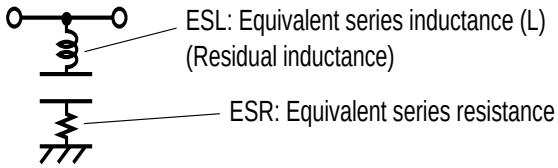
3. Noise Suppression by Low-pass Filters

3.8. The Effect of Equivalent Series Resistance

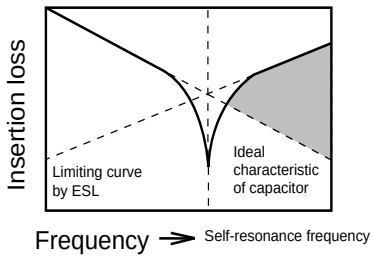
The Effect of Equivalent Series Resistance

20

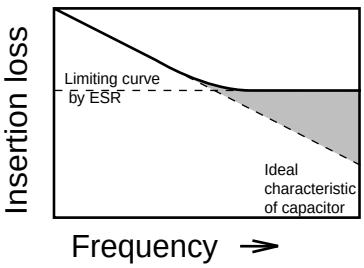
(a) Capacitor's equivalent circuit with ESL and ESR



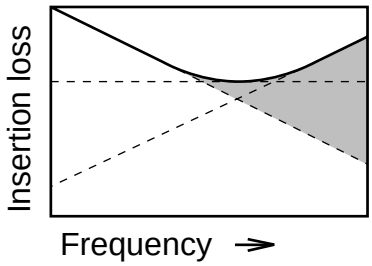
(b) Affect by ESL



(c) Affect by ESR



(d) Insertion loss frequency characteristic of actual capacitor affected by ESL and ESR



The second factor that causes deterioration in the characteristic of capacitors is equivalent series resistance (ESR). The insertion loss will be lower due to ESR caused by the electrode and material.

The ESR is very low in ceramic capacitors but higher in aluminum electrolytic capacitors.

[Notes]

Research Solutions & Resources

Home

What's New?

Electrochemistry Resources CONTENTS

The Bookstore CONTENTS

Search

Contact Us!

Tell Us !
How useful was this information?

The Inductance of Your Electrode!

RESOURCES > EIS > INDUCTANCE > CALCULATOR

- **Even a single, straight piece of wire has some inductance!**

We generally associate inductance with a loop or coil of wire. However, even a straight piece of wire, **or your electrode**, has some self-inductance. This can be important if you are dealing with low impedances at high frequencies.

- My good, old 1962-63 edition of the CRC Handbook is a wealth of information! It contains a section on "*Radio Formulae*" (ref 1) that gives the inductance of a straight piece of round wire or rod:

$$L = 2l\left[2.303\log(4l/d) - 1 + \mu/4 + (d/2l)\right]$$

In this equation, **L** is the inductance in nH (10^{-9} henry), ***l*** is the length and ***d*** is the diameter of the wire/rod (both in cm). **μ** is the permeability of the material (=1.0, except for iron and other ferromagnetic materials).

Wire Diameter (***d***) = [ ] cm

Wire Length (***l***) = [ ] cm

Permeability (**μ**) = [1.0] **(1.0 EXCEPT for IRON)**

Calculate Inductance | Clear Calculator

Inductance = [ ] nH

Uses JavaScript. You must have a JavaScript capable browser, and have scripting enabled.
Download the newest Internet Explorer or Netscape now !

The calculator, above, actually uses a slightly more complicated form of the equation (ref 2). The differences are small, even for fairly "squat" electrodes with ***d*** / ***l*** of 0.2 (x~1.005).

$$x = \sqrt{1 + (d/2l)^2}$$

$$L = 2l\left[\ln\left[(2l/d)(1+x)\right] - x + \mu\delta + d/2l\right]$$

More About ...

Instrumental artifacts
Inductance at Low Frequency
Inductance at High Frequency

More About EIS

How useful was this information? Give us your FEEDBACK !

REFERENCES

- (1) "Handbook of Chemistry and Physics, 44th Ed.", Chemical Rubber Publishing Co., Cleveland, OH, 1962. Yes, I got it NEW.
- (2) http://en.wikibooks.org/wiki/Electronics:Inductors#Construction has a more complete equation.
- (3) http://www.ee.scu.edu/eefac/healy/indwire.html
- (4) http://home.san.rr.com/bushnell/self_inductance.htm

© Copyright 2004-2005 Research Solutions & Resources.
All rights reserved. Updated undefined.

Validated by W3C

Dielectric Constant Reference Guide

Dielectric Constant (k) is a number relating the ability of a material to carry alternating current to the ability of vacuum to carry alternating current. The capacitance created by the presence of the material is directly related to the Dielectric Constant of the material.

Knowing the Dielectric Constant (k) of a material is needed to properly design and apply instruments such as level controls using radar, RF admittance, or capacitance technologies. There are also analytical reasons to know the (k) of a material.

How to use this guide

CLIPPER CONTROLS has compiled an extensive list of products with Dielectric Constants. Many of these Dielectric Constants are given at specific temperatures. If your product's temperature is significantly different from those listed there is a good chance that the Dielectric Constant may be different from the values listed.

The products in this reference are listed in alphabetical order and are grouped in sections by the first letter of their name. Proper chemical names were used, and any trade names are the trademark of their respective owners. If you know the correct spelling of the name of the product you wish to review then use the "Search" feature on the web browser to locate the name in the list. You may also click on the letter from the alphabetical table to go directly to the beginning of that alphabetic section.

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z #

-A-

ABS Resin, Lump 2.4-4.1

ABS Resin, Pellet 1.5-2.5

Acenaphthene (70° F) 3.0

Acetal (70° F) 3.6

Acetal Bromide 16.5

Acetal Doxime (68° F) 3.4

Acetaldehyde (41° F) 21.8

Acetamide (68° F) 41

Acetamide (180° F) 59.0

Acetanilide (71° F) 2.9

Acetic Acid (68° F) 6.2

Acetic Acid (36° F) 4.1

Acetic Anhydride (66° F) 21.0

Acetone (77° F) 20.7

Acetone (127° F) 17.7

Acetone (32° F) 1.0159

Acetonitrile (70° F) 37.5

Acetophenone (75° F) 17.3

Acetoxime (24° F) 3

Acetyl Acetone (68° F) 23.1

Acetyl Bromide (68° F) 16.5

Acetyl Chloride (68° F) 15.8

Acetyle Acetone (68° F) 25.0

Acetylene (32° F) 1.0217

Acetylmethyl Hexyl Ketone (66° F) 27.9

Acrylic Resin 2.7 - 4.5

Acteal 21.0-3.6

Air 1

Air (Dry) (68° F) 1.000536

Alcohol, Industrial 16-31

Alkyd Resin 3.5-5

Allyl Alcohol (58° F) 22.0

Allyl Bromide (66° F) 7.0

Allyl Chloride (68° F) 8.2

Allyl Iodide (66° F) 6.1

Allyl Isothiocyanate (64° F) 17.2

Allyl Resin (Cast) 3.6 - 4.5

Alumina 9.3-11.5

Alumina 4.5

Alumina China 3.1-3.9

Aluminum Bromide (212° F) 3.4

Aluminum Fluoride 2.2

Aluminum Hydroxide 2.2

Aluminum Oleate (68° F) 2.4

Aluminum Phosphate 6.0

Aluminum Powder 1.6-1.8

Amber 2.8-2.9

Aminoalkyd Resin 3.9-4.2

Ammonia (-74° F) 25

Ammonia (-30° F) 22.0

Ammonia (40° F) 18.9

Ammonia (69° F) 16.5

Ammonia (Gas?) (32° F) .0072

Ammonium Bromide 7.2

Ammonium Chloride 7.0

Amyl Acetate (68° F) 5.0

Amyl Alcohol (-180° F) 35.5

Amyl Alcohol (68° F) 15.8

Amyl Alcohol (140° F) 11.2

Amyl Benzoate (68° F) 5.1

Amyl Bromide (50° F) 6.3

Amyl Chloride (52° F) 6.6

Amyl Ether (60° F) 3.1

Amyl Formate (66° F) 5.7

Amyl Iodide (62° F) 6.9

Amyl Nitrate (62° F) 9.1

Amyl Thiocyanate (68° F) 17.4

Amylamine (72° F) 4.6

Amylene (70° F) 2.0

Amylene Bromide (58° F) 5.6

Amylenetetrararboxylate (66° F) 4.4

Amylmercaptan (68° F) 4.7

Aniline (32° F) 7.8

Aniline (68° F) 7.3

Aniline (212° F) 5.5

Aniline Formaldehyde Resin 3.5 - 3.6

Aniline Resin 3.4-3.8

Anisaldehyde (68° F) 15.8

Anisaldoxine (145° F) 9.2

Anisole (68° F) 4.3

Anitmony Trichloride 5.3

Antimony Pentachloride (68° F) 3.2

Antimony Tribromide (212° F) 20.9

Antimony Trichloride (166° F) 33.0

Antimony Trichloride 5.3

Antimony Tricodide (347° F) 13.9

Apatite 7.4

Argon (-376° F) 1.5

Argon (68° F) 1.000513

Arsenic Tribromide (98° F) 9.0

Arsenic Trichloride (150° F) 7.0

Arsenic Trichloride (70° F) 12.4

Arsenic Triiodide (302° F) 7.0

Arsine (-148° F) 2.5

Asbestos 3.0 - 4.8

Ash (Fly) 1.7 - 2.0

Asphalt (75° F) 2.6

Asphalt, Liquid 2.5-3.2

Azoxyanisole (122° F) 2.3

Azoxybenzene (104° F) 5.1

Azoxyphenitole (302° F) 6.8

Return to top

-B-

Bakelite 3.5-5.0

Ballast 5.4-5.6

Ballmill Feed (Cement) 4.5

Balm, Refuse 3.1

Barium Chloride 9.4

Barium Chloride (Anhyd) 11.0

Barium Chloride (2h20) 9.4

Barium Nitrate 5.8

Barium Sulfate (60° F) 11.4

Barley Flour 3.0 - 4.0

Barley Powder 3.0-4.0

Beeswax 2.7 - 3.0

Benzal Chloride (68° F) 6.9

Benzaldehyde (68° F) 17.8

Benzaldoxime (68° F) 3.8

Benzene (68° F) 2.3

Benzene (275° F) 2.1

Benzene (700° F) 1.0028

Benzil (202° F) 13.0

Benzonitrile (68° F) 26.0

Benzophenone (122° F) 11.4

Benzophenone (68° F) 13.0

Benzotrichloride (68° F) 7.4

Benzoyl Chloride (70° F) 22.1

Benzoyl Chloride (32° F) 23.0

Benzoylacetone (68° F) 29.0

Benzyl Acetate (70° F) 5.0

Benzyl Alcohol (68° F) 13.0

Benzyl Benzoate (68° F) 4.8

Benzyl Chloride (68° F) 6.4

Benzyl Cyanide (68° F) 18.3

Benzyl Cyanide (155° F) 6.0

Benzyl Salicylate (68° F) 4.1

Benzylamine (68° F) 4.6

Benzylethylamine (68° F) 4.3

Benzylmethylamine (67° F) 4.4

Beryl 6.0

Biphenyl 20

Biwax 2.5

Bleaching Powder 4.5

Bone Black 5.0-6.0

Bornyl Acetate (70° F) 4.6

Boron Bromide (32° F) 2.6

Boronyl Chloride (202° F) 5.2

Bromaceytal Bromide 12.6

Bromal (70° F) 7.6

Bromine (68° F) 3.1

Bromine (32° F) 1.0128

Bromo-2-Ethoxypentane (76° F) 6.5

Bromoacetyl Bromide (68° F) 12.6

Bromoaniline (68° F) 13

Bromoanisole (86° F) 7.1

Bromobenzene (68° F) 5.4

Bromobutylene (68° F) 5.8

Bromobutyric Acid (68° F) 7.2

Bromoctadecane 3.53

Bromodecane (76° F) 4.4

Bromodeodecane (76° F) 4.1

Bromodocosane (130° F) 3.1

Bromodoecane (75° F) 4.07

Bromoform (68° F) 4.4

Bromoheptane (76° F) 5.3

Bromohexadecane (76° F) 3.7

Bromohexane (76° F) 5.8

Bromoisovaleric Acid (68° F) 6.5

Bromomethane (32° F) 9.8

Bromonapthalene (66° F) 5.1

Bromooctadecane (86° F) 3.5

Bromopentadecane (68° F) 3.9

Bromophropionic Acid (68° F) 11.0

Bromotoluene (68° F) 5.1

Bromotridecane (50° F) 4.2

Bromoundecane (15° F) 4.7

Bronyl Chloride (94° F) 5.21

Butane (30° F) 1.4

Butanol (1) (68° F) 17.8

Butanone (68° F) 18.5

Butycic Anhydride (20° F) 12.0

Butyl Chloral (64° F) 10.0

Butyl Chloride (68° F) 9.6

Butyl Oleate (77° F) 4.0

Butyl Stearate (80° F) 3.1

Butylacetate (66° F) 5.1

Butylamine (70° F) 5.4

Butyraldehyde (79° F) 13.4

Butyric Acid (68° F) 3.0

Butyric Anhydride (68° F) 12.0

Butyronitrile (70° F) 20.7

Return to top

-C-

Cable Oil (80° F) 2.2

Cabondioxide (68° F) 1.000921

Calcim Fluoride 7.4

Calcite 8.0

Calcium 3.0

Calcium Carbonate 6.1-9.1

Calcium Fluoride 7.4

Calcium Oxide, Granule 11.8

Calcium Sulfate 5.6

Calcium Sulfate (H2o) 5.6

Calcium Superphosphate 14-15

Camphanedione (398° F) 16.0

Camphene (68° F) 2.7

Camphene (104° F) 2.3

Campher, Crystal 10-11

Camphoric Imide 4 (80° F) 5.5

Camphorpinacone (68° F) 3.6

Caprilic Acid (18° F) 3.2

Caproic Acid (160° F) 2.6

Caprolactam Monomer 1.7 - 1.9

Caprylic Acid (65° F) 3.2

Carbide 5.8 - 7.0

Carbide, Powder 5.8-7.0

Carbon Black 2.5 - 3.0

Carbon Dioxide (32° F) 1.6

Carbon Dioxide, Liquid 1.6

Carbon Disulfide, Liquid 2.6

Carbon Disulphide (68° F) 2.6

Carbon Disulphide (180° F) 2.2

Carbon Tetrachloride (68° F) 2.2

Carnauba Wax 2.9

Carvenone (68° F) 18.4

Carvol (64° F) 11.2

Carvone (71° F) 11.0

Casein 6.1 - 6.8

Casein Resin 6.7

Cassiterite 23.4

Castor Oil (60° F) 4.7

Castor Oil (80° F) 2.6

Castor Oil (Hydrogenated) (80° F) 10.3

Cedrene (76° F) 3.2

Cellophane 3.2-6.4

Celluloid 3.3-11

Cellulose 3.2-7.5

Cellulose Acetate 3.2-7

Cellulose Acetate (Molding) 3.2 - 7.0

Cellulose Acetate (Sheet) 4.0 - 5.5

Cellulose Acetate Butyrate 3.2 - 6.2

Cellulose Nitrate (Proxylin) 6.4

Cement 1.5 - 2.1

Cement (Plain) 1.5 - 2.1

Cement, Portland 2.5-2.6

Cement, Powder 5-10

Cereals (Dry) 3.0 - 5.0

Cerese Wax 2.4

Cesium Iodine 5.6

Cetyl Iodide (68° F) 3.3

Charcoal 1.2-1.81

Chinaware, Hard 4-7

Chloracetic Acid (140° F) 12.3

Chloracetone 29.8

Chloral (68° F) 4.9

Chlorhexanone Oxime 3

Chlorine (-50° F) 2.1

Chlorine (32° F) 2.0

Chlorine (142° F) 1.5

Chlorine, Liquid 2

Chloroacetic Acid (68° F) 21.0

Chloroacetone (68° F) 29.8

Chlorobenzene (77° F) 5.6

Chlorobenzene (100° F) 4.7

Chlorobenzene (230° F) 4.1

Chlorobenzine, Liquid 5.5-6.3

Chlorocyclohexane (76° F) 7.6

Chloroform (32° F) 5.5

Chloroform (68° F) 4.8

Chloroform (212° F) 3.7

Chloroheptane (71° F) 5.5

Chlorohexanone Oxime (192° F) 3.0

Chlorohydrate (68° F) 3.3

Chloromethane -4 12.6

Chloronaphthalene (76° F) 5.0

Chlorooctane (76° F) 5.1

Chlorophetane 5.4

Chlorotoluene (68° F) 4.7

Chlorotoluene, Liquid 4-4.5

Cholesterin 2.86

Cholestral (80° F) 2.9

Chorine (170° F) 1.7

Chrome, Ore 7.7-8.0

Chrome, Pure 12

Chromite 4.0-4.2

Chromyl Choride (68° F) 2.6

Cinnamaldehyde (75° F) 16.9

Cis-3-Hexene (76° F) 2.1

Citraconic Anhydride (68° F) 40.3

Citraconic Nitrile 27

Clay 1.8 - 2.8

Clinker (Cement) 2.7

Coal Tar 2.0-3.0

Coal, Powder, Fine 2-4

Cocaine (68° F) 3.1

Coffee Refuse 2.4-2.6

Coke 1.1 - 2.2

Compound 3.6

Copper Catalyst 6.0 - 6.2

Copper Oleate (68° F) 2.8

Copper Oxide 18.1

Corderite 2.5 - 5.4

Corn 5-10

Corn (Dry Granulars) 1.8

Corn, Refuse 2.3-2.6

Corning Glass 6.5

Cotton 1.3-1.4

Cotton Seed Oil 3.1

Co2 (32° F) 1.6

Creosol (63° F) 10.6

Cresol (75° F) 5.0

Cresol, Liquid 9-11

Crotonic Nitrice (68° F) 28.0

Crystale 3.5-4.7

Cumaldehyde (59° F) 11.0

Cumene (68° F) 2.4

Cumicaldehyde (58° F) 10.7

Cupric Oleate 2.8

Cupric Oxide (60° F) 18.1

Cupric Sulfate 10.3

Cupric Sulfate (Anhyd) 10.3

Cupric Sulfate ($5h_2o$) 7.8

Cyanoacetic Acid (40° F) 33.0

Cyanoethyl Acetate (68° F) 19.3

Cyanogen (73° F) 2.6

Cyclohedane (20° F) 2.0

Cyclohenanone (68° F) 18.2

Cycloheptasiloxane (68° F) 2.7

Cyclohexane (68° F) 2.0

Cyclohexane, Liquid 18.5

Cyclohexanecarboxylic Acid (88° F) 2.6

Cyclohexanemethanol (140° F) 9.7

Cyclohexanol (77° F) 15.0

Cyclohexanone (68° F) 18.2

Cyclohexanone Oxime (192° F) 3.0

Cyclohexene (68° F) 18.3

Cyclohexylamine-5 5.3

Cyclohexylphenol (130° F)

4.0 Cyclohexyltrifluoromethane-1 (68° F) 11.0

Cyclopentane (68° F) 2.0

Cymene 62 2.3

Return to top

-D-

D-Cocaine 3.1

D.M.T. (Dacron Powder) 1.33

Decahydronaphtolene (68° F) 2.2

Decamethylcyclopentasiloxane (68° F) 2.5

Decamethyltetrasiloxane (68° F) 2.4

Decanal 8.1

Decane (68° F) 2.0

Decanol (68° F) 8.1

Decylene (62° F) 2.7

Decyne (68° F) 2.2

Deuterium (68° F) 1.3

Deuterium Oxide (77° F) 78.3

Dextrin 2.2-2.4

Diacetoxybutane (76° F) 6.64

Diallyl Sulfide (68° F) 4.9

Diamond 5.5 - 10.0

Diaphenylmethane 2.7

Diaplmitin 3.5

Dibenzofuran (212° F) 3.0

Dibenzyl Sebacate (68° F) 4.6

Dibenzylamine (68° F) 3.6

Dibroheptane (24° F) 5.08

Dibromobenzene (68° F) 8.8

Dibromobutane (68° F) 5.7

Dibromoethylene (Cis-1, 2) (32° F) 7.7

Dibromoheptane (76° F) 5.1

Dibromohexane (76° F) 5.0

Dibromomethane (50° F) 7.8

Dibromopropane (68° F) 4.3

Dibromopropyl Alcohol (70° F) 9.1

Dibutyl Phthalate (86° F) 6.4

Dibutyl Sebacate (86° F) 4.5

Dibutyl Tartrate 109 9.4

Dichloracetic Acid (20° F) 10.7

Dichloracetic Acid (72° F) 8.2

Dichloracetone (68° F) 14.0

Dichlorobenzene (127° F) 2.8

Dichloroethane (68° F) 16.7

Dichloroethane (1,2)(77° F) 10.3

Dichloroethylene (62° F) 4.6

Dichloromethane (68° F) 9.1

Dichlorostyrene (76° F) 2.6

Dichlorotoluene (68° F) 6.9

Dictyl Phthalate 5.1

Dicyclohexyl Adipate (95° F) 4.8

Diebenzylamine (68° F) 3.6

Diethyl Benzalmalonate (32° F) 8.0

Diethyl Disulfide (66° F) 15.9

Diethyl Dl-Malate (64° F) 10.2

Diethyl Glutarate (86° F) 6.7

Diethyl I-Malate 9.5

Diethyl Ketone (58° F) 17.3

Diethyl L-Malate (68° F) 9.5

Diethyl Malonate (70° F) 7.9

Diethyl Oxalate (70° F) 8.2

Diethyl Oxaloacetate (66° F) 6.1

Diethyl Racemate (68° F) 4.5

Diethyl Sebacate (86° F) 5.0

Diethyl Succinate (86° F) 6.6

Diethyl Succinosuccinate (66° F) 2.5

Diethyl Sulfide (68° F) 7.2

Diethyl Sulfite (68° F) 15.9

Diethyl Tartrate (68° F) 4.5

Diethyl Zinc (68° F) 2.6

Diethyl 1-Malate (68° F) 9.5

Diethyl-Dimalate 10.2

Diethylamine (68° F) 3.7

Diethylaniline (66° F) 5.5

Dihydrocaroone (66° F) 8.7

Dihydrocarvone (66° F) 8.5

Diimylamine (64° F) 2.5

Diioamylene (62° F) 2.4

Diiodoethylene 1 (80° F) 4.0

Diiodomethane (77° F) 5.3

Diisoamyl (62° F) 2.0

Diisoamylene 2.4

Diisobutylamine (71° F) 2.7

Dimethoxybenzene (73° F) 4.5

Dimethyl Ethyl (68° F) 11.7

Dimethyl Ethyl Carbinol (68° F) 11.7

Dimethyl Malonate (68° F) 10.4

Dimethyl Oxalate (68° F) 3.0

Dimethyl Pentane (20° F) 1.912

Dimethyl Phthalate (75° F) 8.5

Dimethyl Sulfate (68° F) 55.0

Dimethyl Sulfide (68° F) 6.3

Dimethyl-1-Hydroxybenzene (62° F) 4.8

Dimethyl-2-Hexane (68° F) 2.4

Dimethylamine (32° F) 6.3

Dimethylaniline (68° F) 4.4

Dimethylbromoethylene (68° F) 6.7

Dimethylheptane (68° F) 1.9

Dimethylpentane (68° F) 1.9

Dimethylquinoxaline (76° F) 2.3

Dimethyltouidine (68° F) 3.3

Dinitrogen Oxide (32° F) 1.6

Dinitrogen Tetroxide (58° F) 2.5

Dioctyl Phthalate (76° F) 5.1

Dioxane 1,4 (77° F) 2.2

Dipalmitin (161° F) 3.5

Dipentene (68° F) 2.3

Dipenylamine (125° F) 3.3

Diphemylethane (230° F) 2.4

Diphemylethane (62° F) 12.6

Diphenyl 1(66° F) 2.5

Diphenyl Ether (82° F) 3.9

Diphenylamine (124° F) 3.3

Diphenylethane (110° F) 2.38

Diphenytmethane (62° F) 2.6

Dipropyl Ketone (62° F) 12.6

Dipropylamine (70° F) 2.9

Distearin (172° F) 3.3

Docosane (122° F) 2.0

Dodecamethylcyclohexisloxane (68° F) 2.6

Dodecamethylpentasiloxane (68° F) 2.5

Dodecane (68° F) 2.0

Dodecanol (76° F) 6.5

Dodecyne (76° F) 2.2

Dolomite 6.8-8.0

Dowtherm (70° F) 3.4

Return to top

-E-

Ebonite 2.5-2.9

Emery Sand 16.5

Epichlorchydrin (68° F) 22.9

Epoxy Resin (Cast) 3.6

Ethanediamine (68° F) 14.2

Ethanethiol (58° F) 6.9

Ethanethiolic Acid (68° F) 13.0

Ethanol (77° F) 24.3

Ethelene Diamine (18° F) 16.0

Ethelene Oxide -1 13.9

Ethoxy-3-Methylbutane (68° F) 4.0

Ethoxybenzene (68° F) 4.2

Ethoxyethyl Acetate (86° F) 7.6

Ethoxynapthalone (66° F) 3.3

Ethoxypentane (73° F) 3.6

Ethoxytoluene (68° F) 3.9

Ethyl Acetate (77° F) 6.0

Ethyl Acetoacetate (71° F) 15.9

Ethyl Acetoneoxalate (66° F) 16.1

Ethyl Acetophenoneoxalate (66° F) 3.3

Ethyl Alcohol (77° F) 24.3

Ethyl Alcohol (See Ethanol)

Ethyl Amyl Ether (68° F) 4.0

Ethyl Benzene (68° F) 2.5

Ethyl Benzoate (68° F) 6.0

Ethyl Benzoylacetate (68° F) 12.8

Ethyl Benzoylacetoacetate (70° F) 8.6

Ethyl Benzyl Ether (68° F) 3.8

Ethyl Bromide (64° F) 4.9

Ethyl Bromoisobutyrate (68° F) 7.9

Ethyl Bromopropionate (68° F) 9.4

Ethyl Butyrate (66° F) 5.1

Ethyl Carbonate (68° F) 3.1

Ethyl Carbonate (121° F) 14.2

Ethyl Cellulose 2.8 - 3.9

Ethyl Chloracetate (68° F) 11.6

Ethyl Chloroformate (68° F) 11.3

Ethyl Chloropropionate (68° F) 10.1

Ethyl Cinnamate (66° F) 5.3

Ethyl Cyanoacetate (68° F) 27.0

Ethyl Cyclobutane (68° F) 2.0

Ethyl Dodecanoate (68° F) 3.4

Ethyl Ether (-148° F) 8.1

Ethyl Ether (-40° F) 5.7

Ethyl Ether (68° F) 4.3

Ethyl Ethoxybenzoate (70° F) 7.1

Ethyl Formate (77° F) 7.1

Ethyl Formylphenylacetate (68° F) 3.0

Ethyl Fumarate (73° F) 6.5

Ethyl Hydroxy-Tetracarboxylate 5.9

Ethyl Hydroxy-Tetrocarboxylate 2.7

Ethyl Hydroxymethylenephenylacet 5.00

Ethyl Hydroxymethylenomalonate 6.6

Ethyl Iodide (68° F) 7.4

Ethyl Isothiocyanate (68° F) 19.7

Ethyl Levulinete (70° F) 12.1

Ethyl Maleate (73° F) 8.5

Ethyl Mercaptan (68° F) 8.0

Ethyl Nitrate (68° F) 19.7

Ethyl Oleate (80° F) 3.2

Ethyl Palmitate (68° F) 3.2

Ethyl Phenylacetate (70° F) 5.4

Ethyl Propionate (68° F) 5.7

Ethyl Salicylate (70° F) 8.6

Ethyl Silicate (68° F) 4.1

Ethyl Stearate (104° F) 3.0

Ethyl Thiocyanate (68° F) 29.6

Ethyl Trichloracetate (68° F) 7.8

Ethyl Undecanoate (68° F) 3.6

Ethyl Valerate (68° F) 4.7

Ethyl 1-Brobutyrate (68° F) 8.0

Ethyl 2-Iodopropionate (68° F) 8.8

Ethylamine (70° F) 6.3

Ethylaniline (68° F) 5.9

Ethylbenzene (76° F) 3.0

Ethylene Chloride (68° F) 10.5

Ethylene Chlorohydrin (77° F) 26.0

Ethylene Cyanide (136° F) 58.3

Ethylene Diamine (64° F) 16.0

Ethylene Gylcol (68° F) 37.0

Ethylene Iodide 3.4

Ethylene Oxide 25 14.0

Ethylene Tetraflouride 1.9-2.0

Ethylenechlorohydrin (75° F) 25.0

Ethylenediamine (64° F) 16.0

Ethylic Resin 2.2-2.3

Ethylpentane (68° F) 1.9

Ethyltoluene (76° F) 2.2

Etibine (-58° F) 2.5

Eugenol (64° F) 6.1

Return to top

-F-

Fab (From Box, 8% Moisture) 1.3

Fenchone (68° F) 12.0

Fermanium Tetrachloride (76° F) 2.4

Ferric Oleate (68° F) 2.6

Ferrochromium 1.5-1.8

Ferromanganese 5.0-5.2

Ferrous Oxide (60° F) 14.2

Ferrous Sulfate (58° F) 14.2

Flour 2.5-3.0

Flourine (-332° F) 1.5

Flourspar 6.8

Fluorotoluene (86° F) 4.2

Fly Ash 1.9 - 2.6

Formalin 23

Formamide (68° F) 84.0

Formic Acid (60° F) 58.0

Forsterite 6.2

Freon 11 (70° F) 3.1

Freon 113 (70° F) 2.6

Freon 12 (70° F) 2.4

Fuller's Earth 1.8 - 2.2

Furan (77° F) 3.0

Furfural (68° F) 42.0

Furfuraldehyde (68° F) 41.9

Return to top

-G-

Gasoline (70° F) 2.0

Gerber Oatmeal (In Box) 1.5

Germanium Tetrachloride (77° F) 2.4

Glass 3.7-10

Glass (Silica) 3.8

Glass, Bead 3.1

Glass, Granule 6-7

Glass, Raw Material 2.0-2.5

Glucoheptitol (248° F) 27.0

Glycerin, Liquid 47-68

Glycerol (77° F) 42.5

Glycerol (32° F) 47.2

Glycerol Phthalate (Cast Alkyd) 3.7 - 4.0

Glyceryl Triocetate (70° F) 6.0

Glycol (77° F) 37.0

Glycol (122° F) 35.6

Glycolic Nitrile (68° F) 27.0

Grain 3-8

Graphite 12-15

Guaiacol 0 11.0

Gypsum 2.5-6.0

Return to top

-H-

Hagemannie Ester (68° F) 10.6

Halowax 4.5

Heavy Oil 3

Heavy Oil, C 2.6

Helium-3(58° F) 1.055

Helium, Liquid 1.05

Heptadecanone (140° F) 5.3

Heptane (68° F) 1.9

Heptane, Liquid 1.9-2.0

Heptanoic Acid 2.5

Heptanoic Acield (71° F) 2.59

Heptanone (68° F) 11.9

Heptaonic Acid (160° F) 2.6

Heptyl Alcohol (70° F) 6.7

Hexamethyldisiloxane (68° F) 2.2

Hexane (-130° F) 2.0

Hexanol (77° F) 13.3

Hexanone (59° F) 14.6

Hexdecamethylcycloheptasiloxane (68° F) 2.7

Hexyl Iodide (68° F) 6.6

Hexylene (62° F) 2.0

Hexyliodide (68° F) 6.6

Hydrazine (68° F) 52.0

Hydrochloric Acid (68° F) 4.60

Hydrocyanic Acid (70° F) 2.3

Hydrocyanic Acid (32° F) 158.0

Hydrogen (440° F) 1.23

Hydrogen (212° F) 1.000284

Hydrogen Iodide (72° F) 2.9

Hydrogen Bromide (24° F) 3.8

Hydrogen Bromide (-120° F) 7.0

Hydrogen Chloride (82° F) 4.6

Hydrogen Chloride (-188° F) 12.0

Hydrogen Cyanide (70° F) 95.4

Hydrogen Fluoride (32° F) 84.2

Hydrogen Fluoride (-100° F) 17

Hydrogen Iodide(72° F) 2.9

Hydrogen Peroxide (32° F) 84.2

Hydrogen Peroxide 100% 70.7

Hydrogen Peroxide 35% 121.0

Hydrogen Sulfide (-84° F) 9.3

Hydrogen Sulfide (48° F) 5.8

Hydrofluoric Acid (32° F) 83.6

Hydroxy-4-Methy-2-Pentanone (76° F) 18.2

Hydroxymethylene Camphor (86° F) 5.2

Hydroxymethylenehydroxymethyleneacetoacetate 7.8

Hydroxymethylenebenzyl Cyanide (68° F) 6.0

Hydrozine (68° F) 52.9

Return to top

Ido-Iodohexadecane (68° F) 3.5

Idoheptane (71° F) 4.9

Idohexane (68° F) 5.4

Idomethane (68° F) 7.0

Idopoctane (76° F) 4.6

Idotoluene (68° F) 6.1

Ilmenite 6.0 - 7.0

Inadol (140° F) 7.8

Indonol (60° F) 7.8

Iodine (107° F) 118.0

Iodine 11

Iodine (250° F) 118.0

Iodine (Granular) 4.0

Iodioctane 4.6

Iodloctane (24° F) 4.62

Iodobenzene (68° F) 4.6

Iodoheptane (22° F) 4.92

Iodohexane (20° F) 5.37

Iodomethane (20° F) 7.0

Iodotoluene (20° F) 6.1

Iron Oxide 14.2

Iso Butyl Alcohol 18.7-31.7

Iso Butyl Iodide 5.8

Iso Butyl Nitrate 11.9

Iso Butylamine 4.5

Iso Butyric Acid 2.7

Iso Butyronitrile 20.8

Iso Valeric Acid (68° F) 2.6

Iso-Butyl Alcohol (-112° F) 31.7

Iso-Butyl Alcohol (32° F) 20 .5

Iso-Butyl Alcohol (68° F) 18.7

Iso-Butyl Iodide (68° F) 5.8

Iso-Butyl Nitrate (66° F) 11.9

Iso-Butylacetate (68° F) 5.6

Iso-Butylamine (70° F) 4.5

Iso-Butyric Acid (68° F) 2.7

Iso-Butyronitrile 23.9- 20.8

Iso-Butyronitrile (75° F) 20.8

Iso-Iodohexadecane 3.5

Iso-Propyl Alcohol (68° F) 18.3

Iso-Propyl Nitrate (66° F) 11.5

Iso-Valeric Acid (68° F) 2.7

Isoamyl Valerate (19° F) 3.6

Isoamyl Acetate (68° F) 5.6

Isoamyl Alcohol (74° F) 15.3

Isoamyl Bromide (76° F) 6.1

Isoamyl Butyrate (68° F) 3.9

Isoamyl Chloracetate (68 F) 7.8

Isoamyl Chloride (64° F) 6.4

Isoamyl Chloroacetate 7.8

Isoamyl Chloroformate (68° F) 7.8

Isoamyl Lodide (65° F) 5.6

Isoamyl Propionate (68° F) 4.2

Isoamyl Salicylate (68° F) 5.4

Isoamyl Valerate (66° F) 3.6

Isoamylpropionate 4.2

Isobuthyl Resin 1.4-2.1

Isobutyl Acetate (68° F) 5.6

Isobutyl Alcohol (68° F) 18.7

Isobutyl Benzoate (68° F) 5.9

Isobutyl Bromide (20° F) 4.0

Isobutyl Bromide (68° F) 6.6

Isobutyl Butyrate (68° F) 4.0

Isobutyl Chloride (68° F) 7.1

Isobutyl Chloroformate (68° F) 9.2

Isobutyl Cyanide (74° F) 13.3

Isobutyl Formate (66° F) 6.5

Isobutyl Iodide (68° F) 5.8

Isobutyl Nitrate (66° F) 11.9

Isobutyl Rininoleate (70° F) 4.7

Isobutyl Valerate (66° F) 3.8

Isobutylamine (70° F) 4.5

Isobutylbenzene (62° F) 2.3

Isobutylbenzoate (68° F) 5.9

Isobutylene Bromide (68° F) 4.0

Isobutyric Acid (68° F) 2.6

Isobutyric Acid (122° F) 2.7

Isobutyric Anhydride (68° F) 13.9

Isobutyronitrile (77° F) 20.8

Isocapronitrile (68° F) 15.7

Isooctane 2.1-2.3

Isophthalic Acid 1.4

Isoprene (77° F) 2.1

Isopropyl Alcohol 18.3

Isopropyl Benzene (68° F) 2.4

Isopropyl Nitrate 11.5

Isopropylamine (68° F) 5.5

Isopropylether (77° F) 3.9

Isoquinoline (76° F) 10.7

Isosafrol (70° F) 3.4

Return to top

-J-

Jet Fuel (Jp4) (70° F) 1.7

Jet Fuel (Military Jp4) 1.7

Return to top

-K-

Kent Wax 6.5-7.5

Kerosene (70° F) 1.8

Kynar 2.0

Return to top

-L-

Lactic Acid (61° F) 22.0

Lactronitrile (68° F) 38.4

Lad Oxide 25.9

Lead Acetate 2.5

Lead Carbonate (60° F) 18.1

Lead Chloride 4.2

Lead Nitrate 37.7

Lead Nomoxide (60° F) 25.9

Lead Oleate (64° F) 3.2

Lead Oxide 25.9

Lead Sulfate 14.3

Lead Sulfite 17.9

Lead Tetrachloride (68° F) 2.8

Lime 2.2 - 2.5

Limonene (68° F) 2.3

Linde 5a Molecular Sieve, Dry 1.8

Linoleic Acid (32° F) 2.6 - 2.9

Linseed Oil 3.2-3.5

Liquified Air 1.5

Liquified Hydrogen 1.2

Lityium Chloride 11.1

Lonone (65° F) 10.0

LPG 1.6-1.9

Return to top

-M-

M-Bromoaniline (66° F) 13.0

M-Bromotoluene (137° F) 5.4

M-Chloroanaline (66° F) 13.4

M-Chlorotoluene (68° F) 5.6

M-Creosol 5

P-Cresol (24° F) 5.0

O-Cresol (77° F) 11.5

M-Dichlorobenzene (77° F) 5.0

M-Dinitro Benzene (68° F) 2.8

M-Nitrotoluene (68° F) 23.8

M-Sylene 2.4

M-Toluidine (64° F) 6.0

M-Xylene (68° F) 2.4

Maganese Dioxide 5-5.2

Magnesium Oxide 9.7

Magnesium Sulfate 8.2

Malachite 7.2

Maleic Anhydride (140° F) 51.0

Malolic Anhydride 51

Malonic Nitrile (97° F) 47.0

Mandelic Nitrile (73° F) 18.1

Mandelitrile (73° F) 17.0

Mannitol (71° F) 3.0

Margarine, Liquid 2.8-3.2

Melamine Formaldehyde (MF)

(MF) Molding Resin 5.5 - 6.0

(MF) With Alpha Cellulose Filler 7.2 - 8.2

(MF) With Asbestos Filler 6.1 - 6.7

(MF) With Cellulose Filler 4.7 - 7.0

(MF) With Flock Filler 5.0 - 6.0

(MF) With Macerated Fabric Fille 6.5 - 6.9

Melamine Resin 4.7-10.9

Menthol (42° F) 3.95

Menthol (107° F) 4.0

Menthonol (43° F) 2.1

Menthonol (110° F) 2.1

Mercuric Chloride 3.2

Mercurous Chloride 9.4

Mercury (298° F) 1.00074

Mercury Chloride 7-14

Mercury Diethyl (68° F) 2.3

Mesityl Oxide (68° F) 15.4

Mesitylene (68° F) 2.4

Mesitylene 3.4

Methal Cyanoacetate (69° F) 29.4

Methallmine (77° F) 9.4

Methane (-280° F) 1.7

Methane, Liquid 1.7

Methanol (77° F) 32.6

Methlene Idide 5.1

Methoxy-4-Methylphenol (60° F) 11.0

Methoxybenzene (76° F) 4.3

Methoxyethyl Stearate (140° F) 3.4

Methoxyphenol (82° F) 11.0

Methoxytoluene (68° F) 3.5

Methyl Acetate (77° F) 6.7

Methyl Acetophenoneoxalate (64° F) 2.8

Methyl Alcohol (-112° F) 56.6

Methyl Alcohol (32° F) 37.5

Methyl Alcohol (68° F) 33.1

Methyl Benzoate (68° F) 6.6

Methyl Butane (68° F) 1.8

Methyl Butyl Ketone (62° F) 12.4

Methyl Butyrate (68° F) 5.6

Methyl Chloride (77° F) 12.9

Methyl Chloroacetate (68° F) 12.9

Methyl Ether (78° F) 5.0

Methyl Ethyl Ketone (72 ° F) 18.4

Methyl Ethyl Ketoxime (68° F) 3.4

Methyl Formate (68° F) 8.5

Methyl Heptanol (68° F) 5.3

Methyl Iodide (68° F) 7.1

Methyl Kexyl Ketone (62° F) 10.7

Methyl Methacrylate (Cast) 2.7 - 3.2

Methyl Nitrobenzoate (80° F) 27.0

Methyl O-Methoxybenzoate (70° F) 7.8

Methyl P-Toluate (91° F) 4.3

Methyl Propionate (66° F) 5.4

Methyl Propyl Ketone (58° F) 16.8

Methyl Salicylate (68° F) 9.0

Methyl Thiocyanate (68° F) 35.9

Methyl Valerate (66° F) 4.3

Methyl 5 Ketocyclohexylene (68° F) 24.0

Methyl-1-Cyclopentanol (35° F) 6.9

Methyl-2 4-Pentandeiol (86° F) 24.4

Methyl-2-Pentanone (68° F) 13.1

Methylal (68° F) 2.7

Methylaniline (68° F) 6.0

Methylbenzylamine (65° F) 4.4

Methylcyclohexanol (68° C) 13.0

Methylcyclohexanone (192° F) 18.0

Methylcylopentane (68° F) 2.0

Methylene Iodide (70° F) 5.1

Methyleneaceloacetate (70° F) 7.8

Methylenemalonate (72° F) 6.6

Methylenephenylacetate (68° F) 5.0

Methylether, Liquid 5

Methylhexane (68° F) 1.9

Methylisocyanate (69° F) 29.4

Methyloctane (69° F) 30.0

Methylomine (21° F) 10.5

Methylphenyl Hydrazin (66° F) 7.3

Methylpyridine (2) (68° F) 9.8

Metnoxy-Four-Methyl Phenol 11

Mica 2.6-3.2

Mica 7.0

Mica (Glass Bonded) 6.9 - 9.2

Micanite 1.8-2.6

Mills (Dry Powder) 1.8

Mineral Oil (80° F) 2.1

Monomyristin (158° F) 6.1

Monopalmitin (152° F) 5.3

Monostearin (170° F) 4.9

Morpholine (77° F) 7.3

Return to top

-N-

N-Butyl Alcohol (66° F) 7.8

N-Butyl Bromide (68° F) 6.6

N-Butyl Formate (-317°F) 2.4

N-Butyl Iodide (77° F) 6.1

N-Butylacetate (19° F) 5.1

N-Butyricaid (68° F) 2.9

N-Hexane (68° F) 1.9

N-Methylaniline (68° F) 6.0

N-Pentane (68° F) 1.8

Naphthy Ethyl Ether (67° F) 3.2

Napthalene (185° F) 2.3

Napthalene (68° F) 2.5

Napthonitrile (70° F) 6.4

Napthyl Ethyl Ether (67° F) 3.2

Neon (68° F) 1.000127

Neoprene 6-9

Nitric Acid (14° C) 50.0 +/- 10.0

Nitroanisole (68° F) 24.0

Nitrobenzal Doxime (248° F) 48.1

Nitrobenzene (68° F) 35.7

Nitrobenzene (77° F) 34.8

Nitrobenzene (176° F) 26.3

Nitrobenzyl Alcohol (68° F) 22.0

Nitrocellulose 6.2-7.5

Nitroethane (68° F) 19.7

Nitrogen (336° F) 1.454

Nitrogen (68° F) 1.000580

Nitroglycerin (68° F) 19.0

Nitromethane 22.7-39.4

Nitromethane (68° F) 39.4

Nitrosodimethylamine (68° F) 54.0

Nitrosyl Bromide (4° F) 13.0

Nitrosyl Chloride (10° F) 18.0

Nitrotoluene (68° F) 1.96

Nitrous Oxide (32° F) 1.6

Nonane (68° F) 2.0

Nylon 4.0 - 5.0

Nylon Resin 3.0 - 5.0

Return to top

-O-

O-Bromotoluene (137° F) 4.3

O-Chlorophenol (66° F) 8.2

O-Chlorotoluene (68° F) 4.5

O-Cresol (77° F) 11.5

O-Dichlorobenzene (77° F) 7.5

O-Nitro Analine (194° F) 34.5

O-Nitrotoluene (68° F) 27.4

O-Toluidine (64° F) 6.3

O-Xylene (68° F) 2.6

Octadecanol 3.42

Octadecanol (136° F) 3.4

Octamethylcyclotetrasiloxane (68° F) 2.4

Octamethyltrisiloxane (68° F) 2.3

Octane (24° F) 1.061

Octane (68° F) 2.0

Octanone (68° F) 10.3

Octene (76° F) 2.1

Octyl Alcohol (64° F) 3.4

Octyl Iodide (68° F) 4.9

Octylene (65° F) 4.1

Oil, Peanut (52° F) 3.0

Oil, Almond (68° F) 2.8

Oil, Cotton Seed (57° F) 3.1

Oil, Grapeseed (61° F) 2.9

Oil, Lemon (70° F) 2.3

Oil, Linseed 3.4

Oil, Olive (68° F) 3.1

Oil, Paraffin (68° F) 2.2 - 4.7

Oil, Petroleum (68° F) 2.1

Oil, Pyranol (68° F) 5.3

Oil, Sesame (55° F) 3.0

Oil, Sperm (68° F) 3.2

Oil, Terpentine (68° F) 2.2

Oil, Transformer (68° F) 2.2

Oleic Acid (68° F) 2.5

Oleric Acid 2.4-2.5

One-Dichloroethane 10.7

One-Diethoxyethane 3.8

Opal Wax 3.1

Organic Cold Molding Compound 6.0

Oxygen (-315° F) 1.51

Oxygen (68° F) 1.000494

Return to top

-P-

P-Bromotoluene (137° F) 5.5

P-Chlorophenol (130° F) 9.5

P-Chlorotoluene (68° F) 6.1

P-Cresol (70° F) 5.6

P-Cresol (137° F) 9.9

P-Cymene (63° F) 2.3

P-Dibromobenzene (190° F) 4.5

P-Dichlorobenzine (68° F) 2.86

P-Dichlorobenzine (120° F) 2.4

P-Nitro Analine (320° F) 56.3

P-Nitrotoluene (137° F) 22.2

P-Toludine 3.0

P-Toluidine (130° F) 5.0

P-Xylene (68° F) 2.3

Paint 5-8

Palmitic Acid (160° F) 2.3

Paper (Dry) 2.0

Paraffin 1.9-2.5

Paraffin Wax 2.1-2.5

Paraldehyde (68° F) 14.5

Paraldehyde (77° F) 13.9

Parawax 2.3

Parrafin Chloride 2.0-2.3

Penanthiene (68° F) 2.8

Pentachlorgethane (60° F) 3.7

Pentadiene 1,3 (77° F) 2.3

Pentane (68° F) 1.8

Pentanol (77° F) 13.9

Pentanone (2) (68° F) 15.4

Pentene (1) (68° F) 2.1

Pentochlorethane 3.7

Perlite 1.3 - 1.4

Petroleum 2.0-2.2

Phenanthrene (230° F) 2.7

Phenathiene (68° F) 2.8

Phenathrene (110° F) 2.72

Phenetole (70° F) 4.5

Phenol (118° F) 9.9

Phenol (104° F) 15.0

Phenol (50° F) 4.3

Phenol Ether (85° F) 9.8

Phenol Formaldehyde Resin (PFR) 4.5 - 5.0

(PFR) With Asbestos Filler 5.0 - 7.0

(PFR) With Glass Fiber Filler 6.6 - 7.0

(PFR) With Mica Filler 4.2 - 5.2

(PFR) With Mineral Filler (Cast) 9.0 - 15.0

(PFR) With Sisal Fiber 3.0 - 5.0

(PFR) With Wood Flour Filler 4.0 - 7.0

Phenol Resin 4.9

Phenol Resin, Cumulated 4.6-5.5

Phenoxyacetylene (76° F) 4.8

Phentidine (70° F) 7.3

Phenyl Acetate (68° F) 6.9

Phenyl Ether (86° F) 3.7

Phenyl Iso Thiocyanate (68° F) 10.7

Phenyl Isocyanate (68° F) 8.9

Phenyl Urethane 2.7

Phenyl-L-Lropane (68° F) 2.7

Phenyl-One-Iropane 2.7

Phenyl-1-Propane (68° F) 1.7

Phenylacetaldehyde (68° F) 4.8

Phenylacetic (68° F) 3.0

Phenylacetonitrile (80° F) 18.0

Phenylethanol (68° F) 13.0

Phenylethyl Acetate (58° F) 4.5

Phenylethylene (77° F) 2.4

Phenylhydrazine (72° F) 7.2

Phenylsalicylate (122° F) 6.3

Phosgene (32° F) 4.7

Phosphine (-76° F) 2.5

Phosphorus (93° F) 4.1

Phosphorus Oxychloride (72° F) 14.0

Phosphorus Pentachloride (320° F) 2.8

Phosphorus Tribromide 3.9

Phosphorus Tribromide (68° F) 3.9

Phosphorus Trichloride (77° F) 3.4

Phosphorus, Red 4.1

Phosphorus, Yellow 3.6

Phosphoryl Chloride (70° F) 13.0

Phosphrous 4.1

Phtalide (166° F) 36.0

Phthalic Acid 5.1-6.3

Phthalide (74° F) 36.0

Pinacolin (62° F) 12.8

Pinacone (75° F) 7.4

Pine Tree Resin, Powder 1.5-1.8

Pinene (68° F) 2.7

Piperidine (68° F) 5.9

Plaster 2.5 - 6.0

Plastic Grain 65-75

Plastic Pellets 1.1-3.2

Plastic Sulphur, Unground 1.5

Platinum Catalyst 6.5 - 7.5

Poly Propylene 1.5

Polyacetal 3.6-3.7

Polyacetol Resin 2.6-3.7

Polyacrylic Ester 3.5

Polyamide 2.5-2.6

Polybutylene 2.2-2.3

Polycaprolactam 2.0 - 2.5

Polycarbonate 2.9-3.0

Polycarbonate Resin 2.9 - 3.0

Polyester Resin 2.8 - 4.5

Polyester Resin (Flexible) 4.1 - 5.2

Polyester Resin (Glass Fiber Filled) 4.0 - 4.5

Polyester Resin (Ridgid Cast) 2.8 - 4.1

Polyether Chloride 2.9

Polyether Resin 2.8-8.1

Polyether Resin, Unsaturated 2.8-5.2

Polyethylene 2.2-2.4

Polyethylene, Pellet 1.5

Polymide 2.8

Polymonochloro Pifluoroethylene 2.5

Polypropylene 1.5

Polypropylene Powder 1.25

Polypropylene, Pellet 1.5-1.8

Polystyrene Resin 2.4 - 2.6

Polystyrol 2.0-2,6

Polysulphonic Acid 2.8

Polytetra Fluoroethylene 2.0

Polyvinyl Alcohol 1.9-2.0

Polyvinyl Chloride 3.4

Polyvinylchloride Resin 5.8 - 6.8

Porcelain 5.0-7.0

Porcelain With Zircon 7.1 - 10.5

Potassium Aluminum Sulphate 3.8

Potassium Carbonate (60° F) 5.6

Potassium Chlorate 5.1

Potassium Chloride 4.6

Potassium Chloronate 7.3

Potassium Iodide 5.6

Potassium Nitrate 5.0

Potassium Sulfate 5.9

Potssium Chloromate 7.3

Pottasium Chloride 5.0

Propane (Liquid) (32° F) 1.6

Propanediol (68° F) 32 .0

Propanol (177° F) 20.1

Propene (68° F) 1.9

Propionaldehyde (62° F) 18.9

Propionic Acid (58° F) 3.1

Propionic Anhydride (60° F) 18.0

Propionitrile (68° F) 27.7

Propy Butyrate (68° F) 4.3

Propyl Acetate (68° F) 6.3

Propyl Alcohol (68° F) 21.8

Propyl Benzene (68° F) 2.4

Propyl Bromide (68° F) 7.2

Propyl Butyrate (68° F) 4.3

Propyl Chloroformate (68° F) 11.2

Propyl Ether (78° F) 3.4

Propyl Formate (66° F) 7.9

Propyl Nitrate (64° F) 14.2

Propyl Propionate (68° F) 4.7

Propyl Valerate (65° F) 4.0

Propylene Liquid 11.9

Psuedocumene (60° F) 2.4

Pulegone (68° F) 9.5

Pulezone (66° F) 9.7

PVC, Powder 1.4

Pyrex 4.8

Pyrex Glass 4.3 - 5.0

Pyridine (68° F) 12.5

Pyroceram 3.5-4.5

Pyrrole (63° F) 7.5

Return to top

-Q-

Quartz 4.2

Quinoline (77° F) 9.0

Quinoline (-292° F) 2.6

Return to top

-R-

Reburned Lime 2.2

Refractory (Cast) 6.7

Refractory (For Casting) 1.8 - 2.1

Resorcinol 3.2

Rice (Dry) 3.5

Rice Bran 1.4-2.0

Rouge 1.5

Rouge (Jewelers) 1.5 - 1.6

Rubber 3.0

Rubber (Chlorinated) 3.0

Rubber (Hard) 2.8

Rubber (Isomerized) 2.4 - 3.7

Rubber Cement 2.7-2.9

Rubber Chloride 2.1-2.7

Rubber, Raw 2.1-2.7

Rubber, Sulphurized 2.5-4.6

Ruby 11.3

Rutile 6.7

Return to top

-S-

Safrol (70° F) 3.1

Salicylaldehyde (68° F) 13.9

Salt 3.0 - 15.0

Sand (Dry) 5.0

Sand (Silicon Dioxide) 3 - 5.0

Santowax (70° F) 2.3

Selenium 6.1-7.4

Selenium 11

Selenium (482° F) 5.4

Selevium (249° F) 5.4

Sesame 1.8-2.0

Shellac 2.0-3.8

Silica Aluminate 2

Silica Sand 2.5-3.5

Silicon 11.0 - 12.0

Silicon Dioxide 4.5

Silicon Tetrachloride (60° F) 2.4

Silicone Molding Compound (SMC)

(SMC) (Glass Fiber Filled) 3.7

Silicone Oil 2.2-2.9

Silicone Resin, Liquid 3.5-5.0

Silicone Rubber 3.2-9.8

Silicone Varnish 2.8-3.3

Silk 2.5-3.5

Silver Bromide 12.2

Silver Chloride 11.2

Silver Cyanide 5.6

Slaked Lime, Powder 2.0-3.5

Slate 6.0-7.5

Smithsonite 9.3

Soap Powders 1.2 - 1.7

Sodium Carbonate 5.3 - 8.4

Sodium Carbonate (Anhyd) 8.4

Sodium Carbonate ($10h_20$) 5.3

Sodium Chloride 5.9

Sodium Chloride (Salt) 6.1

Sodium Cyanide 7.55

Sodium Dichromate 2.9

Sodium Nitrate 5.2

Sodium Oleate (68° F) 2.7

Sodium Perchlorate 5.4

Sodium Phosphate 1.6-1.9

Sodium Porchlorate 5.4

Sodium Sulphide 5

Sorbitol (176° F) 33.5

Soy Beans 2.8

Stannec Chloride (72° F) 3.2

Starch 3-5

Starch, Paste 1.7-1.8

Stearic Acid (160° F) 2.3

Stearine 2.3

Steatite 5.5 - 7.5

Styrene (77° F) 2.4

Styrene (Modified) 2.4 - 3.8

Styrene (Phenylethane) (77° F) 2.4

Styrene Resin 2.3-3.4

Succinamide (72° F) 2.9

Succinic Acid (78° F) 2.4

Sucrose 3.3

Sucrose (Mean) 3.3

Sugar 3.0

Sugar, Granulated 1.5-2.2

Sulfur 1.6 - 1.7

Sulfur Dioxide (-4° F) 17.6

Sulfur Dioxide (32° F) 15.0

Sulfur Monochloride (58° F) 4.8

Sulfur Trioxide (64° F) 3.1

Sulfurous Oxychloride (72° F) 9.1

Sulfuryl Chloride (72° F) 10.0

Sulphur(244° F) 3.5

Sulphur (450° F) 3.5

Sulphur Dioxide (32° F) 15.6

Sulphur Trioxide (70° F) 3.6

Sulphur, Liquid 3.5

Sulphur, Powder 3.6

Sulphuric Acid (68° F) 84.0

Sulfuric Acid (25° C) 100.0

Supphuric Oxychloride (72° F) 9.2

Syrup 50-80

Syrup Wax 2.5-2.9

Return to top

-T-

Tantalum Oxide 11.6

Tartaric Acid (68°F) 6.0

Tartaric Acid (14° F) 35.9

Teflon 2.0

Teflon (4f) 2.0

Teflon, FEP 2.1

Teflon, PCTFE 2.3-2.8

Teflon, PTFE 2

Tepineol 2.8

Terpinene (70° F) 2.7

Terpineol (72° F) 2.8

Tetrabromoethane (72° F) 7.0

Tetrachloroethylene (70° F) 2.5

Tetradecamethyltetradecamethylcycloheptasiloxan 2.7

Tetradecamethylhexosiloxane (68° F) 2.5

Tetradecanol (100° F) 4.7

Tetraethyl Amylenetetracarboxylate 4.40

Tetraethyl Hexane-1-Phenyl Tetracarboxylate (66° F) 5.9

Tetraethyl Pentane Diphenyl Tetracarboxylate (68° F) 2.7

Tetraethyl Propane Tetracarboxylate (66° F) 5.2

Tetraethyl Propylene Tetracarboxylate (66° F) 6.0

Tetraethyl Silicate (68° F) 4.1

Tetrafluoroethylene 2.0

Tetrahydro-B-Napthol (68° F) 11.0

Tetranitromethane (68° F) 2.2

Tetratriacontadiene (76° F) 2.8

Thallium Chloride 46.9

Thinner 3.7

Thioacetic Acid (68° F) 13.0

Thionyl Bromide (68° F) 9.1

Thionyl Chloride (68° F) 9.3

Thiophene (60° F) 2.8

Thiophosphoryl Chloride (70° F) 5.8

Thorium Oxide 10.6

Thrichloroethylene (61° F) 3.4

Thujone (32° F) 10.0

Tide (Loose From Box) 1.6

Tin Tetrachloride (68° F) 2.9

Titamium Tetrachloride (68° F) 2.8

Titanium Dioxide 110.00

Titanium Oxide 40-50

Titanium Tetrachloride (68 Degrees F) 2.8

Tobacco 1.6 - 1.7

Tobacco Dust (6% Moisture) 1.7

Toluene (68° F) 2.4

Toluene, Liquid 2.0-2.4

Toluidine (68° F) 6.0

Tolunitrile (73° F) 18.8

Tolyl Methyl Ether (68° F) 3.5

Totane (111° F) 5.5

Tourmaline 6.3

Trans-3-Hexene (76° F) 2.0

Transmission Oil (80° F) 2.2

Tribromopropane (68° F) 6.4

Tributylphosphate (86° F) 8.0

Trichlorethylene 3.4

Trichloroacetic Acid (140° F) 4.6

Trichloroethane 7.5

Trichloroethylene (61° F) 3.4

Trichlorololuene (70° F) 6.9

Trichloropropane (76° F) 2.4

Trichlorotoluene (69° F) 6.9

Trichloroxoluene 6.9

Tricosanone (176° F) 4.0

Tricresyl Phosphate (104° F) 6.9

Triethyl Aconitate (68° F) 6.4

Triethyl Aluminum (68° F) 2.9

Triethyl Ethanetricarboxylate (66° F) 6.5

Triethyl Isoaconitate (68° F) 7.2

Triethylamine (21° F) 3.2

Triethylamine (77° F) 2.4

Trifluoroactic Acid (68° F) 39.0

Trifluorotoluene (86° F) 9.2

Trimethyl Borate (68° F) 8.2

Trimethyl-3-Heptene (68° F) 2.2

Trimethylamine (77° F) 2.5

Trimethylbenzene (68° F) 2.3

Trimethylbutane (68° F) 1.9

Trimethylpentane 1.9

Trimethylpentane (68° F) 2.9

Trimethylsulfanilic Acid (64° F) 89.0

Trinitrobenzene (68° F) 2.2

Trinitrotoluene (69° F) 22.0

Triolein (76° F) 3.2

Triphenylmethane (212° F) 2.3

Tripolmitin (140° F) 2.9

Tristearin (158° F) 2.8

Turpentine (Wood) (68° F) 2.2

Two-Dichloroethane 10.7

Return to top

-U-

Undecane (68° F) 2.0

Undecanone (58° F) 8.4

Urea 5-8

Urea (71° F) 3.5

Urea Formaldehyde (Uf Degrees F)

Urea Formaldehyde (Cellulose Filler) 6.4 - 6.9

Urea Resin 6.2-9.5

Urethan (121° F) 14.2

Urethane (74° F) 3.2

Urethane Resin 6.5 - 7.1

Return to top

-V-

Valeraldehyde (58° F) 11.8

Valeric Acid (68° F) 2.6

Valeronitrile (70° F) 17.7

Vanadium Oxybromide (78° F) 3.6

Vanadium Oxychloride (78° F) 3.4

Vanadium Sulfide 3.1

Vanadium Tetrachloride (78° F) 3.0

Vaseline 2.2-2.9

Veratrol (73° F) 4.5

Vinyl Alcohol Resin 2.6-3.5

Vinyl Butyral 3.3 - 3.9

Vinyl Chloride (Acetate) 3.0 - 3.1

Vinyl Chloride (Flexible) 3.5 - 4.5

Vinyl Chloride (Ridgig) 2.8 - 3.0

Vinyl Chloride Resin, Hard 5.8-6.4

Vinyl Chloride Resin, Soft 2.8-4.0

Vinyl Ether (68° F) 3.9

Vinyl Formal 3.0

Vinyllidene Chloride 3.0 - 4.0

Vycor Glass 3.8

Return to top

-W-

Water 4-88

Water (32° F) 88.0

Water (68° F) 80.4

Water (212° F) 55.3

Water (390° F) 34.5

Water (80° F) 80.0

Water (Steam) 1.00785

Wax 2.4-6.5

Wheat Flour 3.0 - 5.0

Wheat Flour (Dry Powder) 1.6

White Mica 4.5-9.6

Wood, Dry 2-6

Wood, Pressed Board 2.0-2.6

Wood, Wet 10-30

Return to top

-X-

Xylene (68° F) 2.4

Xylene, Liquid 2.2-2.6

Xylenol 17

Xylenol (62° F) 3.9

Xylidine (68° F) 5.0

Return to top

-Y-

-Z-

Zinc Oxide 1.7-2.5

Zinc Sulfide 8.2

Zircon 12.0

Zirconium Oxide 12.5

Zirconium Silicate 5.0

Return to top

-Numeric-

1, 2-Dichloroethane (77° F) 10.7

1-Diethoxyethane (75° F) 3.8

1-Heptene (68° F) 2.1

1-Octanol (68° F) 10.3

2-Methyl-1-Propanol (77° F) 17.7

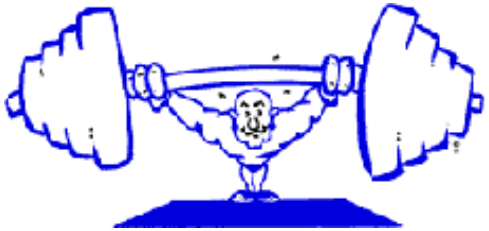
3 Dimethyl-2-Butanone 13.1

3-Chloro-1, Dihydroxyprone (68° F) 31.0

3-Dimethyl-2-Butanone (293° F) 13.1

Return to top

Phone (415) 808-AHOY Fax (415) 808-2470 Email us
©1998-2005 Clipper Controls Inc. All rights reserved.
For more information please read our Legal Terms and Notices.



Compression In Mastering

Part I - Technical Guidelines for the Uses of Compressors
We define compression, and give technical guidelines for its use in mixing and mastering.

Part II - The Perils of Compression
The esthetic ramifications of compression, the current loudness race, the bleak future of DVD audio quality, and more to the point, what to do about it!

Part III - Tools To Help Keep Us From Overcompressing
Including an honor roll of CDs that sound very good and aren't overcompressed.

Mastering engineers discuss compression in mixing

Part I
Technical Guidelines for Use of Compressors

Here's an E-mail discussion I had with a U.K. engineer who discovered our web site.

> *I had a look at your web page and what a page! Exactly a page I have been looking for! I can't wait until you get it ready. I may say, could you add something about using compression, like how much to do by yourself and how much to leave to the mastering lab?*

First, let's define:

Com-pres'-sion

Email to Friend

1) Reduction of audio dynamic range, so that the louder passages are made softer, or the softer passages are made louder, or both. Examples include the limiters used in broadcasting, or the compressor/limiters used in recording studios.
2) Digital Coding systems which employ data rate reduction, so that the bit rate (measured in kilobits per second) is less. Examples include the MPEG (MP3) or Dolby AC-3 (now called *Dolby Digital)* systems.

This article is about *compression,* using concept #1 above. It's not good to refer to two different concepts with the same word, so please encourage people to use the term "Data Reduction System" or "Coding system" when referring to concept #2.

Now, for the basic two rules:

Rule #1: **There are no rules.** If you want to use a compressor/limiter of any type, shape and size in your music, then go ahead and use it.

Rule #2: **When in doubt, don't use it!**

How can you tell when you have *enough* compression*?*

- Discussing sound in print is like describing colors to a blind person, but let me try. Here's a simplistic example... supposing there are two sonic qualities of music, one called *punchy,* the other *smooth*... Let's say that some music sounds better *punchy*, other music sounds better *smooth.* Let's also assume for this example that you can achieve *punchy* or *smooth* sound through different amounts and types of compression, or not using compression at all.
- In general, *try to avoid overall compression in the mix stage if:*
  - you're mixing *punchy* music (the type of music that needs *punch*), perhaps using some individual compression on certain instruments or singers--and the mix already sounds *punchy* (good) to you.
  - you're mixing *smooth* music, and your mix already sounds *smooth.*
  - you play a well-recorded CD of similar music, and your CD in the making already sounds good (or better than) the CD in the player.
  - your music already seems to accomplish the sound you are looking for.

We'll discuss compression *esthetics* in more detail in Part II.

Technical reasons to avoid overall compression on your album
Save decisions on overall compression and individual tune equalization for an expert CD mastering house because:

1. The mastering house will have a more appropriate compressor with the proper attack, ratio, and release times exactlyright for your music. If you mixed to digital tape, they will probably use a 24-bit digital compressor for the purpose.
2. They will likely be more experienced than you about the compromises, advantages and disadvantages of applying overall compression.
3. The mastering house can program that compressor with precision, adjusting it optimally for each tune in question. You're working out of context (without having the perspective of the entire album) by attempting these sorts of decisions during mixing.
4. The mastering house will be able to monitor your "CD in the making" using a calibrated monitoring system so that they know exactly how loud your "CD in the making" is compared to other CDs of similar music. For more information on *loudness,* see my article Level Practices in Digital Audio.
5. A good mastering house will be able to do all of this in a non-destructive, non-cumulative manner. In other words, after making a reference CD, they will be able to undo anything you are unhappy with, whether it be compression, EQ or levels. Whereas, most digital audio editing stations can only perform destructive EQ or compression, only with 16-bit word length, with a consequent loss of resolution as long internal words are either dithered (resulting in a veil if further processed), rounded (slightly better than truncated), or truncated to 16 bit. For further information,see my article The Secrets of Dither.
6. For the same technical reasons, it is not a good idea to use a digital compressor (or any digital processor) on your material before sending it for mastering. If you do feel the need to nsert one of these boxes, for example, to give a demo CD to your client, be sure to also make a non-processed version to prepare for the mastering house. It is likely that the mastering house will have a fresher-sounding, more effective approach at polishing your material, and it's self-defeating if they have to try to undo what was done.
7. If you apply overall compression to your music, and your choice of compressor was wrong (e.g., the compressor you chose caused subtle pumping or breathing, loss of transients, loss of life or liveliness, etc. These are typical symptoms of "compressor misuse" on tapes I have received), the mastering house will have a difficult or impossible time attempting to undo the damage. As I've mentioned, mastering is like whittling soap; it is hard to undo compression. However, I do have some tricks up my sleeve (grin) that can restore some life to squashed tapes.

Part II
The Perils of Compression
or
The Ghost of CD Past

Introduction

24-bits, 96kHz, multichannel sound. These are some of the exciting features of the upcoming super audio disc. Before we can use these new capabilities to their fullest, we must learn not torepeat our past mistakes. Some of our engineering practices with the CD have done a serious disservice to the consumer. This article looks at one of those practices--overcompression of dynamics, and makes some recommendations that will turn the DVD (and new 16-bit CDs) into the true media of the future.

Dynamic Range - The Ups and Downs of Music

Before we can study the art of compression, we must learn to appreciate the power of music's dynamic range. How does music grab our interest? For short periods (about the length of a "single" played on the radio or in the disco), power and loudness can grab our initial attention. But at home, *variety of dynamics maintains our interest* for long periods of time. Good music written for a long-term musical experience contains a judicious mixture of variety and similarity in dynamics. A production which is relentlessly loud (or relentless in its sameness) can become boring very fast. At the age of 10, I learned the lesson of Franz Joseph Haydn's *Surprise Symphony,* the first composition to teach the importance of dynamic contrast. Musical genres that depend on constant sameness become old very fast. Disco died because it became boring, and I'm convinced that overcompression (which eliminates dynamic contrast) contributed to its death, by creating a continuously loud, boring dynamic. I wonder if the current slack in music sales is related to overcompression and its tendency to give everything a monotonous loudness--is the public voting against compression with its pocketbook?

Any genre that does not grow in musicality will quickly die, and dynamic contrast plays a big role in musicality. Today's Rap music has taken a 250-year-old lesson from classical composition, by beginning to incorporate a melodic and harmonic structure. The genre can further grow and avoid sounding tiresome by expanding its dynamic range, adding surprises. Silence and low level material creates suspense that makes the loud parts sound even more exciting. Five big firecrackers in a row just don't sound as exciting as four little cherry bombs followed by an M80. This is what we mean by dynamic range. Radio, TV and Internet distribution are currently too compressed to transmit the joy of wide dynamic range, but it sure turns people on at home, and also in the movie theater.

Films provide an ideal framework to study the creative use of dynamic range. The public is not consciously aware of the effect of sound, but it plays a role in a film's success. I think the movie *The Fugitive* succeeded because of its drama, but despite an aggressive, compressed, fatiguing sound mix. From the beginning bus ride, with its super-hot dialog and effects, all the crashes were constantly loud and overstated, completely destroying the impact of the big train crash. I can hear the director shouting, "more more more" to the mix engineers. Haven't

they heard of the term "suspense?" In contrast, the sound mix of *Titanic*, is a masterpiece of natural dynamic range. The dialog and effects at the beginning of the movie are played at natural levels, truly enhancing the beauty, drama and suspense for the big thrills at the end. Kudos to director James Cameron and the Skywalker Sound mix team for their restraint and incredible use of dynamic range. That's where the excitement lies for me.

Compressors as Tools To Manipulate Dynamic Range

Compression is a tool; when used by skilled hands, it has produced some of the most beautiful recordings in the world.A lot of music genres are based on the sound of compression, from Disco to Rap to Heavy Metal. And a skilled engineer may intentionally use creative compression to paint a mix and form new special effects; this *intended distortion* has been used in every style of modern music. This is analogous to the work of the greatest visual artists; many painters are quite capable of producing a natural-looking landscape, but have abandoned that medium to create abstractions which at first glance look like the fingerpaint work of a six-year-old. But a skilled observer realizes what the master artist is communicating. The keys here are *intent and skill.* Too often, in music, unskilled *compression* can *squash* sound, remove the life, vigor and impact, and replace it with boring mush. Many engineers don't know what uncompressed, natural-sounding audio sounds like. It actually takes more work and skill to make a natural-sounding recording than an artificial one. In audio as in the visual arts, first learn to paint naturally; then and only then can you truly understand the art of creating distortion. Learn where compression is useful, and where it does a disservice to the music. A compressed production may sound good on a boombox, but when reproduced on a high-fidelity system, it can sound overbearing and ultimately lifeless. That's why we may need to separately mix "single" and album cuts.

Compressors are commonly used in recording (tracking), mixdown, and in mastering. Everyone has his own style of working with compressors and there are no rules. However, before you make *your* rules, start by working *without any* compressors! This learning process will teach you to make better-sounding music later on; the compressor becomes a tool to handle problems, not a crutch or substitute for good recording techniques. First, learn about the natural dynamics and impact of musical instruments, then begin to alter them with compressors (which can include using compression to create special effects). Every 5 years or so, give yourself a reality check...try making a recording or mix with little or no compression. You'll rediscover what I call the *microdynamics* of music. It's a real challenge, but a refresher course may point out that *less compression* will buy you a more open, more musical sound than you've previously been getting.

Watch For These Compression-related Pitfalls

Tracking

When tracking vocalists (who have a habit of belting now and then), a well-adjusted compressor can sound reasonably transparent, and most engineers agree the cure is better than the disease. But watch out for a "closed-in" sound, "clamping down" when the vocalist gets loud, or loss of clarity or transparency. Compare IN versus BYPASS before committing to tape. Match levels to make a fair comparison. If you notice too much degradation, maybe it's time to consider a different compressor or change the settings you are using. The sound should be open and clear... remember that no amount of equalization in the mixdown can substitutefor capturing a clear sound quality in the first place. This is true for all the lead instruments, including trumpets and electric guitars. If possible, put the uncompressed sound on a spare track--it may save your life. If there's any "rule," most engineers would agree to save the decision on drum and percussion compression until mixing. There are always exceptions--every piece of music is unique. Just remember, you cannot undo the damage of overcompression, so be careful about compression during tracking.

Mixdown

There are two possible places to apply compressors during mixdown. The first is on individual instruments or stereo pairs; the other is on the console mix buss.

For individual tracks: The same precautions apply to the use of compressors in mixing as in tracking. Start fresh each time--free yourself of preconceptions. Although you compressed the bass on 9 out of the last 10 albums, maybe this time you won't need a compressor. Each musician is an individual. In general, the better the bass player, the less compression you will need to use, and the greater the chance that compression will "choke up" his sound. Get to know the sound of your instrumentalists. What is your mixing philosophy? Are you trying to capture the sound of your instrumentalists or intentionally creating a new sound? While more and more music is created in the control room, it's good practice to know the real sound of instruments; learn how to capture natural sound before moving into the abstract.

In pop music, compressors are often used to create a tighter band sound, making the rhythm instruments sit in a good, constant place in the mix. But when misused or overused, compressors can take away the sense of natural breathing and openness that makes music swing and sway. Thus, I recommend that during mixing, after you've inserted a few compressors on certain instruments (e.g., the bass, rhythm guitar, vocal) and listened for a while, try comparing with the compressors bypassed (automation makes that process easy; store two fader snapshots so you can switch between them). Many times you'll find the compression was hurting the mix, and not helping the sound, by losing the subtleties of the musician's performance. Learn the negative as well as the positive effects of compressors by proving to yourself that you really needed the compressor, or that degree or type of compression. The process of refining a mix should always include revisiting your compression (and EQ) settings and questioning your work. Most music these days is recorded in

overdubbed sections, but some performances are still captured at once. An engineer once told me that the best sound he got was the monitor mix on the recording day. By the time he got through slicing and dicing and remixing, all the life was taken out of it (what I call *the loss of microdynamics)*. So remember the sound you got during the recording...did you lose the magic in the mix?

Avoid Wimpy Loud Sound

Some of you may say that my conservative advice only applies to acoustic genres like country music or jazz. However, Rock and Roll music is often a casualty of compressor abuse. I receive rock mixes from well-meaning engineers that should be getting louder and louder and reach a climax, but which have lost their intensity, producing *wimpy loud sound*. The dynamics of choruses and verses are reversed. Instead of the chorus sounding lively and dramatic, it's been pulled back. When you go against the natural dynamics of the music, the results are less pleasant, and *less exciting.* I strive to put that kind of life back into sound during the mastering, and my clients are delighted by the results. You can make the mastering engineer's job easier. When you mix rock, listen closely to the climaxes; is it possible that you are killing the music with your compressor? This is a very common problem and only the most skilled mix engineers are able to overcome it, maintaining excitement all the way to the highest peaks. Many mix engineers have trouble handling the duality of rock; they find compressors give them power at mid levels, but they have trouble handling the climaxes; they want them to be loud, but can't seem to do it without overload or overcompression. If you're having those kinds of troubles, don't despair. Mastering engineers dislike getting squashed material, because the damage is really hard to fix (though some of the tools I apply are pretty darn effective). Better to send material that's mixed well and powerfully at the mid levels but at the high levels is not squashed. Even if the climaxes don't sound loud enough, consider it a "work in progress." Let the mastering engineer take it to the next level of performance. Using specialized and unique tools, I can remix your material, giving it the punch it needs at mid levels and strength and volume at high levels.

Buss processors

Let me be a bit dogmatic. Reserve "buss compression" or "overall compression" for the mastering stage. As a mastering engineer, I can unequivocally say that the most frequently abused compressor is in the console mix buss. Lately I've been receiving a lot of mixes that have been squashed to death by unintentional misuse or overuse of buss compression.

Overuse of buss compression

Properly-used buss compressors *can* make music sound louder and more powerful, possibly without deteriorating its character, but is the console mix buss the right place to be working on the loudness character of your music? *Absolutely not.* If you already have a great-sounding mix without buss compression, then don't add buss compression just to "beef it up." More often

than not, the buss compressor you have available will *take away the life of* your music. Turn up the monitor level if the music doesn't sound loud enough! By all means, leave questions of loudness character out of the mixing process and save them for the mastering stage, where they can be dealt with correctly and effectively (more about that in a moment).

Recently a potential client told me that he was using a little bus compression on his mix. I asked him why he wasd oing that. He said, "because I think the levels are a little too low." Please don't compress for that reason; if the "levels are too low," then turn them up! The only possible reason to bus compress during a mix is because "it sounds better" to you. I hope that in this article I have provided some useful ways of how you can judge that the mix really "sounds better" before you overall compress. Mixing "to the compressor" is also a bit like cheating. Your whole judgment becomes geared to what the compressor is doing rather than the act of mixing itself. When in doubt (and even when not in doubt), mix two versions, one with and one without bus compression, and send both to the mastering house. You may be surprised which version the mastering engineer chooses, and which one sounds better after mastering. Also remember, that not all compressors sound that good. The mastering house might be able to employ a digital compressor like the Weiss, which uses 40 bit floating point internal processing, double sampling, has extraordinary attack and release time flexibility. Or an analog compressor, like the Manley Vari-Mu. Both of these are examples of specialized mastering compressors with extraordinary sound.

One possible proper use of a buss compressor is to "tighten" a mix when individual compressors couldn't do the job. However, be careful not to squash the mix. Tuning a buss compressor is an art born of technical knowledge and experience. As always, compare IN and OUT very carefully, and don't be afraid to patch it OUT if that sounds better. Buss compression causes all the instruments to be modulated by the attack and transients of the loudest instrument. A rim shot or cymbal crash can take down the reverberation and the sound of all the other instruments. There are very few console compressors that are capable of doing buss compression without screwing up transparency, transient response or musical dynamics. Excellent circuit design is required, as well as attack and release characteristics idealized for the job of buss compression. Very few outboard compressors can handle that job. If you want to tighten the mix, first try using submix compression on the rhythm section alone. That way you won't abuse the clarity of the drums and vocal.

Stop Emulating Squashed CDs

Many mixing engineers compare their mixes against already-pressed CDs, but becareful what you choose as a standard. Ironically, mastered CDs often do not sound like what comes out of the mix, so how can you emulate something which can only be done post-mix? What you really need is to hear the sound of a good mix *before* it was sent for mastering. But since that's not available, choose from the plenitude of pop records that have been well-mixed and mastered, as listed

in the CD Honor Roll. When choosing a reference album, don't pick it because it's "hotter" than everyone else; instead, listen for impact, clarity, transparency, ambience, warmth, space, depth, beauty, openness, naturalness, and (sometimes) punch. But "punch" is an ambiguous term; any so-called "mastering engineer" with a $2000 processor can give an album a kind of "punchy" sound, but often sacrificing all the other character that makes music worth listening to. Remember this: when two CDs are presented at equal loudness, nine out of ten musicians prefer the sound of an uncompressed presentation to a compressed one. For the first few seconds, a louder presentation may grab you, but relentless sound quickly becomes fatiguing. Many of today's compact discs have already exceeded the loudness limit--the level above which the sound quality goes down hill while the sound "quantity" goes up. You can't get something for nothing.

If You Can't Make It Sound Good, Make it Loud?

Contrary to some people's beliefs, mastering is *not* supposed to be the process of making a record hotter than the competition. Mastering should be the process of making a record *better* than the competition. Currently there is a lot of pressure on mastering engineers to make a record hotter than its neighbor. I'm really surprised that more recording engineers are not up in arms about how mastering engineers (on producer's orders?) are ruining their recordings. I'm quite flattered that one recording engineer called me the first mastering engineer to make his recordings sound *better.* If you're a mastering engineer, wear the red badge of courage; strive for good sound, even if you have to sacrifice a few dB of loudness.

Clients often complain when they haveto turn the volume control up when switching CDs. Why aren't they complaining that they have to turn the volume down when a hot disc comes on? One client told me that she loved the sound of her master, but her test CDR was not hot enough when played in rotation on the CD changer in a local bar. This upset me, because it turned out she was comparing her CD against rather trashy-sounding competition and I didn't want to trash up her sound. In the bar, you can't tell the difference in sound quality, but we're making CDs to sound good at home. I told her that she would sacrifice quality if I made her CD any hotter. Ironically, it was already a hot CD, by the standards of last year and the year before, but obviously not this year! Oh, by the way, eventually I compromised, using my best skills to raise the volume on her CD slightly without sacrificing too much in sound quality, but it saddened me that this had to occur. Her CD would have sounded better if it were not as compressed.

The PARTY Button

CD changers present a real problem in client education. I had to tailor the apparent loudness of this client's CD to work with 5 other hand-picked CDs, but I could pick dozens more that are much louder or softer than hers. We have to teach clients that CDs will always differ in volume and that a CD changer is not a

radio station. The restaurant CD changer really needs a live DJ, but that's not practical. The solution: put a compressor in the restaurant, the car, and the jukebox, and reserve quality listening for the home, without compression. We should lobby manufacturers to put a compressor button on receivers and future DVD players; with DSP technology this is a snap. They could label it the **Party Button**! There should be three compression settings--for background listening at low levels, for moderate listening in a noisy environment, and for parties, whereyou want to switch discs without adjusting the volume control. Panasonic and Sony will sell a million of them, and we engineers will be eternally grateful! The button may be misused by ignorant consumers, but no more than the loudness button I find permanently pushed in 6 out of 10 homes.

Save it for the Mastering

When you're through mixing, your recordingis a diamond in the rough, it's not supposed to sound like a "record" until it is mastered. Just make sure the mixes sound great and wait for mastering to add any post-mix processes. When you make copies for your clients, if they have any problems, tell them to turn up their monitors and wait for the mastering. Don't be tempted to use so-called "mastering processors" ("maximizers")before the mastering begins. As mastering engineers, we want to receive the cleanest, highest resolution, unprocessed, original mix tape or disc. In mastering, the individual songs will be levelled (not "normalized"), and elements of your music defined and clarified, turning your record into a work of art. The mastering engineer objectively looks at every song in context in a controlled, familiar acoustic environment, using superior tools, monitoring, experience, and artistry.

The Vicious Circle of Loudness Envy

The practice of overcompression is partof a vicious circle of loudness envy. Sadly, the current crop of CDs is louder and even more squashed than its predecessors because few people have stood up to fight the problem. Participants in this unwitting vicious circle include mix engineers, musicians, producers, mastering engineers, and radio program directors, but the problem is introduced during the mastering process. Many people blame the program director for the problem, but I think we're all partly at fault. Regardless of the cause, we all have to participate in a solution before our music turns into mush.

Program directors should realize that the sound on their office CD player has little to do with the disc's on-air quality. PD's may think the loudest record they hear is the best, but they forget that when it gets to the air, on-air processors will squash it (drop the volume) more than other records. Producers are afraid that the PD will reject their record if they have to turn up the volume. But by now, hot CDs have put the PD's volume control at the bottom of its travel, so where dowe go from here? Well, let's get the program directors to make decisions on the merits of the music, not on its loudness character. One way to solve that is to install a compressor in the PD's audition system, one that'll squash music as much as his radio station does. We could call it **The**

Ecumenical Button. Send me suggestions to see how we can get this one done (no kidnapping, please).

From The Sublime to the Ridiculous

roducers don't seem to like making a CD that's even a little softer than the competition, so each succeeding CD is often a little bit hotter. Just how much hotter can CDs get? I can cut a CD that's **16 dB louder** than the ones we made in the early 90's, before digital limiters became popular, but it'll look like a square wave and soundlike audio cream of wheat! Imagine the consumer problems caused by large variations in loudness--switching CDs has literally become Russian Roulette, shooting out our speakers and ears! But ultimately, your hot CD doesn't get any louder for the public; they just turn their monitor down, and scream in disgust at the increasing range they have to move the knob when they change CDs. In addition, sound quality is suffering by an unjustifiable demand for hotter CDs. A fellow mastering engineer reminds me that in the early days of CDs, we didn't have any pressure to make them hotter (because there was little competition), and early pop CDs had good, open sound. They're much softer than current CDs, but if you turn up your volume control you'll see their dynamics are much better-sounding. Why do we have to go backwards in sound quality? We mustn't repeat this mistake with the DVD. Part II of my article on Levels discusses 21st century solutions to this problem.

Let's review the basics. The loudness war may have begun with analog records, but the current problem is many decibels worse than it was in analog. LPs were mixed largely with VU meters, which created a degree of monitoring consistency, but today's peak level meters give entirely too much more room for mischief, and today's digital limiters provide the tools to do the mischief. The net result: great consistency problems in CD level. The peak meter is currently being seriously misused. Remember that *the upper ranges of the peak meter were designed for headroom, not for level.* A compressed piece of music peaking to -6 dBFS can sound much louder than an uncompressed work peaking to 0. Mixing and mastering engineers, use compression for creative purposes, but why not master the CD at a lower peak level, and monitor at the same gain you used for your last CD? There's no reason to fill up all those bits if the CD sounds loud enough. Or, useless limiting if you insist on peaking to 0 dBFS. Too many producers are unskilled meter readers; they seem to need all those lights flashing. Try working at a fixed monitor level, with the meter hidden from view. It'll be a very educational experience.

It's come to the point where mastering engineers should think about working differently. In Part II of my article on Levels, I discuss the 21st century solution to this problem, because the future of our DVD Audio is at stake. Only education can stop this vicious spiral. It's time to fight for quality, not quantity. Sound lower than XYZ hit record?Turn up the monitor!

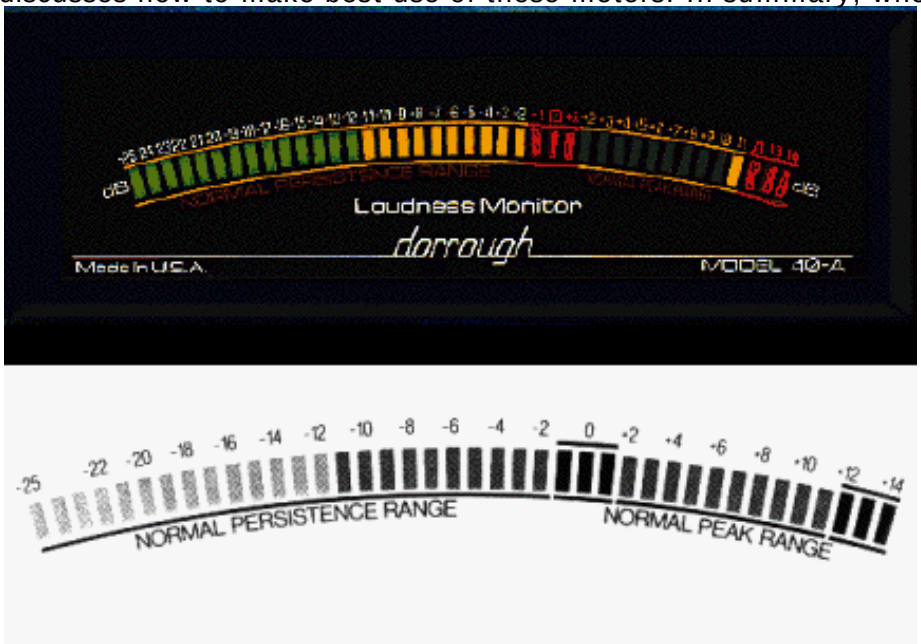
Part III
Tools To Help Keep Us From Overcompressing

How can you tell when *proper compression* is becoming *overcompression?* If you don't have good monitors, it's not easy to know when you've crossed the line. The first sign that you're probably going too far is if you start playing with the compressors simply to achieve overall program "loudness" rather than to help you make a great sound. Remember, the mix room is for mixing. If you know you've got a great sound when the monitor is turned up, then all is ok, the mastering engineer can do the rest. The second sign you're probably overcompressing is if you find you're leaning too much on the compressors to make your mix. A program shouldn't mix itself. It takes a lot of work to mix, and depending on the compressors to do that work for you will probably result in a squashed, lifeless sound.

Here are some practical tools you can use to make better-sounding recordings to send to the mastering studio.

1) As mentioned above, install a *Dolby-level-calibrated monitor control, connected via a single, high-quality D/A converter.* Visit Part II of my article on Levels for more details.

2) *Metering*. Meters with combined peak and average readings are some of the best protections against overcompression. The Dorrough meter is a good example, as are meters from DK and Pinguin. Part II of my article on Levels discusses how to make best use of these meters. In summary, when mastering, try to keep the "average level" on this meter's average scale from exceeding 0, with occasional "high average levels" to +3 (equivalent to +3 on a VU meter). If you do, you will have obtained approximately a 14dB peak



to average ratio. (Peak to average ratio is also known as *Crest Factor).* The meter is also a good visual aid for visiting producers.

3) *Monitoring.* A clean, high-headroom monitor system is essential. If your monitor speakers or amplifier saturate, how can you possibly tell if your material is saturating?

CD Honor Roll

I'd like to thank all my friends on the Sonic Solutions Maillist, The Mastering Webboard**,** the Pro Audio List, and many of my fellow mastering engineers, for support and ideas. We've been preaching to the converted. Now it's time to transmit these points to the rest of the world.

These HTML documents are Copyright 1995-2003, Digital Domain, Inc. The following are trademarks of Digital Domain, Inc.: Digital Domain, Digi-nary. These documents may not be reproduced in any manner without the permission of the copyright owner. We invite the audio and music community to link to this web site, which will be periodically revised. This page last revised December 2003.

Recording and Measurement Microphones

Rod Elliott (ESP)

PCBs for the discrete mic preamp and a complete housing are available - click on the image for details

Introduction

The purpose of this article and small group of projects is firstly to introduce the electret microphone into the ESP projects lineup, and secondly to allow the reader to build a microphone that although uncalibrated, can be used to great effect as a measurement mic with any loudspeaker project, or for very high quality recordings.

Traditionally, measurement microphones are calibrated, so that the exact output level for a given SPL is known, and so that the frequency response is predictable and accurate. These are fine goals, but few hobby speaker builders can afford (or can justify) the expense of a fully calibrated measuring set, or even the microphone by itself.

The measurement mic project here is not calibrated for level or response, but relies on the reasonably predictable performance of electret microphone units. These are readily available, very cheap (less than $5.00 in any currency), and are usually surprisingly good.

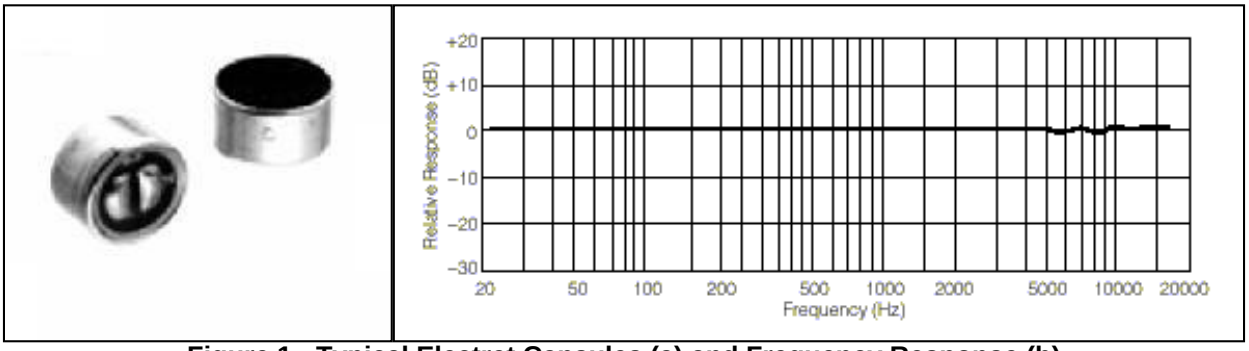


Figure 1 - Typical Electret Capsules (a) and Frequency Response (b)

Figure 1 shows what they look like, and a typical frequency response (Panasonic capsules and response are shown - but this is also typical of many others). This is extremely good performance, and is fairly close to what you can expect.

Additional microphone projects will be presented when I get the opportunity, with some hopefully interesting variations to the concept of a basic microphone as described here. Most applications will require a directional mic, and these will be discussed in the next article. Unfortunately, the casing is much more complex and critical for a directional mic than for an omnidirectional version - fortunately, measurement mics need to be omnidirectional, so that limitation is not a problem ... yet.

Powering

Traditionally, electret mics are powered from a 1.5V cell, in a very simple circuit as shown in Figure 2. Shown is the schematic for a Radio Shack "Boundary" microphone, and this is actually more complex than most - the inductor is not usually used. Like all such simple circuits, this has some very real disadvantages.

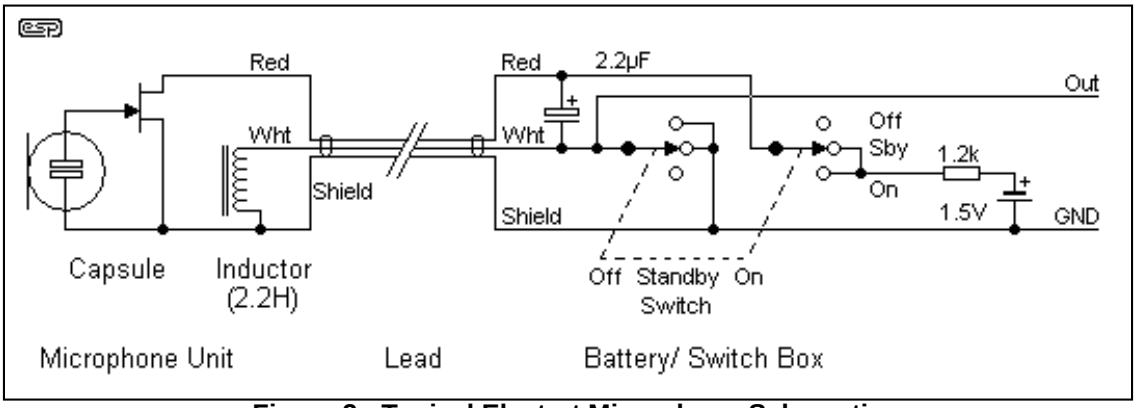


Figure 2 - Typical Electret Microphone Schematic

The disadvantages of the standard method (and commercial electret microphones in general) are ...

1 Output impedance is relatively high (typically about 1k to 5k)
2 Signal output is limited (relatively low sensitivity)
3 Noise is relatively high
4 Sound level handling ability is low (typically <90dB SPL)
5 They are normally available from retail outlets very cheaply

These have caused such mics to have a poor reputation, however, with some additional work excellent results can be achieved. The first objection is easily resolved with an opamp to buffer the output, making sure that the output impedance is kept low. It is easy to achieve an impedance of 100 ohms or so, and this will drive any mixer.

The second objection is resolved by increasing the supply voltage. 1.5V is simply too low to be useful, and a supply of 9V or more is recommended. This in turn solves the next two problems as well, since with more signal from the mic the noise contribution is lower, and a higher supply voltage allows much more output voltage before distortion.

The final objection remains, and we can't change that, but we can use it to our advantage. The idea that something so cheap is capable of excellent performance is somewhat disconcerting - the expectation is that if it is very cheap, it cannot have high performance. This is actually not the case at all.

I have used a modified hyper-cardioid electret mic (which cost less than $50 at the time) and achieved excellent results for recording voice announcements. In many cases, the quality was better than that from a professional studio, even though the recordings were done in an ordinary (but reasonably quiet) room, and having no acoustic treatment whatsoever.

The original mic was housed in a plastic case, with zero shielding (so it picked up lots of electrical noise), and used a 1.5V cell as its power source. After modification, the case was fully shielded, and it used a modified power feed directly from the mixer - not 48V phantom power, just the 15V supply from the mixer itself. Output level and signal handling ability were increased dramatically, and the results were very impressive indeed - all for $50 or so, and a bit of modification.

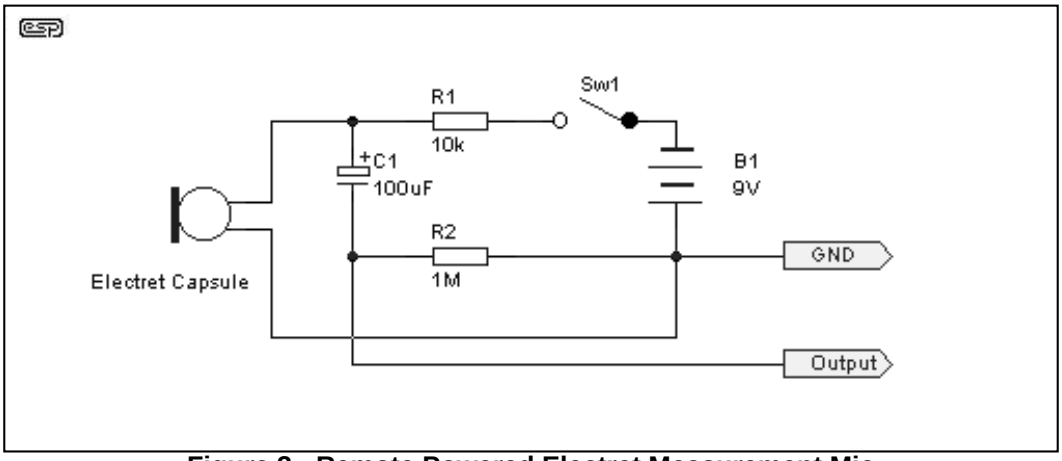


Figure 3 - Remote Powered Electret Measurement Mic

Figure 3 shows a simple remote powered microphone schematic - this can be used directly as a measurement mic with a 9V battery, and will give very good results as long as lead lengths are kept short. Generally, a circuit such as that shown should only be used with a maximum of a metre or so of low capacitance cable. If this is not enough (and it usually won't be), it is necessary to use an amplifier to reduce the output impedance. It is usually worthwhile to include some extra gain as well as shown in the schematics below.

The standard "off the shelf" electret mic may not be calibrated, but typical inserts will be acceptably flat from 20Hz to 10kHz, often with a 3dB rise at 18kHz before rolling off again. A typical response graph is shown above in Figure 1b. While it is obviously impossible to guarantee that the one you get will be the same, it is unlikely that it will be wildly different.

I have included a photo of my prototype probe, and based on initial tests, seems to be remarkably close to a Behringer measurement mic that I have performance wise, but a great deal cheaper. The output is very high with either of the amplifiers shown below - I have measured about 50mV at 70dB SPL, and it has a clean undistorted output at 100dB SPL of almost 1.6V RMS. Based on this, it is obviously not suited to measuring extreme SPLs, but as a measurement mic it is perfect.

Figure 4a - Behringer Measurement Microphone

Figure 4b - My Prototype Measurement Mic Probe

The long tube ensures that there is minimal diffraction interference from the casing, and the final unit has a turned aluminium casing, and is phantom powered. Since the PCB needs to be very small, it was not possible to finish the unit until I had PCBs made - these are approximately 12.5 x 50mm (or 0.5" x 2"), and are now available. The prototype amp was built first, and works extremely well (see Project Proper, below), but alas is too big to fit into the casing.

Figure 4c - Photo of Completed Preamp

Figure 4c above shows what the preamp looks like. I attached mine directly to the modified XLR connector, using PCB pins carefully bent to fit the receptacles of the XLR. The connections on the board are designed to align with the XLR connections for exactly this purpose. Despite the use of standard (as opposed to miniature) 100uF/16V caps throughout, the completed assembly does fit into the casing perfectly - there is not a lot of room for error, but it does fit. This was my intention from the outset, and the idea is to be able to use a standard 19mm (3/4") inside diameter tube to match the connector diameter.

Note: I have produced a small number of housings, and the "exploded" photos are shown below. These are available now, and include the XLR connector (machined to fit the housing), a PCB and PCB pins, copper tape, and two ferrite beads - all basic hardware. You will need to supply the microphone and the components for the PCB, as well as some insulation and a short length of heatshrink tubing (and/ or a suitable adhesive) to secure the mic element to the end of the tube.

The main housing is machined from a solid aluminium bar, and is 25mm (1") in diameter. Photos of the component parts are shown below. Pricing is available on the Purchase PCBs page, and covers the basic hardware set as shown - the remaining components should cost you less than AU$10, to make a complete phantom powered microphone.

Mic Casing Detail

The Other Parts

In case you were wondering, the steel rule is not part of the deal :-) It is shown so that you can see the relative size. The copper strip is used to wrap around the XLR, so that the case is properly earthed via pin 1. The complete assembly details are available in the secure section, and has step by step photos. Such photos are normally needed, but with a mechanical assembly it is essential. The screws are to secure the mic tube and the XLR connector via the recessed holes visible in the upper photo. The final assembly has a nice solid feel, and is extremely easy to put together. Not shown (but included) are the ferrite beads for the mic's outputs.

Pressure Zone Microphones® (PZM®)

The original Pressure Zone Mic was developed quite some time ago, and I have one of the very first ones that were available (manufactured by Wahrenbrock). Crown Audio has been making these since 1980, and Radio Shack (known as Tandy in Australia) also makes one (now called a boundary microphone) that is easily modified to be of near studio quality. Again, all it needs is a decent power supply, and a buffer or amplifier to ensure that the output impedance is kept low (and balanced) to match up with professional mixing desks. The original Radio Shack unit was a true PZM microphone, but it must be noted that the new ones are not - similar, but not the same. See Figure 2 for the schematic of the standard unit.

To make the modifications, the case does not even need to be undone. If you want to see what's inside, the latest ones have 4 screws under the rubber pad on the bottom, and this must be removed. It is attached with double sided tape (such as carpet tape or similar, which can be used to re-attach it when you are done playing). There is a small PCB inside amongst some medium density foam. The top piece can be removed, and the mic terminals are then accessible. The 2.2H iron cored choke (of rather dubious quality) is simply left disconnected, by not using the white wire in the shielded pair to the switch unit and battery holder (these will be discarded). The inner wiring from the mic unit is connected to the new preamp board using a suitable connector on the existing cable from the mic. A fixed lead is not recommended, but you don't have much choice at the mic end unless you really do want to dismantle it. The shield is the mic negative terminal (and also the case), and the red lead carries the signal.

Figure 5 - Radio Shack (Tandy) Boundary Microphone

Using only the shield and red lead (microphone connection - see Figure 2, above for internal schematic), this mic unit can be connected to the preamp shown below. It can then be phantom powered for theatre or other performance recording or sound reinforcement, and gives a very good account of itself indeed.

The Project Proper

Figure 6 shows the project preamp - a balanced mic line driver. This is suitable for use with phantom or battery power, and is easily adapted for either (as described below). This preamp has a PCB which is available, and is suitable for use with any of the microphones shown in this article. Suitable for measurement or recording, it has high output, low noise, and may be powered from a 9V battery or phantom power from 30V to 48V. The load resistance/ impedance should be at least 1k, but this is actually considerably lower than most preamps, so will not cause a problem.

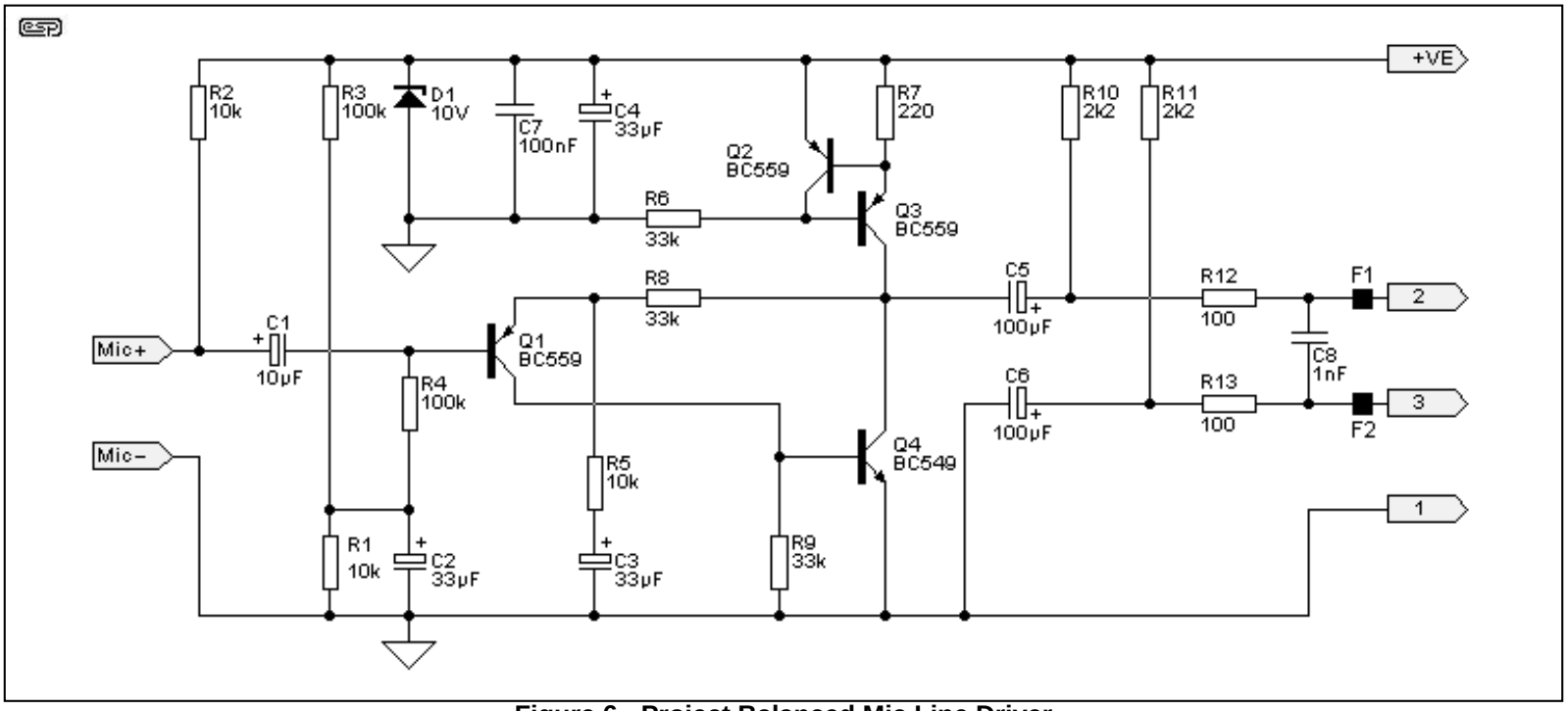


Figure 6 - Project Balanced Mic Line Driver

The preamp is very similar to the DoZ preamp - the topology is identical, but it has been modified to use a lower

supply voltage. The amplifier is a single ended Class-A, current feedback circuit, which has extremely good linearity, wide bandwidth and is unconditionally stable. Ferrite beads (F1 and F2) are recommended at the outputs. The output pins shown are the normal connections to an XLR audio connector, with Pin 1 as ground, Pin 2 is "hot", and Pin 3 is audio return ("cold"). The output actually is balanced, but is asymmetrical - this is very common, and the same basic idea is used by many premium studio microphones.

All resistors should be metal film, and electrolytic caps need to be rated at 16V. The 100uF units shown for C5 and C6 may be reduced to 33uF for recording use, but to maintain response down to 10Hz for measurement, use 100uF as shown.

To use the preamp on battery power, simply leave out R10 and R11, and connect the battery to the +VE terminal. Note that C5 and C6 *must* be reversed if you plan on using battery power, and D1 may also be omitted. Current drain is quite low (about 5mA maximum), so a 9V battery should last quite well. As noted above, PCBs are available for this version - they are tiny, and will fit easily into even small microphone housings.

This preamp can be used for a modified "boundary" mic, to make a high performance measurement mic, or to convert a cheap directional mic into something of near professional quality. It won't be the equal of a Neumann or other expensive studio microphone, but it won't set you back over $1,000 either :-)

Opamp Version

Getting enough current from the phantom supply of a mixer is not a trivial task. If the two 6.8k feed resistors are shorted to ground, the maximum available current is only 14mA, but with no voltage at all. For a workable current of (say) 10mA, the maximum available supply voltage is 14V DC. Getting the current as low as possible is a fine goal, but all opamps need some current to operate, and the powering circuit shown in Figure 7 is a relatively simple way to achieve the desired results.

This is the receiving end of the phantom supply, and it powers the microphone and opamp line driver. This circuit can be expected to handle sound levels up to about 110dB, and possibly more - this is more than sufficient for any microphone that is not used in close proximity to loud vocals or instruments.

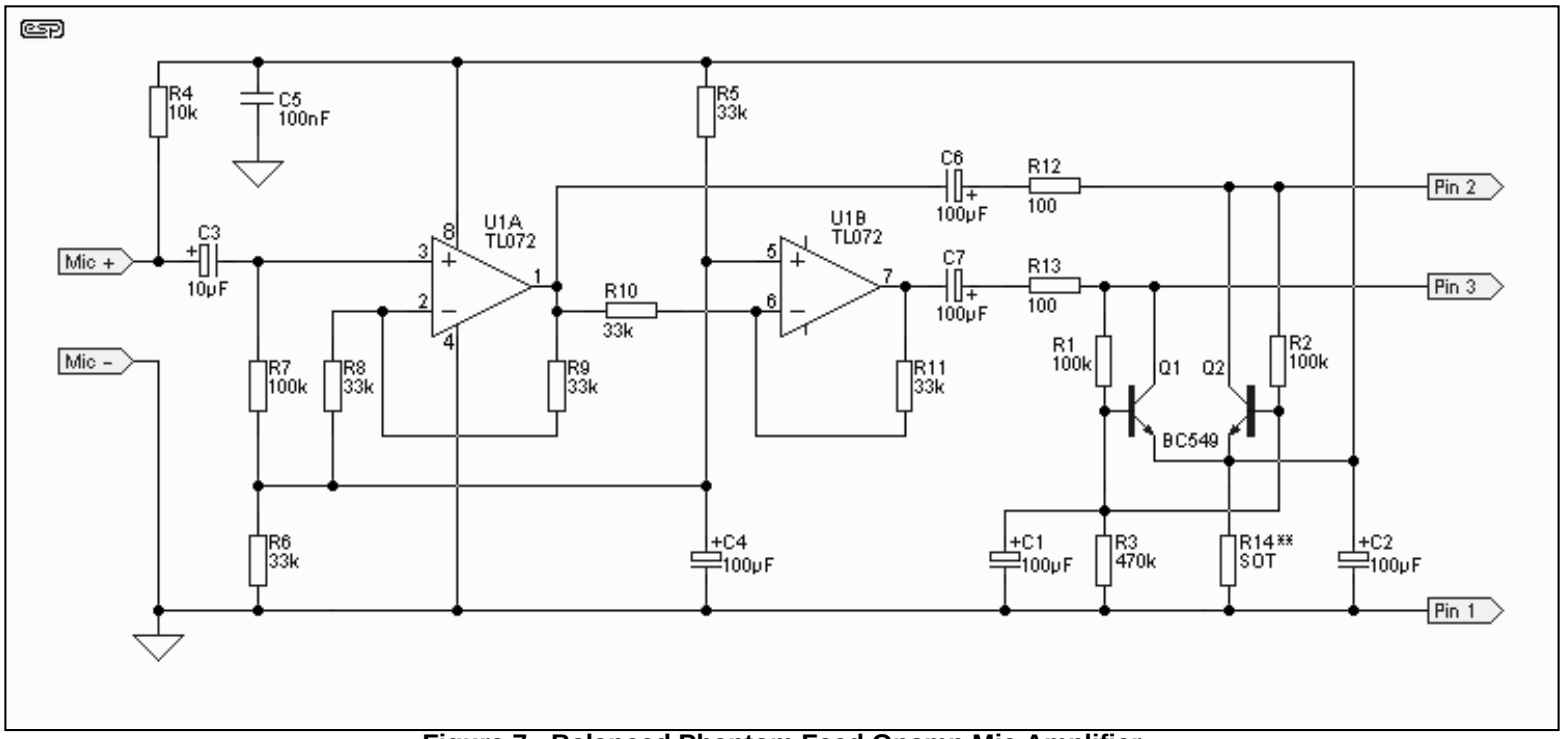


Figure 7 - Balanced Phantom Feed Opamp Mic Amplifier

A similar (but slightly more complex) method for deriving the DC from the signal lines is used by Crown in their PZM microphones, and similar circuits are also used in other (similar) mics. How does it work? It is actually quite simple. Q1 and Q2 are operated as current sinks, and the load is connected to the emitters. Because a current sink (or source) has an extremely high impedance on the collector, there is minimal loading of the signal lines. The DC appearing at the bases is filtered by C1, so the collectors only "see" the DC - the AC signal is left untouched except at extremely low frequencies (less than 1 Hz for the circuit shown above). R14 (marked **) is described as "S.O.T", or select on test. This resistor needs to be chosen so that the DC is about 10V with a normal phantom supply of 48V. A zener may be used instead, which will give a little more voltage range - depending on the opamp used.

R3 is used to ensure that there is some voltage across the transistors, and the differential must be greater than the expected peak output. As shown, the differential is about 3V, so a 6V peak to peak signal can be accommodated - this is equivalent to an output level from the microphone of almost +10dBm! This circuit has been built and tested, and works extremely well. Output level is about the same as the discrete version shown in Figure 6, but it is more complex, and a board cannot be made small enough to fit inside a slim microphone housing. No PCB will be offered for this version unless there is considerable demand.

The most common variant of my simplified DC "extraction" circuit uses a transistor as a capacitance multiplier instead of just a capacitor. This has not been found to be necessary in practice for the schematic as shown, and I have found that it works exceptionally well without the extra complications.

My prototype preamp uses a OPA2134 (a relatively high current opamp), and gets a working voltage of a little over 10V with 48V, 6.8k phantom feed. The output level is extremely high - 2V RMS can be achieved quite easily by speaking directly into the mic ... LOUDLY! In all, this is an extremely capable preamp, with excellent specifications and performance. If demand warrants it, a surface mount version of this preamp may be made available at a later stage, as this is the only way to get the size down so it could be used in a slim microphone housing.

Sound Pressure Level

It must be realised that all electret mics have one limitation we cannot readily change, and that is maximum SPL. Because they have an integral amplifier, there will always be a level where they will distort. A capsule having a 10k feed resistor and supplied from a 15V supply will output well over 1V RMS quite easily, simply by having it close enough to your mouth as you speak loudly. Even professional microphones (including dynamic types) are quite capable of 0dBm in close proximity to a floor tom or a loud singer. As a result, close vocal work, drums and brass instruments (trumpet, sax, etc) are capable of extremely high SPL, and are not really suitable candidates for electret mics. It is possible to get good performance at up to 110dB SPL quite easily, however - possibly more.

The sensitivity can be reduced, simply by reducing the value of the feed resistor. Again, there is a limit, as the internal FET amplifier can be driven into distortion regardless of what you do on the outside of the capsule. It is feasible to modify the capsule itself - but this is only possible with some models unless you are willing to make a few sacrifices (you can guarantee that you will ruin a couple in the process).

This I shall leave to the individual constructor. One possibility if using a Panasonic mic insert, is that the external track on the PCB can be cut, and this gives access to the source of the internal FET. As shown in Figure 6, this little track can be cut, and the mic is then rewired with the FET as a source follower. I do not know what the maximum SPL that a mic modified in this way will take, but it is considerably higher than an unmodified electret capsule. Expect the microphone polarity to be reversed if you do this. The standard insert produces a positive voltage for an air compression because the electret is wired to do so.

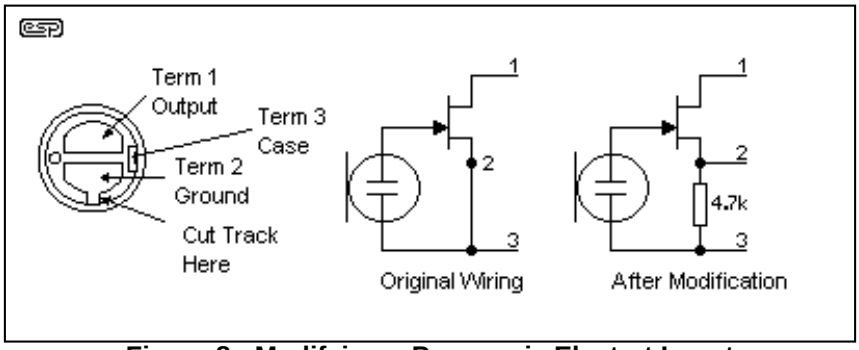


Figure 8 - Modifying a Panasonic Electret Insert

This modification was originally suggested by Siegfried Linkwitz for his cosine burst generator (see Project 58). Note that this is only known to work with the Panasonic capsule, but others may be able to be adapted in a similar manner.

The drain connection is connected to the power source with no series resistor, and the output now comes from the source (Terminal 2). There is no amplification, so expect the output level to be quite a bit lower than normal. Output impedance with a 4.7k source resistor is somewhat less than 4.7k, but it still requires an opamp (or transistor) buffer to prevent current saturation in the FET if connected to any typical mic preamp.

NOTE:
Pressure Zone Microphone® and PZM® are registered trademarks of Crown International Inc. Pressure Recording Process™ is a trademark of E.M. Long Associates.

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright © Rod Elliott 05 Jul 2002

The Audio Pages

Elliott Sound Products; Project 96

48V Phantom Feed Supply for Microphones

Rod Elliott (ESP)

Page Updated 13 June 2003

PCBs are available for this project - click the image for details

Introduction

For reasons that I find somewhat puzzling, there are very few decent 48V phantom power supply schematics available on the Net. Those that I have seen are either very crude, or require the use of a special transformer that (naturally) is all but unobtainable (or both). In contrast, a 15-0-15V transformer can be obtained almost anywhere, but alas, does not have enough voltage after rectification and filtering.

Fortunately, this is not a problem, as a voltage doubler supply will give more than enough voltage, and is easy to build. The design featured here uses just that, and allows the use of a readily available transformer and other low-cost parts, to give a supply with extremely good performance. As shown, there is no short circuit protection, but a phantom supply is unlikely to be shorted anyway, so this is not a limitation.

Of the designs that *are* available on the Net, there *is* the supply shown in the P30 mixer (see Figure 3), which is similar to the design shown here, but is not as refined.

Another version of phantom supply (reasonably common) uses an oscillator and voltage multiplier to provide the 48V (or thereabouts) supply, but these are not suitable (IMO) due to very poor regulation and an inherent inability to provide enough current. The maximum current that can be drawn from a standard phantom circuit (using 6.81k resistors) is 14mA into a short circuit. Since all phantom powered systems need some operating voltage (a typical value being around 10V), the maximum practical current drawn by each powered circuit is around 11mA. Naturally, some will draw less than this. The voltage multiplier supply will be struggling tosupply even this meagre current, and will have high battery drain as well.

There are also quite a few suggestions that you can use as little as 18V for phantom powering (using a pair of 9V batteries), but unless the microphone or DI box (for example) is specified to

operate at such a low voltage, then I would not recommend it - headroom will be reduced dramatically, and distortion will be a problem at even relatively low levels. Some equipment will not work at all. The lowest recommended phantom supply voltage is 30V, and I consider even that to be too low for most things.

Output current of this design is rated at 100mA at 48V, although you will be able to get more - 200mA is not unreasonable, and even then, output ripple can be expected to be well below 1mV. Simulation gives a figure of 10uV peak to peak at 200mA, but this is likely to be rather optimistic.

It is probable that few people will ever need the maximum suggested output current, since 200mA is capable of supplying up to 20 phantom powered microphones at once.

Description

The circuit is shown in Figure 1, and as described avove, uses a voltage doubler rectifier. From there, a pair of resistors provide additional smoothing to the secondary filter caps. R3 is used to balance the voltage across C3 and C4, and must not be omitted.

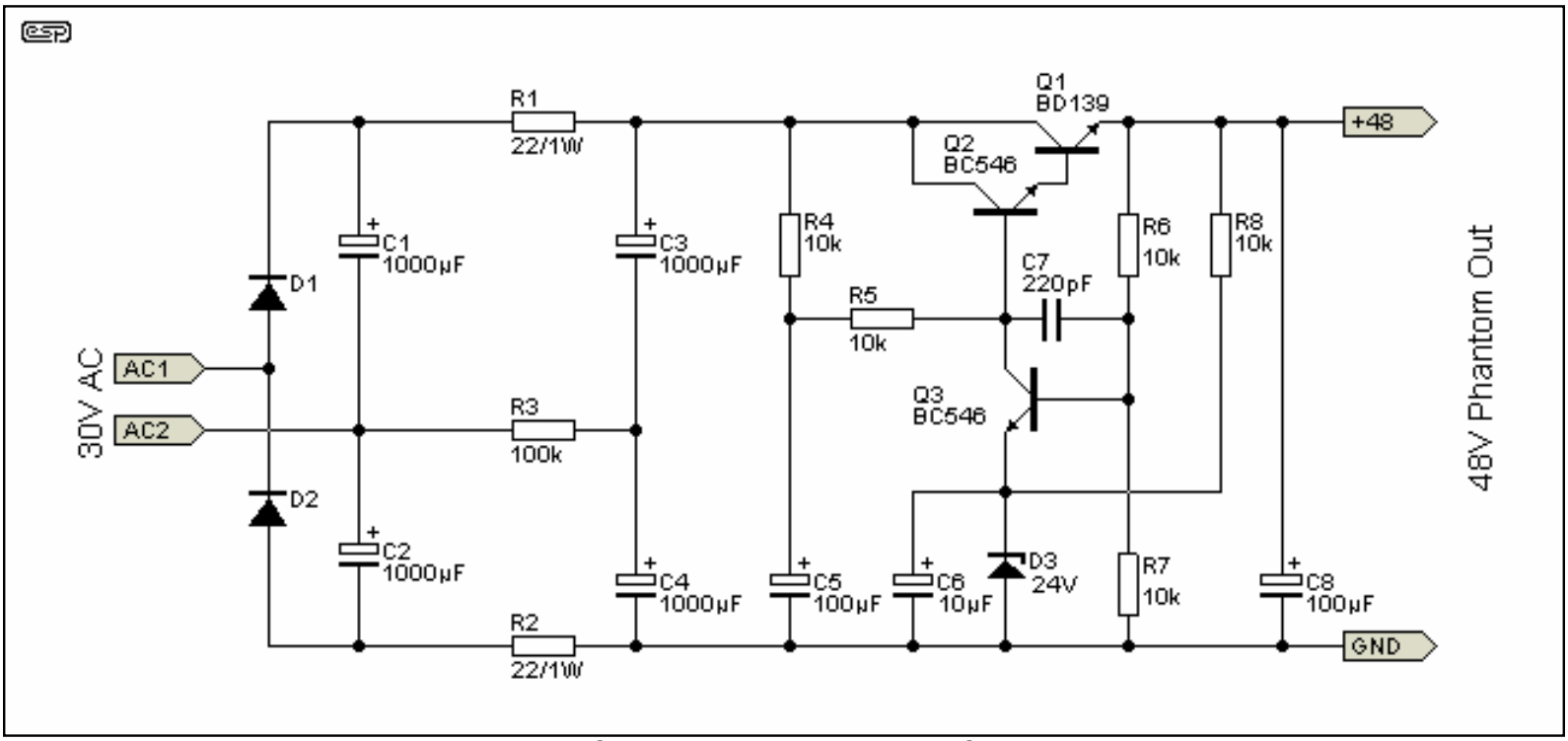


Figure 1 - 48V Power Supply

The regulator was a very common topology prior to the introduction of 3-terminal regulator ICs, and is used here so that high voltage regulators are not needed. These are much harder to get than the standard versions, and still require additional circuitry because 48V versions are not made. Although the circuit looks complex, it is very easy to build (especially if the PCB is used).

The zener diode is the reference voltage, and 1/2 the output voltage is compared to the zener voltage by Q3 (the error amplifier). If output voltage increases, Q3 is turned on harder, removing

base drive from Q1 (and hence Q2), reducing the output voltage to the preset value.

As can be seen, there are no adjustments, and this means that the 48V may be a little higher (or lower) than rated. This is not a problem however, and all phantom feed microphones will handle the variation without any problems at all.

Load regulation is far better than you might expect, with typically 100mV variation between full load (100mA) and no load. At 200mA load, the voltage falls by less than 150mV compared to the no-load voltage. Line (input) regulation is also quite good, with less than 200mV output change with +20% and -20% input voltage, with a load of 100mA.

The next problem is how to actually send the phantom power to the microphone and not the mixer input circuits. The latter will not be impressed with 48V DC applied, and will most likely voice their displeasure by failing instantly. The standard value of 6.81k (0.1% tolerance) for phantom feed circuits can be reduced to 6.8k (a standard E12 series resistor value), and I suggest that using a multimeter to match the resistors to at least 0.1% is the easiest and cheapest alternative. Each pair of resistors should be matched to within 10 ohms (or less if possible) of each other for best results. This is better than 0.1%, and ensures that common mode performance is not compromised.

In case you were wondering about my claim that 10 ohms is better than 0.1%, a worst case pair of 0.1% 6.81k resistors could have a difference of 13.62 ohms - one resistor at the maximum positive tolerance, and the other at maximum negative tolerance. Fairly obviously, the closer the match the better, and the multimeter used does not have to be absolutely accurate, since you are measuring for a difference rather than an absolute resistance value. If your multimeter refuses to measure to the number of digits needed, see the appendix for an alternative method you can use.

Figure 2 shows the basic phantom powering scheme. Only one channel is shown - subsequent channels are identical, up to a typical maximum of 10 (20 at a pinch) for a single supply module.

Although shown using bipolar electrolytic caps, some constructors will no doubt want to use something 'better', but polyester or similar caps at those values will be very large! Assuming a mic circuit input impedance of 1.2k (fairly typical), the two 22uF caps as shown will give a -3dB frequency of 12Hz - this is needed to get flat response to 20Hz. Naturally, if the lowest frequency you need is higher, then lower capacitance is acceptable. Likewise, if the mic preamp input impedance is higher than 1.2k, less capacitance can also be used.

It is worth noting that many mixers use polarised electrolytics at the inputs of phantom power circuits. While this is quite ok while phantom power is applied, the caps will be unbiased when phantom power is not being used. This is not a good idea (IMO), and can lead to audible distortion or colouration. For a "cost no object" design, use 10uF/50V (or higher) polyester caps, enclosed in their own shielding can. These can be wired into the PCB without too much difficulty. Keep wires tightly twisted for maximum noise rejection.

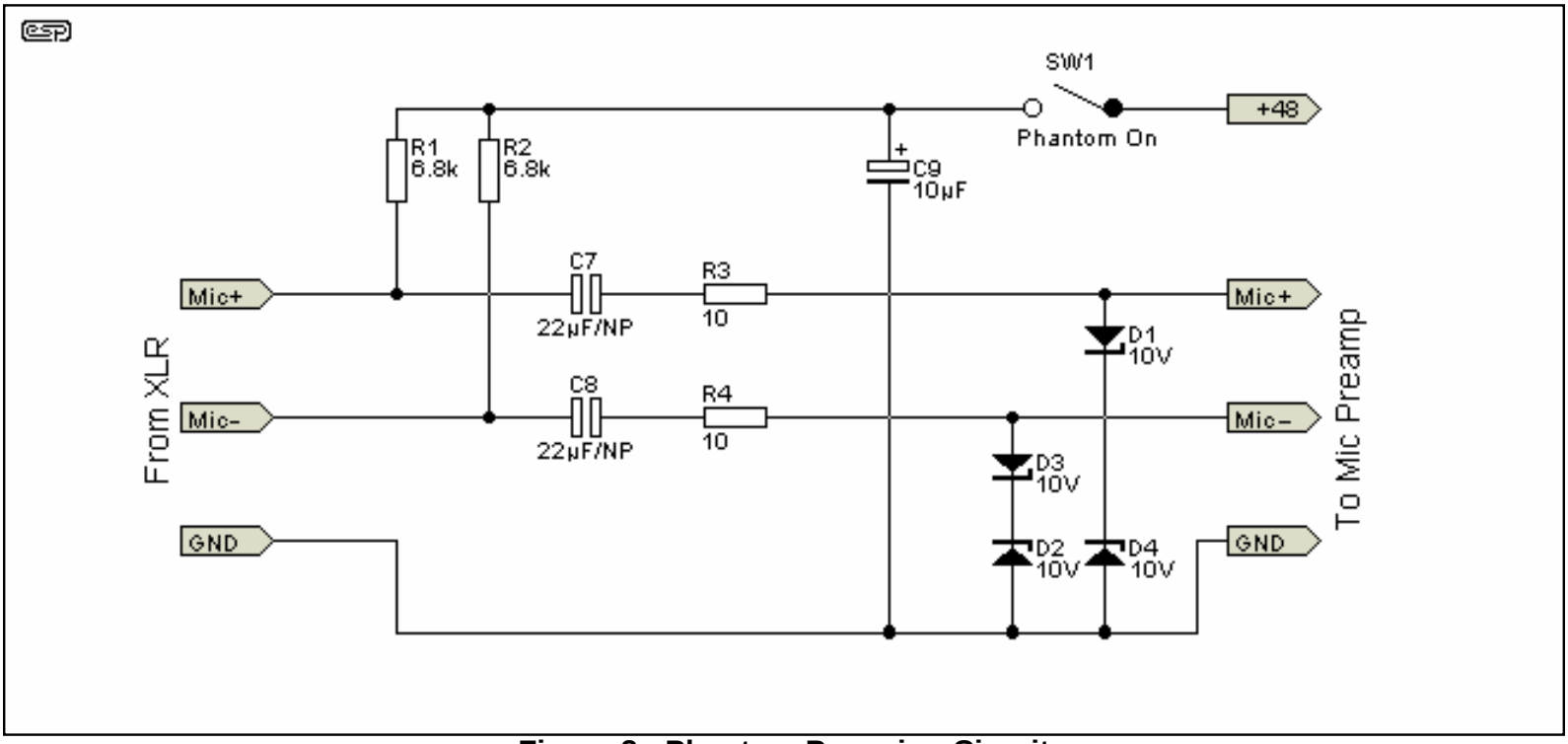


Figure 2 - Phantom Powering Circuit

Zener diodes must be used as shown to limit the maximum voltage applied to the mic input circuits. The worst possible scenario is when a mic lead is connected to a mic while phantom power is on. The cable is effectively a capacitor, and the sudden discharge of the cable and coupling caps can create a high current through the zeners, which must be capable of withstanding the surge without failure. Fortunately, 1W zeners are rugged enough to take it, and this is almost an industry standard circuit. Maximum surge current for 1W 10V zeners is typically around 450mA, and this is unlikely to be exceeded in practice. 10V zeners are specified because it is virtually impossible for any microphone to exceed that level, and the mic preamp will clip well before you achieve the 7V RMS input voltage limit imposed by the zeners. In addition, 10V zeners have a higher current capacity than higher voltages for an additional safety margin. The 10 ohm series resistors will have little or no influence on input level or noise, and help to limit the peak zener current.

Some mixers boast a "silent" phantom switch to alleviate the typical loud BANG through the mixer when phantom power is turned on or off. The phantom distribution PCB (2 channels) has this feature, but it is not shown here.

Construction

Somewhat naturally, I suggest the PCBs be used, as this makes construction very easy. The PCBs both measure only 64mm x 38mm (2.5" x 1.5"), so will be easy to retrofit into all but the most compact mixer. Where there is just no space at all, an external box can house the regulator and distribution board(s).

If you don't wish to buy the PCBs, you may use Veroboard, but unless great care is taken with the

earthing arrangements, noise will almost certainly be worse than quoted. All electrolytic caps should be rated at 50V or higher (**NOTE:** C5 in Figure 1 *must* be rated at 63V!). A standard 50V ceramic is recommended for C7, which is used to ensure that the regulator does not oscillate. Not shown (or needed) are film caps in parallel with the electrolytics - if you wanted to, these may be added, but with the filtering shown high frequency noise should be non-existent.

The power transformer does not need to be anything too fancy, but I suggest a separate box for it to minimise hum and noise. Typically, a 20VA 15-0-15V (or a multitap transformer with a 30V connection) will be more than enough. These are available readily in Australia, but I can't speak for the rest of the world. If the worst comes to the worst, you can use a pair of single winding 15V transformers, with the windings in series to give 30V. I recommend a conventional 'EI' transformer if possible, as these have less capacitance between primary and secondary, and will allow less HF mains noise through.

It is extermely important that the transformer is not used to power other supplies or equipment. The centre-tap *must* not be connected to anything, and needs to be insulated to prevent contact. The supply circuit uses the full 30V AC in a floating configuration, and connection to another supply or rectifier will cause a short on the winding.

Q1 on the supply must be fitted with a heatsink, and worst case power dissipation will be around 5W. This may not seem like a great deal, but a 10°C/W heatsink (typical of a large PCB mount type) will get to 50°C above ambient temperature (i.e. too hot!) at 200mA output.

At 100mA load, this is reduced to about 3W, which is a little more manageable. Even so, there is no such thing as a heatsink that is too big, so I suggest that you use the largest one that you can. Forced air (fan) cooling will not be necessary.

The PCB is laid out such that the power transistor can be attached directly to the chassis if desired (using insulating washers and heatsink compound, of course), and this alleviates the need for a separate heatsink. For only one or two phantom powered mics, a heatsink is not essential, but a small one is cheap insurance. In this case, R1 and R2 in Figure 1 may be increased in value, and this will provide even better filtering. 100 ohms will be more than satisfactory for a two microphone system.

Appendix - Resistor Matching

The age old methods are sometimes the best. A Wheatstone Bridge used to be the standard method of accurately measuring resistance, inductance and capacitance, but new digital instruments have taken over. This is a shame (I think), because the old methods actually taught you something as you used them - not the case with any digital instrument.

A Wheatstone Bridge is easy to set up, and while it is not especially accurate in absolute terms, it

can be made extremely sensitive to *variations* between components. Using a power supply, 9V battery (or even a small AC voltage from a transformer), the circuit shown in Figure 3 will resolve a difference of 1 ohm easily with a suitable test setup. A difference of 10 ohms will give a 4.38mV output signal from a 12V input.

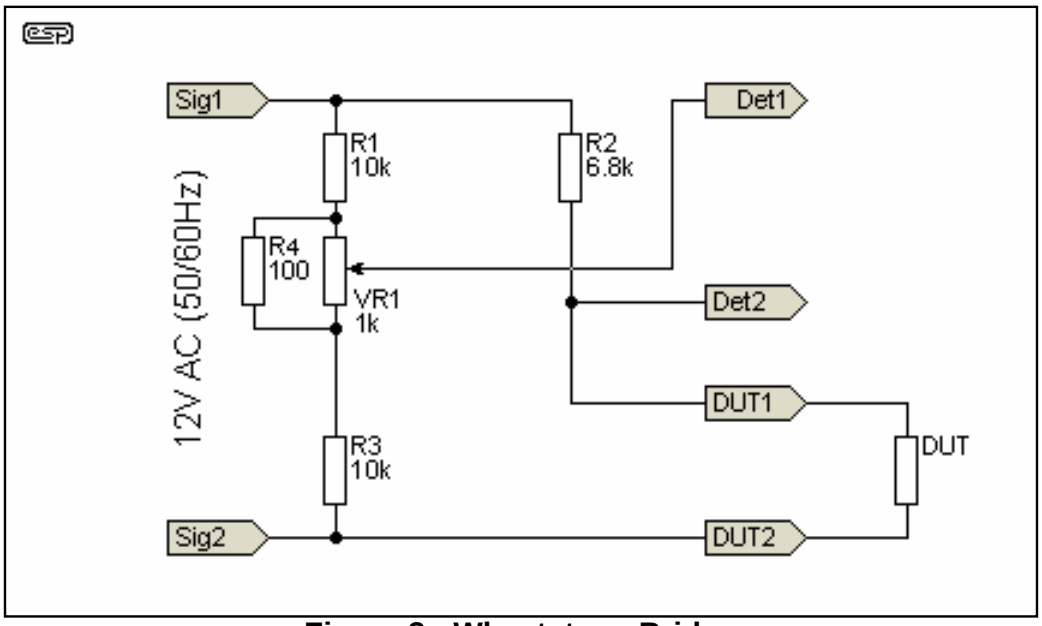


Figure 3 - Wheatstone Bridge

Because the bridge circuit does not care if you use AC or DC, an amplifier can be used to detect the null, and it can be made almost unbelievably sensitive. In use, wire up the circuit as shown, and use a meter or amplifier and headphones (or speaker) to monitor the difference signal. **Be warned** - the amp input signal will be 6V RMS (for a 12V input) with the DUT (Device Under Test) removed! Install a test resistor, and adjust VR1 for a complete null (no signal). That is now your reference resistor. Connect the remaining resistors into the circuit, and aim for an output signal of less than 4mV - do not re-adjust VR1.

The signal input (applied unsurprisingly to Sig1 and Sig2) *must* be floating if you use an amp for monitoring, since one side of the amp input will be earthed. A 50 or 60Hz 12V transformer is fine for this. Press the TEST button only when the DUT is securely attached, or the output will be very loud indeed. You can connected a pair of diodes across the Monitor terminals - in parallel and opposite polarity. This will keep the maximum level down to something more sensible.

The Wheatstone Bridge is such a useful device that you can soon expect to see a complete article describing its use for comparing resistors, capacitors and inductors - actually, you can match any passive components and even complete networks by this means. Wheatstone bridges are also used for precision temperature measurement, strain gauges (use to test mechanical movement in structures) and has many other uses.

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright © Rod Elliott 22 Dec 2002

The Audio Pages

Elliott Sound Products **Project 112**

Yorick - A Dummy Head Recording Microphone

Rod Elliott (ESP)

PCBs are available for this project. Click the PCB image for details, or select P93 or P88.

Introduction

"Alas Poor Yorick ..." Well, there's not a lot of alas in this - indeed a lack of alas, one might say. Yorick just happens to be the name I chose for my dummy head recording microphone for reasons that shall remain obscure Using a pair of P93 electret microphone amps, this dummy head can be phantom powered or you can use batteries. It will run happily on a standard 9V alkaline battery.

> Speaking of Shakespeare, something that is not at all well understood is this snippet ... While writing 'King Richard III', poor Bill's word processor had a hard drive failure, and the actual quote (lament, more like it) was "Now is the winter of our disk content." - having just lost everything, I know how he must have felt. This was utterly misconstrued over the years, until we now have the mangled version that everyone seems to think is correct. Of course this is true - would I lie? To *you* ?

The project came about when I came across a polystyrene foam wig holder (replete with wig), and its final use was sealed on the spot. It still took me a year to get around to it, but hey, these things take time. With the addition of a pair of electret mics and a couple of RCA sockets, the dummy head was complete, just needing a suitable mic amplifier to finish it off. P93 was an immediate candidate, although for my application I used P88.

More than anything else, this is intended to be a fun project. While those who are especially interested in making recordings of <*insert favourite application here*> will find it very useful, the project also shows how powerful our auditory processing really is. To be able to listen to a recording made in natural surroundings with almost no reverb - which we normally do not hear in most rooms, even though it's there all the time - is remarkable. To hear the difference, simply mono the signal, and suddenly the reverb and general background noise becomes most intrusive.

Figures 1 & 2 - The Completed Dummy Head Recording System, Front and Rear Views

The completed dummy head unit is shown above. The details are given below. He's not particularly handsome by any stretch of the imagination, but that's not really a requirement for the application. The control box is shown below.

Figure 3 - Prototype Control Box

The prototype shown is based on the Project 88 hi-fi preamp (details shown below). It can run from a pair of 9V batteries, or an external DC source may be used. Any external supply needs to be (electrically) very quiet. Any hum or noise from the supply will have some impact on overall noise performance. I used NE5532 opamps in mine, but be aware that these (as well as many other high quality audio opamps) draw considerable current - around 20mA in total for two dual opamps. This means that battery life is fairly limited. Expect no more than about 10 hours continuous operation from alkaline batteries, less from zinc-carbon or re-chargeable types. As you can see, I added a pair of XLR connectors for flexibility, but this is completely optional.

Description

There's not really very much to it. Simply take a suitable head such as the one shown, and make a couple of ear holes using a long, thin screwdriver as a boring tool. These should be located in approximately the position of real ear holes, but if you plan to add a wig for added realism (oh, really?) then make sure the wig doesn't cover the holes. The outer ends of the holes need to be enlarged so the mic inserts are a snug fit.

Figure 4 - Side View (Sans Wig) Showing Microphone Insert

Most of the expanded polystyrene foam heads are reasonably solid, and have a hollowed out section up the inside. This allows us to run the cables through easily. A rotary tool is useful for making neat cut-outs for the microphone connectors, and for enlarging the ear holes to accept the inserts. Overall insulation is not needed because polystyrene is an excellent insulator, so you only have to make sure that the mic leads (inner cable and shield) can't short together.

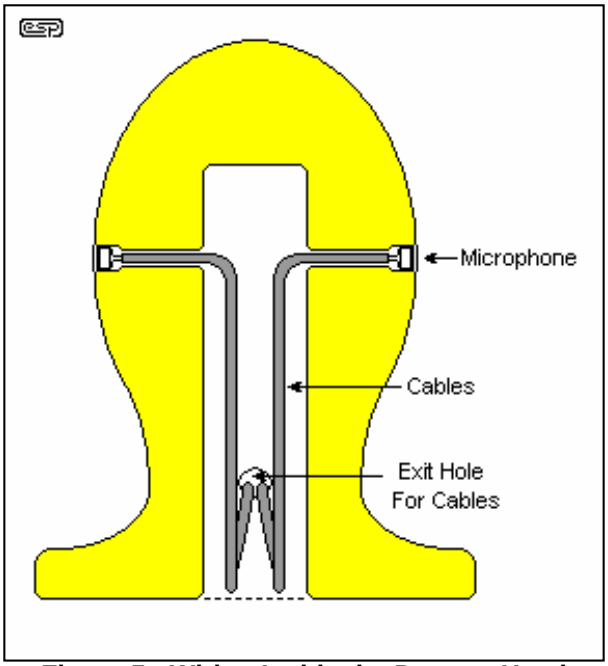


Figure 5 - Wiring Inside the Dummy Head

Getting the cables through the small holes from the mic locations, down the hollow centre and out through the opening for the connectors can be a challenge. You may find it easier to use a thin draw-wire that can be fed through the holes and hooked out with a stiff wire hook. Ultimately, I have to leave that part to you, as the best

method will depend on what you have available. Because the hollow section is a reasonable size (at least in the heads I've seen), you have plenty of room to stash any excess cable.

Mounting the mic connectors may be easy or hard, depending on the connectors you have, and what you can find in your junk box. In my case, I have a large bag of the RCA connectors you see in the photos above, and it was possible to just screw them directly into the polystyrene. No, this is not the most secure fastening, but I don't plan to hang heavy cables from them, and tests so far seem to indicate that they'll be fine for as long as I need. Some care is needed, but that's not an issue.

I do recommend that the two connectors are electrically isolated from each other. Because of the small signal level, a ground loop will be created if the leads join at more than one place, and this will create hum. If you plan on gluing anything, remember that solvent glues will dissolve polystyrene foam (it just vanishes!), so only use adhesives that are designed for such applications (I don't actually know of any, which is why I tried the screws). PVA wood glue will work and does not dissolve polystyrene, but it probably won't bond very well either. The connectors could be attached to a strip of plastic and attached with cable ties. There are many possibilities, so see what works best for you.

Control Box Construction

There are actually two equally valid possibilities for the control box. As mentioned above, the P93 mic amp is ideally suited, and for reference this is shown below.

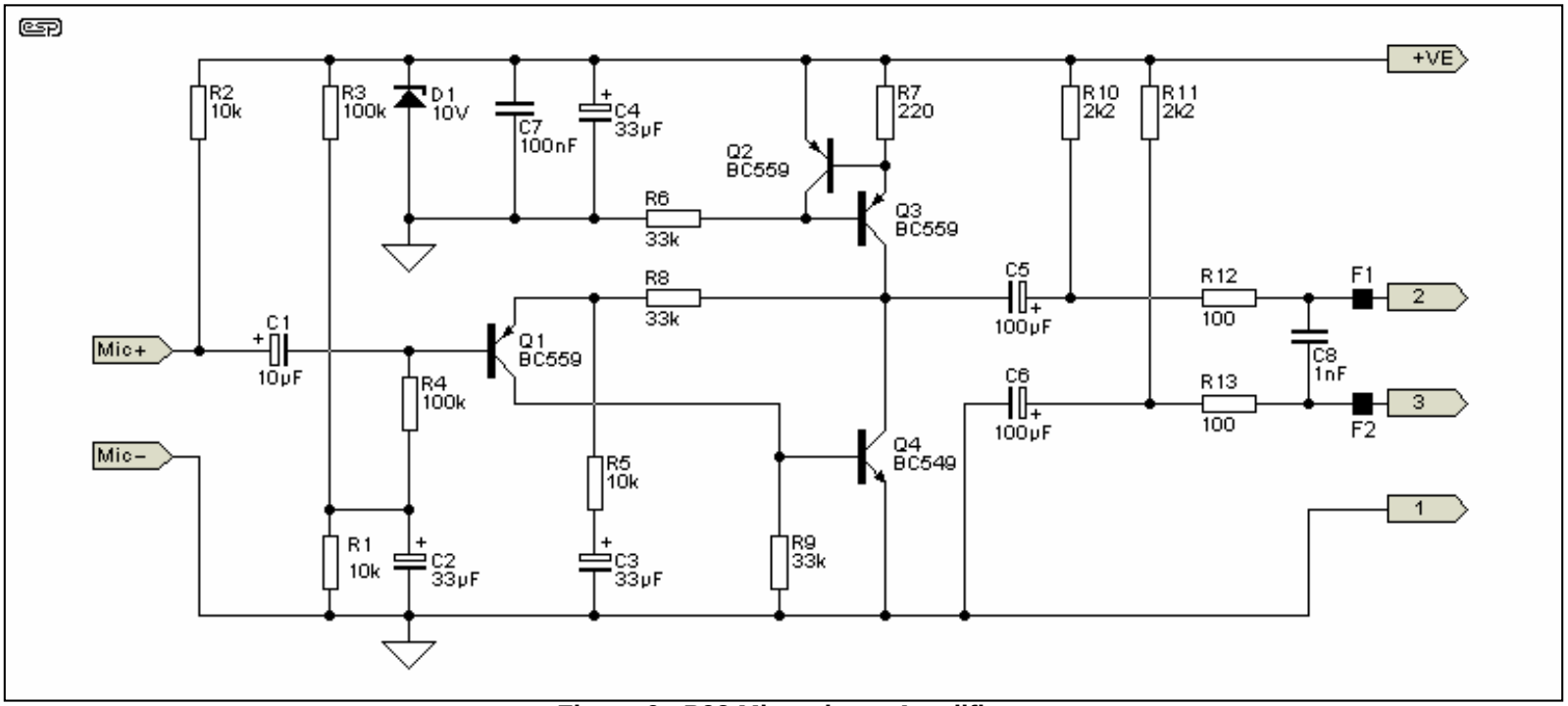


Figure 6 - P93 Microphone Amplifier

This is unchanged from the original project. The only difference is that the P93 board is not installed close to the mic capsules. Although it would be possible to do so, having XLR connectors on the dummy head would mean a substantial mounting because of the force needed to connect and disconnect. RCA connectors require very little force, making mounting easier.

Another method is to use the P88 preamp, and this is ideal if you don't need the balanced output and phantom feed capability. The gain needs to be increased, but it is a very flexible alternative.

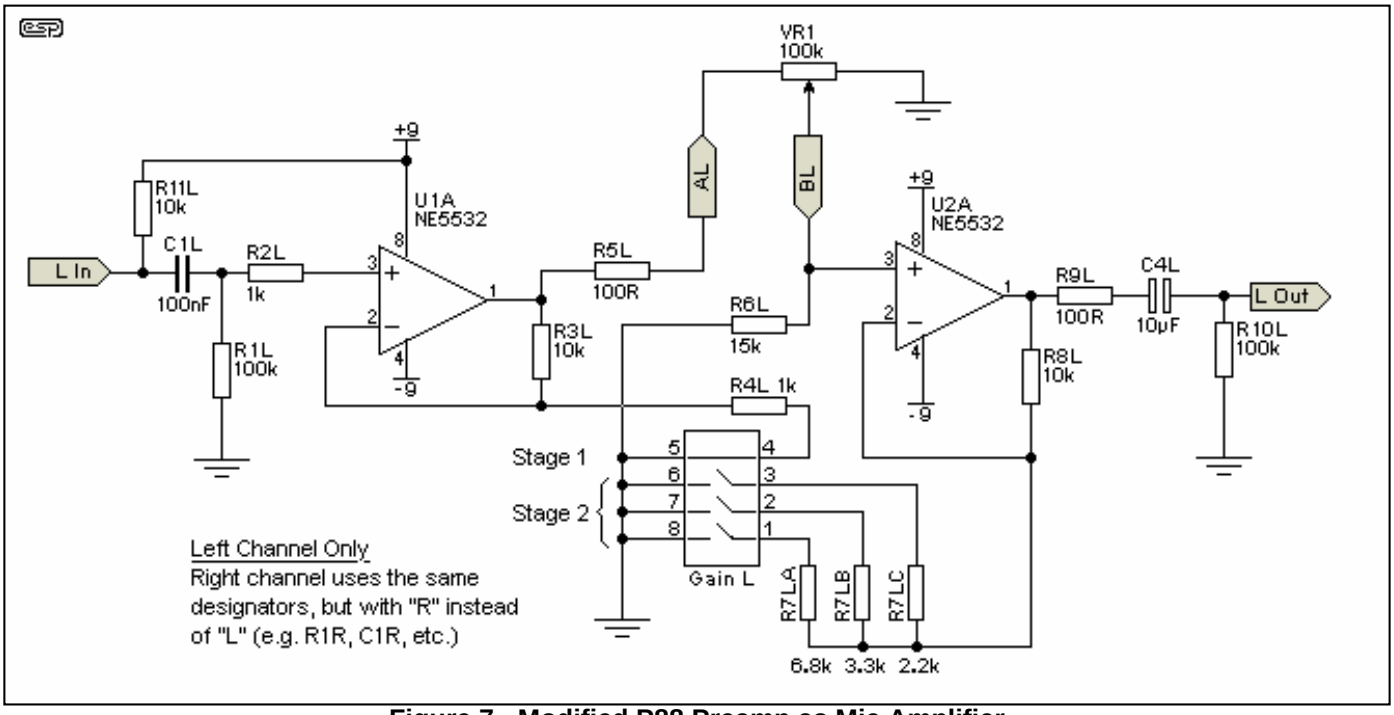


Figure 7 - Modified P88 Preamp as Mic Amplifier

The use of low noise opamps is highly recommended, but be warned that devices like the OPA2134 and NE5532 draw about 5mA per opamp (20mA per package for dual opamps), so you will be looking at close to 20mA for the amp. While 9V batteries can supply this easily enough, you will only get about 10 hours of use from a set. TL072s will give much longer battery life, but at the expense of noise.

As shown, the first stage gain is increased to 11, so a typical 20mV signal level will give you 220mV across the level pot. The gain of the second stage is also increased, and like the P88 can be made switchable to accommodate different recording systems. With the values shown, you can get the following ...

Switch	Stage 2 Gain	Total Gain	dB (Total)
1	2.47	27	29
2	4.03	44	33
3	5.54	61	36
2+3	8.57	94	39
1+2+3	10.05	110	41

The above is with the volume control at maximum. Switch #4 should be closed at all times for this application. You can experiment with the resistor values if you need more or less gain. The level control (VR1) can be a stereo (dual gang) pot, since in most cases you will not need to adjust the levels independently. Remember that the opamps are DC coupled, so expect to get some DC offset at the output pins. A 10uF bipolar cap may be used from the AL (and AR) terminals leading to the pot. This will prevent DC induced pot noise as the level is adjusted.

Note that R11 is not provided for on the PCB, but it is an easy matter to add it. Details will be included shortly in the P88 construction guide to make it a simple matter to adapt the preamp.

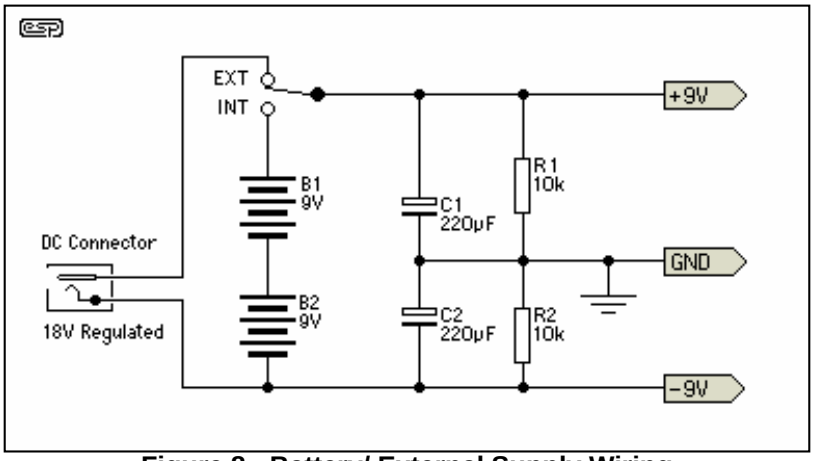


Figure 8 - Battery/ External Supply Wiring

The above is a suggested wiring scheme for the supply for the modified P88 board. While the battery centre tap could be earthed, then you would need a double pole switch to change from battery to external power. Feel free to do so if it makes you feel any better.

The supply will not be completely symmetrical because of the mic feed resistor. You can change R2 in the supply above to 5.6k or thereabouts to make the supplies equal, but the opamps don't really care that much.

Testing

Connect to a suitable power supply - remember that the supply earth (ground) must be connected! When powering up for the first time, use 100 ohm safety resisors in series with each supply to limit the current if you have made a mistake in the wiring. Refer to the individual project pages (P88 or P93, depending on which you used) for full test procedures.

Once the preamps are working, connect the complete recorder system to a tape machine or a computer sound card (via a mixer in most cases). It is quite uncanny to listen to the normal sounds around you via headphones from the dummy head.

Project Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright © Rod Elliott 29 Nov 2005

Copyright (c) 1999-2002 by Rick Chinn. All rights reserved.
Revision date: 1/06/2003

Phantom Powering for Condenser Microphones

History

Condenser (capacitor) microphones differ from dynamic and ribbon microphones because they are not self-generating. That is, they cannot generate electricity in response to an impinging sound wave. A condenser microphone modifies an external source of electricity to reflect the effects of a sound wave striking its diaphragm. Dynamic and ribbon microphones use magnetism to generate electricity in response to a sound wave: they are self-generating. Furthermore, both of these types of microphones are inherently low impedance devices. It is possible to connect a dynamic microphone element directly to a balanced, low-impedance mixer input. Many commercially made dynamic microphones do just that.

On the other hand, a condenser micorphone is an inherently high-impedance device. How high? Verrrrrrry high. On the order of a billion ohms (1 Gigaohm). This is high enough that the inherent capacitance of a foot of shielded cable would audibly reduce the output of the microphone. All condenser microphones have an impedance converter, in the form of a vacuum tube or field-effect transistor (FET), built into the microphone, and located extremely close to the microphone element. The impedance converter and the microphone element itself require an external power source. <To be strictly correct, electret condenser microphones are a bit different as the microhone element does not require a power source for operation (it is more or less permanently self polarized). Regardless, the impedance converter still requires an external source of power.>

What is it, exactly?

The obvious external power source for any modern microphone is a battery. About the only electronic advantage that a battery has is that its output is pure DC. The only other advantage (to the battery company) is that you have to keep on buying them.

Tube microphones require several different voltages for operation. This invariably means a multi-conductor cable and non-standard (not XLR) connectors. A tube microphone will always have an associated external power supply.

In the late 1960’s, Neumann (you know, the folks that brought you the U47 and U87 microphones) converted its microphones to solid-state, adopting a system of remote powering that they called, and trademarked, Phantom Powering. Because of the trademark, some manufacturers use terms like Simplex Powering, etc. Over the years, the trademark has become genericized and now refers to any device that

is powered according to DIN standard 45 596.

So, where is the Phantom of Phantom Powering? Because (like the Phantom in the old comic strip) it's there when you need it, and invisible when you don't. This technology is not new, it actually predates rocket science. Like so many other things in audio, it was brought to you by the telephone company, who used it to get an extra circuit from a pair of wires. In effect, so does your phantom powered microphone.

What is important is: phantom powering is a compatible system. Your dynamic/ribbon microphones as well as your condenser microphones work side-by-side, from the same microphone inputs, without further thought on your part.

Technically speaking, phantom powering refers to a system where the audio signal is applied to the balanced line in differential-mode <across pins 2 and 3>, and the DC power is applied common-mode <pins 2 and 3 plus pin 1>. The audio travels via pins 2 and 3 differentially, and the power travels between pins 2 and 3 simultaneously and pin 1. Microphones that do not require power simply ignore the DC present between pin 2, pin 3, and pin 1. If you measure with a voltmeter between pin 2 and pin 3, you will read 0 Volts DC. This is what your dynamic microphone sees. Measuring between pin 2 and pin 1, or between pin 3 and pin 1, you will read the phantom power voltage, usually 48V, without a microphone connected. The dynamic microphone, as well as your balanced mixer input, ignores this voltage.

Lately, the term phantom power has been perverted to refer to any remote powering system. In the strict sense of the DIN standard, this is not true. Furthermore, microphones or transducers that claim to use this system are not compatible with the DIN standard and will almost certainly be damanged if connected into such a system. Fortunately, these systems use tip-ring-sleeve phone plugs or miniature XLR connectors and they are usually associated with instrument pickup applications.

Phantom powering is defined in DIN standard 45 596 or IEC standard 268-15A..

What works?

To be compatible in a phantom powered system, a device (microphone, preamp with a microphone-style output, or direct box) must have a balanced and floating, low-impedance output. This includes all microphones commonly used for sound reinforcement and recording such as the Shure SM-58, SM57, Electro-Voice RE-15, RE-16, RE-20, ND series, Beyer M88, M160, M500, AKG D224, D12, D112, and MANY others.

If you are fortunate enough to own any tube condenser microphones, such as the AKG C12, Neumann U47 or U67, these microphones may be connected in a phantom powered system and will operate without regard to the presence or absence of phantom power. They will always require their external power supply (which must be plugged in and turned on).

Older condenser microphones (Sony C22, EV 1777, PL76, etc.) that were either made before the adoption of phantom power, had the phantom power feature omitted as a cost savings measure, or omitted it for product differentiation reasons (stupid, really stupid). These microphones will operate, but they will always require their own batteries. Consider them a battery-powered dynamic microphone. There is one tube condenser microphone that can operate using phantom powering. It is the Microtech Gefell UM900. For further information, consult their website

What doesn't work?

The list is short:

1. Microphones with unbalanced outputs.
2. Microphones with grounded center-tapped outputs. Many old ribbon microphones were supplied connected this way. Have a technician lift the ground from the center tap. The microphone is now phantom-power compatible. Connecting a microphone with a grounded center-tap, when phantom power is present, could damage the microphone. Connecting the microphone first, and applying phantom power second is permissible and will not harm the microphone. We recommend the latter (connect first, then apply power) as a safety measure if you don't know for sure that the microphone has a grounded center-tap.
3. High-impedance microphones.
4. Microphones that exhibit leakage between either pin 2 or pin 3 and pin 1. These microphones will sputter and crackle when phantom power is applied and will work fine when you turn off the phantom power. Get the microphone repaired.
5. Musical Instrument Stuff with Direct Outputs. Sometimes they work, sometimes they don't. It's a crapshoot at best.

Many pieces of electronic equipment aimed at the musician marketplace have direct outputs intended to make connection into a professional sound system easier. In theory at least, you eliminate needing a direct box. The problem is that many of these manufacturers apparently haven't heard of phantom power.

The problem is that their balanced output is transformerless, and (worse) can't tolerate the 48V power present. Usually, the output coupling capacitors are 25V parts, and most often connected so that their + terminal goes to the opamp that drives the output connector. When connected into a phantom powered system, the 48V reverse biases the capacitor and causes trouble. This trouble can be temporary and sometimes it can be permanent.

If you're a musician and this scares you, it should. What can you do? Call the manufacturer and ask. Have a service tech who understands the problem check the schematic. A really bright service tech can study the circuitry without a schematic by looking at the unit and figuring out what the manufacturer did circuit-wise.

I have seen at least one brand of amplifier (G-K) with outputs that are supposedly phantom power proof that aren't. In this case, I spoke to their chief engineer, and we both agreed that what he was doing should work, but in practice it doesn't. If you are connecting to a G-K/Gallien-Kruger amplifier, you should use a direct box.

Electronic keyboards are also suspect. I've had other musicians tell horror stories of blown instruments caused by assuming that the balanced outputs on the instrument were phantom power proof. When in doubt (and that's my usual state), use a direct box.

A Simple Modification

This modification makes a balanced transformerless output phantom power proof. I came up with this idea when I worked at Mackie Designs. It works. It has been incorporated into all of their mixers that have XLR connectors for their outputs. I'll describe the mod in words. This should be crystal clear to a competent service tech.

I'm assuming that the circuitry of the balanced output is driven by a pair of opamps, with their outputs arranged to be out-of-phase with each other. It doesn't matter if the output is cross-coupled or not.

Find the output coupling capacitors. They need to be at least 25V units, polarized so that the plus side connects to the XLR connector.

There should be a resistor on the XLR side of the coupling capacitor going to ground. There should be one for pin 2 and one for pin 3. If not, add them. The value needs to be 10k or lower (equal value).

The resistors put enough of a load on the phantom power coming from the mic preamp to reduce the phantom voltage to something that the capacitor can withstand.

What about T-Powering?

T-powering is another remote powering system. It is sometimes known as A-B powering or as "Modulation Lead Powering." It uses pins 2 and 3 to carry both power and audio. It is NOT compatible with dynamic, ribbon, or phantom powered microphones. For further information, please refer to my article about T-powering.

Do/Don't Chart

Do	Don't

If you are plugging in a condenser microphone, do verify that your microphone can be phantom powered.	Worry about your other microphones as long as their outputs are balanced and floating.
Ensure that the microphone's output is low impedance, balanced and floating. This is especially important for vintage ribbon microphones like the RCA 44BX and 77DX.	Connect microphones or devices that do not conform to the DIN 45 596 standard.
Mute the sound system when turning the phantom power on or off, or when connecting or disconnecting microphones. Lest you forget, the resulting loud, nasty POP may be your last.	Don't connect A-B or T-system microphones (another remote powering system) without suitable adaptors.
Turn the phantom power off when connecting ribbon microphones (cheap insurance). It's OK to turn it back on after making the connection. Be sure that there is some way to bleed the charge off of the power supply if you chose to do this. See the discussion on patch bays, below.	Worry about your tube condenser microphones. They're compatible; they just can't be phantom powered.
Do read the discussion about phantom powering causing magnetized transformer cores.	Put your microphone lines thru a TT or TRS patch bay. See further discussion, below.

Phantom Power and TT/TRS Patch Bays

If you're building a studio, this is a bad idea. Avoid it.

Use XLR connectors instead. Yes, they're not as high density as a TT bay. Don't hot patch microphones when the phantom powering is turned on. Turn the phantom power off first. Even this may not be sufficient.

The problem is that the tip of the plug can contact ground as the plug is being inserted. With a transformerless input, this can discharge/charge the input blocking capacitors thru the microphone's output transformer. This could magnetize the core of the transformer, zap a ribbon microphone, or (remote possibility) damage a dynamic microphone.

With a transformer coupled input, you risk magnetizing the core of the console's input transformer, ruining its distortion characteristics.

If the microphone's output is transformer coupled, the microphone's output transformer could also have its core magnetized, ruining its distortion characteristics.

This is not good.

There are several ways to disable or turn off the phantom power in a mixer.

Disconnect the input to the rectifier/filter system in the phantom power supply. Mackie mixers with a global phantom power switch use this method. There should be a bleeder resistor somewhere within the supply to discharge the filter capacitors.

Break the feed to the 6k81 phantom feed resistors, either individually or in groups. In transformerless consoles, there should be a discharge path for the input coupling capacitors.

Using a voltmeter, or the console schematics, determine which method your console uses and how long it takes for the voltage at the microphone inputs to fall to zero volts after the phantom power has been switched off. If you can't do this yourself, get someone who can and have them explain it to you until you understand it. Have them read this article.

When patching microphones through a patchbay, either turn off the phantom powering first, and/or unplug the sensitive microphones (ribbons) before repatching any of these lines.

Phantom Power and Magnetized Transformer Cores

A discussion on the Ampex listserve led to some further thought *vis a vis* the issue of hot patching (connecting or disconnecting a microphone without first turning off the phantom powering). Although the theory says that you can do this, practice says that you need a few more details before deciding what the practice shall be in your own facility.

This issue at hand is the magnetization of the transformer core caused by unbalanced current pulses through one of its windings. Once this occurs, it is quite difficult (although not impossible) to undo. Once the core has been magnetized, the transformer's low distortion characteristics go by the wayside.

First, let's look at the different ways that this can happen:

1. Hot patch a phantom powered microphone. If you can guarantee that both pin 2 and pin 3 make at *exactly* the same time, then this is not a problem. Problem is, you can't. This likely isn't as bad as #2 or #4. If the phantom feed system uses a resistor bridge across pins 2 and 3 of the XLR, then this cause is pretty much a non-issue.
2. Using a shorted mike cord with phantom power turned on. Varying degrees of problem result depending on the nature of the short and what happens at the instant of plug insertion.
3. A dufus with an ohmmeter
4. Plug in a mike having a grounded center tapped output with phantom power turned on. This is the same mechanism as #1.

So, if your board has input transformers, and/or if the mike has an output transformer, then it's good medicine to turn that phantom power off and wait a bit before plugging things in. If you know for sure that both the board and mike use resistor bridges to apply and get powering to/from the line, then it doesn't matter if you turn the phantom powering off first or not. If this isn't the case, it would be worth the piece of mind to modify the phantom power switching to ensure that there was a discharge path for the input stuff, and/or monitor it with a meter to see how long it takes things to die.

I guess there IS a GOOD reason to have phantom power switches on every input!

Phantom Power and Mic Splitters

There are several ways to split a microphone's output. They range from cheap (wye-cord or paralleled inputs) to expensive (BSS active mic splitter). Somewhere in the middle there is the venerable old transformer coupled splitter.

For sure, the active splitter is the best solution. The microphone only sees one load, which is the splitter's input impedance. Each of the outputs is buffered from the rest. The only rub is that the splitter's sonics are imposed on everyone that takes one of the split outputs.

In this case, the splitter supplies phantom power to the microphone and could care less what the mixers are doing.

Transformers are the next-best solution. Unfortunately the usual 1:1:1 transformer still parallels the input impedances of the two consoles. From a microphone loading perspective, this is no better than simply paralleling the mixer inputs. If you run into a nasty grounding situation, the transformer coupled

splitter will save your bacon. For phantom power purposes, what usually happens is that the FOH mixer gets a direct feed from the microphone and the monitor mixer gets a transformer-isolated split (in reality, a parallel feed) from the mic. The FOH mixer supplies phantom and everyone is happy. In a really nasty scene, both mixers might get transformer splits, and the splitter itself has phantom feed resistors and a 48V DC source.

In many instances, simply connecting the mixers in parallel works just fine. You can get into trouble if you aren't really careful about how things get grounded, especially with regard to the FOH console's ground system and the monitor console's ground system. If either of the boards has pin-1 trouble, this may be exacerbated. I've used this method myself for over 30 years and only occasionally had problems. In each instance, I can trace the source of trouble back to not following the rules. The big rub with parallel connected mixers is the old question: "Who supplies phantom?"

Some consoles don't like having phantom power applied via their microphone inputs when their internal supply is not switched on.

This is especially true of mixers that don't have individual phantom on/off switches or at least switches for groups of inputs. What happens here is that the manufacturer took the easy way out to turn the phantom power supply on and off: they switched the AC feed to the phantom supply rectifier. The voltage regulator that supplies 48V to all of the phantom feed resistors is always connected. When the mixer that isn't supplying phantom has it's inputs 'backwashed' with phantom power, the regulator doesn't quite know how to deal with this voltage coming at it from what it thought was its output.

The cure here is a simple series diode connected at the output of the regulator, and a large electrolytic capacitor to ensure that the output impedance of the regulator stays low in spite of the series diode.

The Mackie SR series mixers that I had a hand in have this series diode. Other mixers may, and it pays to check the schematic for this feature.

Now, having said that, let's explore the situation when both mixers supply phantom.

Quite simply, we have two 48V supplies driving the mic through 6800 ohm resistors on each leg. As far as the microphone is concerned, it's no different than if we halved the value of the feed resistors to 3400 ohms each. So, again from the microphone's perspective, the mic is seeing 48V thru 3400 ohms in parallel with 3400 ohms. This means that twice the current is available to every microphone. Looking inside the microphone, we typically have a zener diode with or without some extra current limiting resistance. So, the effect of the extra current is that the zener diode works a bit harder.

I have used sound systems where both the monitor and FOH boards supplied phantom powering to the microphones. I've done this with just about everything from cheap electrets to Neumanns and AKGs. It works, and the microphones have all said, "so what." The Neumanns are especially immune to this. The 9-52V AKGs (old C451's, C414's) have to work their zeners a bit harder, but it's no big deal. They're a

worst case example. My C451's and C414's still work just fine, thank you.

Links

- Rectifier and filter circuit for a DIY supply.
- Voltage regulator for the phantom supply
- Applying phantom power to a mixer's inputs
- Using a non-phantom processor with a phantom powered mike.
- A complete phantom power supply for any mixer with a balanced input.
- All about T-powered condenser microphones
- P48 to T power adaptor #1
- P48 to T power adaptor #2
- Home

Model SM58 Dynamic Microphone Specification Sheet

MODEL SM58®
UNIDIRECTIONAL DYNAMIC MICROPHONE

OVERVIEW

The Shure SM58® is a unidirectional (cardioid) dynamic vocal microphone designed for professonal vocal use in sound reinforcement and studio recording. A highly effective, built-in, spherical filter minimizes wind and breath “pop” noise. A cardioid pickup pattern isolates the main sound source while minimizing unwanted background noise. The SM58 has a tailored vocal response for a sound which is a world standard. Rugged construction, a proven shock mount system, and a steel mesh grille ensure that even with rough handling, the SM58 will perform consistently. Outdoors or indoors, singing or speech—the SM58 is the overwhelming choice of professionals worldwide.

FEATURES

- *Frequency response tailored for vocals, with brightened midrange and bass rolloff*
- *Uniform cardioid pickup pattern isolates the main sound source and minimizes background noise*
- *Pneumatic shock-mount system cuts down handling noise*
- *Effective, built-in spherical wind and pop filter*
- *Supplied with break-resistant stand adapter which rotates 180°*
- *Legendary Shure quality, ruggedness, and reliability*

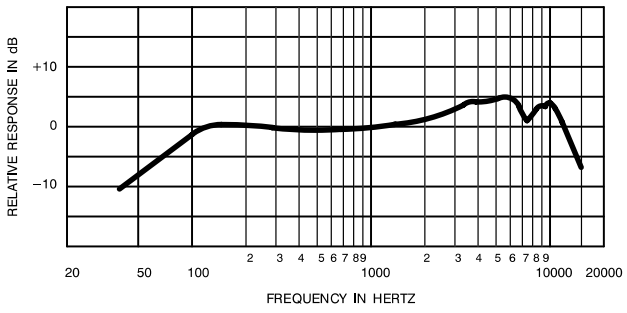
Models

SM58
SM58S (With On/Off Switch)

SPECIFICATIONS

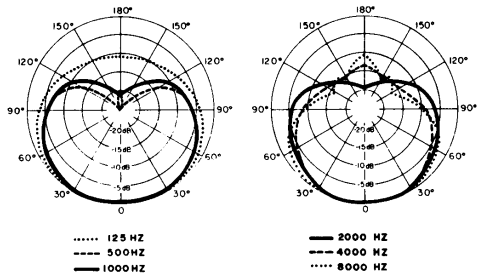
Type
Dynamic (moving coil)

Frequency Response
50 to 15,000 Hz



TYPICAL FREQUENCY RESPONSE

Polar Pattern
Unidirectional (cardioid), rotationally symmetrical about microphone axis, uniform with frequency (see Figure 3)



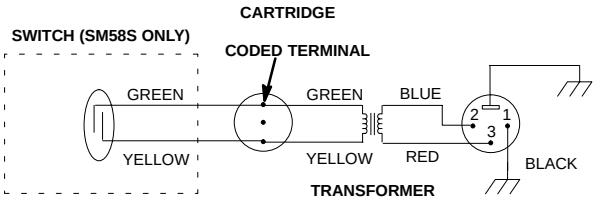
TYPICAL POLAR PATTERNS

Sensitivity (at 1,000 Hz Open Circuit Voltage)
–54.5 dBV/Pa (1.85 mV)
1 Pa = 94 dB SPL

Impedance
Rated impedance is 150 Ω (300 Ω actual) for connection to microphone inputs rated low impedance

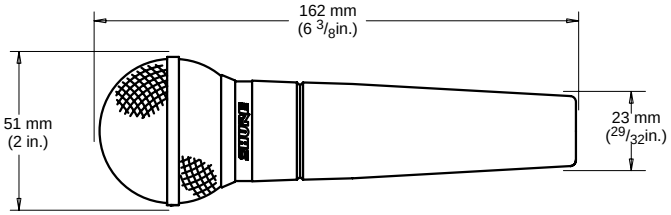
Polarity
Positive pressure on diaphragm produces positive voltage on pin 2 with respect to pin 3

Internal Connections



INTERNAL CONNECTIONS

Overall Dimensions



OVERALL DIMENSIONS

Connector

Three-pin professional audio connector (male XLR type)

Case

Dark gray, enamel-painted, die cast metal; matte-finished, silver colored, spherical steel mesh grille

Swivel Adapter

Positive-action, break-resistant, adjustable through 180°, with standard $^{5}/_{8}$ in.–27 thread

Net Weight

298 grams (10.5 oz)

FURNISHED ACCESSORIES

Swivel Stand Adapter A25D
Storage Bag 26A13

REPLACEMENT PARTS

Cartridge R59
Screen and Grille Assembly RK143G

OPTIONAL ACCESSORIES

Windscreen . . . . . A58WS Series (8 colors available)
Desk Stand S37A, S39A
Isolation Mount A55M
Dual Mount A26M
Cable (7.6 m [25 ft] C25E, C25F

Microphones part 1 - The Carbon Era

- Microphones part 1 - The Carbon Era 1915-1925
- Microphones part 2 - The Electrical Era 1925-1945
- Microphones part 3 - The Modern Era 1945-present
- Microphone History Sources

1915 Western Electric 323
from Wilson House

1915 single-button mic
from AT&T Archives

1915 single-button mic
from AT&T Archives

1928 film clip on how the
telephone works

The single-button carbon microphone was widely used in the "candlestick" telephone made by Western Electric, models 299 and 323, from 1915 to the 1920s. This type of microphone was suitable for the limited voice range of telephones, but carbon produced too much noise for music recording.

1920 Western Electric 387
from Wilson House

Hoover with radio mics
from *Literary Digest*
1929/05/04

The spring-mounted double-button carbon microphone was designed to keep the carbon granules inside the transmitter from moving and touching the diaphragm during operation. Western Electric introduced the first model 373W in 1920, followed by improved the improved model 387, the smaller 389 (used on top of a candlestick stand as the 1441CJ announce mic), and the model 600A.

1925 WE 1B used by Hoover
from *Literary Digest*
1925/12/19

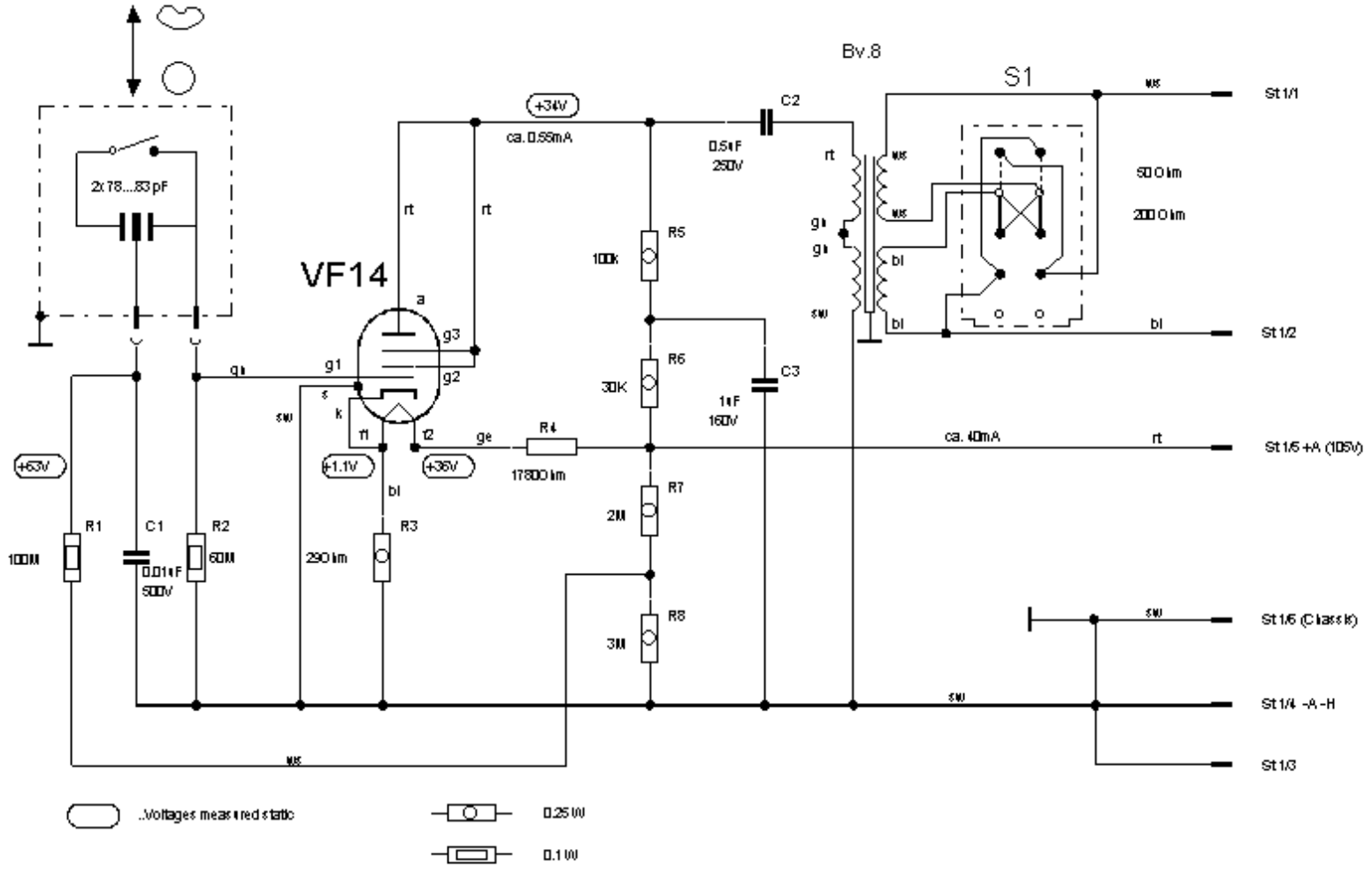
373-W in the WE 1B mount
#25603 from AT&T Archives

The Western Electric 1A or 1B housings enclosed the same carbon transmitters used in the spring-mounted versions. The spring-mounted and enclosed versions were widely used in radio broadcasting 1921-1931.

- Microphones part 1 - The Carbon Era 1915-1925
- Microphones part 2 - The Electrical Era 1925-1945
- Microphones part 3 - The Modern Era 1945-present
- Microphone History Sources

© 1999-2006 by Steven E. Schoenherr. All rights reserved.

Return to Recording Technology History Notes | this page revised 4/22/06


Bv.8
S1
VF14
2x78...83pF
+34V
ca. 0.55mA
C2
0.5nF
250V
R5
100k
R6
30K
C3
1nF
160V
R4
17800 km
R7
2M
R8
3M
R1
100M
C1
0.01nF
500V
R2
60M
R3
290 km
+63V
+1.1V
+36V
ca. 40mA
50 Ohm
200 Ohm
St 1/1
St 1/2
St 1/5 +A (105V)
St 1/6 (Class k)
St 1/4 -A -H
St 1/3
..Voltages measured static
0.25 W
0.1 W
VF14M
seen from below socket
St 1
Tuchel T2045
plus of connector
Condenser Mikrophone
U47 - NG
U47-00-00-01-S(4)
Date of original diagram: 28.1.1959
Georg Neumann
Laboratorium für Elektroakustik G.m.b.H.
Berlin-Schoeneberg
redrawn by:
supersonic
COMMUNICATIONS

Elliott Sound Products **Amplifier Basics - How Amps Work (Part 1)**

Copyright (c) 1999 - Rod Elliott (ESP)
Page Last Updated 06 Apr 2005

Articles Index

Main Index

Amp Basics Index

- Part 1 - Valves (Vacuum Tubes)
  - Valve Characteristics
  - Valve Current Amplifier
  - Valve Power Amplifiers
  - Summary

Part 1 - The Valve (Vacuum or Thermionic Tube)

In the beginning the vacuum tube was the only way to amplify, and valves (or "tubes") survive to this day, with a dedicated following of "believers" who are convinced that the development of the transistor (or indeed, any semiconductor) was fundamentally a bad idea. This is not a discussion I intend to follow - I intend simply to state how these devices amplify a signal, and the factors that determine voltage and current gain.

The basic amplifying valve (there are many different types) has three elements. These are

- The anode (Plate) - Collects the electrons released by the cathode
- Cathode (indirectly or directly heated) - The source of the electron flow
- Control Grid (Grid) - controls the flow of electrons

When a positive voltage is applied to the anode with respect to the cathode, an electron stream

is emitted from the cathode and flows to the anode, completing the circuit. The grid is a fine coil of wire, suspended between the other two elements. A negative voltage on the grid (with respect to the cathode) will repel some of the electron stream, causing the current to be reduced. If the voltage on the grid were to be varied, then the cathode to anode current must also vary, and an amplifier is born. Figure 1.1 shows the basic circuit of a valve voltage amplifier.

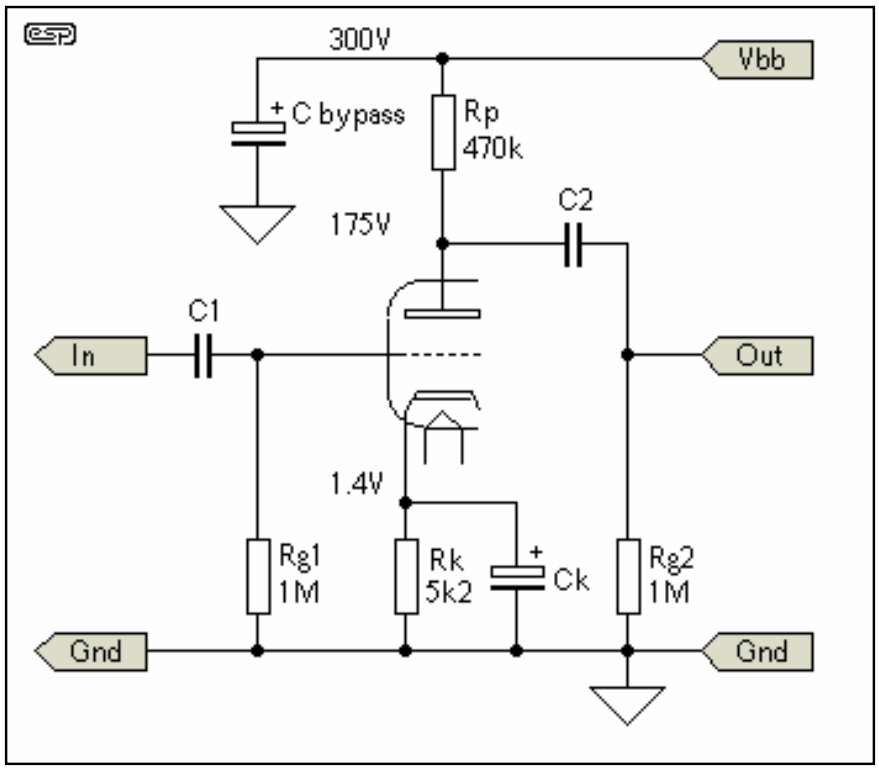


Figure 1.1 - A Basic Valve Voltage Amplifier

This circuit configuration is known as "common cathode", because the cathode reference point (earth) is common to both input and output. By placing a resistor "Rk" in the cathode circuit, a voltage is developed because of the current flow. If the grid is referenced to earth (ground), then the grid is negative with respect to the cathode. The voltages shown on the circuit are typical of a single element of a 12AX7 twin triode. Note that the two valve pins that are not connected are for the heater. This is used to heat the cathode so that it emits electrons more readily.

The cathode resistor will cause the circuit to reach a stable current, where any attempt at increasing the cathode current will cause a greater voltage across the resistor, making the grid effectively more negative and reducing the current. A point of equilibrium is quickly reached, where the circuit operates in a stable manner. This is known as cathode biasing, and is most common with signal level and very low power amps.

By applying a varying voltage (the signal) to the grid, the current between cathode and anode will vary too. Since the anode load is a resistance, a varying voltage will be developed which will (hopefully) be greater than the voltage applied to the grid. The input voltage has been amplified.

Because the signal voltage on the grid is "fighting" the attempt of the cathode resistor to maintain the current through the valve at a constant value, this is a form of feedback. It is also known as cathode degeneration. The name is of no consequence, because as local feedback, it will improve the linearity of the stage but reduce the gain. In reality, the improvement in linearity is only minor, and leaving out the capacitor can increase noise in sensitive circuits - especially hum that is induced into the cathode from the heater, which was nearly always operated from AC in the past, but DC heater supplies are now common.

Where the cathode is directly heated (the filament has the oxide coatings directly applied), DC operation is mandatory, or hum would be inevitable. Directly heated cathodes will always emit electrons unevenly, because of the voltage gradient across the filament. The most postive end will have the greatest emission, since the negative voltage between cathode and grid is lowest.

Note that for indirectly heated cathodes, the heating element is called the heater, but for directly heated cathodes it is more commonly referred to as the filament (as in the heated filament of a light bulb).

Because valves have a relatively low voltage gain, it is common to bypass the cathode resistor with a capacitor to defeat the local feedback and extract as much gain as possible, as is shown in Figure 1.1. The gain (or more correctly, the transfer characteristic) of a valve is sometimes measured in mA/V - which tells us how many milliamps change in anode current will occur with a grid voltage change of 1 Volt. Another common way to describe this is "mu" or amplification factor. Yet another value is common with valves - the "conductance" (A.K.A. mutual conductance or transconductance), which is the opposite of resistance, and is expressed in Mhos (Ohms backwards - seriously!).

One problem with valves has always been the number of different methods used to describe what is essentially the same thing. Depending entirely which book you happen to be reading, you will see the effective gain quoted as mA/V, mutual conductance ("g_m", in Mhos or more commonly uMhos), or the equally obscure term "Amplification Factor", none of which has any direct relevance to the gain you can expect without further calculation.

The output impedance of the circuit of Figure 1.1 is 470k - the same as the plate resistor. Rg2 is the grid resistor for the following stage, and at 1M, loads the output and reduces gain.

Valve Characteristics

There are three main characteristics that are quoted for any given valve. These are:

- Amplification Factor (mu) - This parameter compares the effectiveness of the control grid voltage to the plate voltage in changing the plate current
- Plate Resistance - This is the equivalent resistance of the internal cathode to plate circuit,

since the plate voltage is applied across the valve, and the current flows through it
- Mutual Conductance (AKA Transconductance) in uMhos - This shows how effective the grid is in controlling the plate current.

One important thing to realise about valves is that everything changes. The characteristics vary widely with plate voltage, load resistance, bias current and just about everything else you can think of. Despite this, it is still possible to design a circuit using valves that will be repeatable from one unit to the next, provided the designer knows what s/he is doing.

A typical signal valve (such as the 12AX7 high mu dual triode) has a plate resistance of 2k5, an amplification factor (mu) of 100, and a g_m (using the circuit of Figure 1.1) of about 1600uMhos, which can (by simple mathematics) be converted into a figure of 1.6mA/V, meaning that a change of 1V on the grid will cause a change of 1.6mA in the anode current. This does not actually mean what it says, since the valve might be quite incapable of sustaining an anode current of 1.6mA under all circumstances. However, a change of 0.1V at the grid *can* cause a change in plate current of 0.16mA - the measurement is typically "normalised" to make comparison easier (just like they don't in the supermarket).

Let us now have a look at how the valve amplifies the signal. The transfer curve in Figure 1.2 shows the input waveform applied to the grid, at any convenient frequency. As the signal becomes more positive, the valve draws more current, until at the peak of the waveform, the grid voltage has been made 0.1V more positive than it was before. Therefore, the anode current is 0.16mA greater that it was before. Using Ohm's Law, 0.16mA with a resistance of 470k means that the anode voltage is 75 Volts lower than when in the idle (or quiescent) state.

This would seem to indicate that the valve has a gain of (75 / 0.1) 750 - not a chance! The circuit of Figure 1 will have a typical voltage gain (Av) of around 75. As I said - everything changes. The only way to be certain what a valve will actually do is consult the manufacturer's data, and refer to the transfer curves for the mode of operation you wish to use. Supply voltage, plate current, plate voltage and the impedance of the next stage all have a profound effect on the performance of a valve.

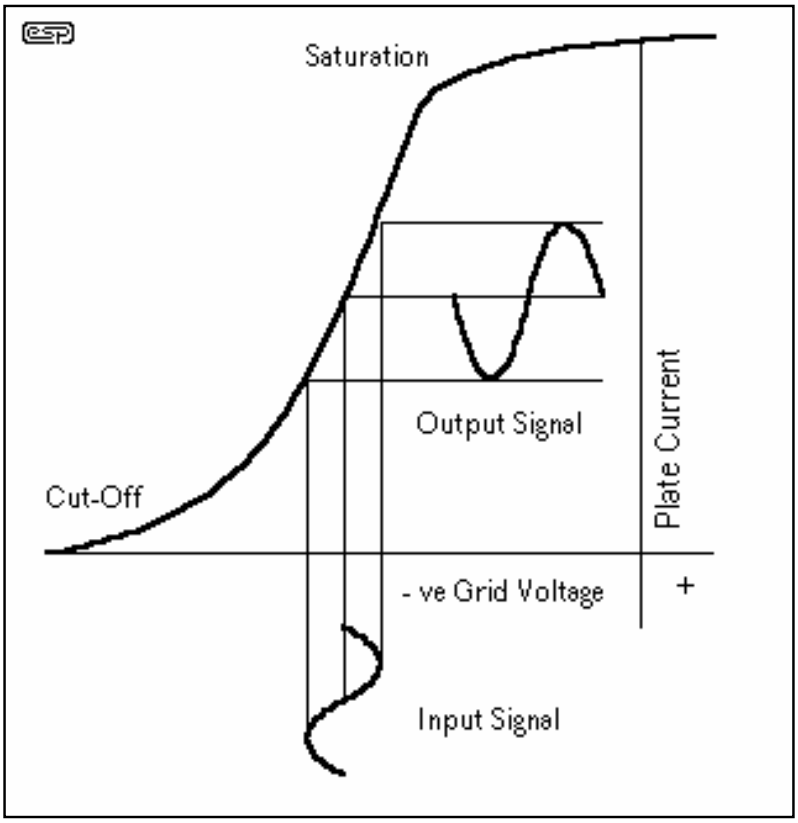


Figure 1.2 - Typical Valve Transfer Curve

As can be seen from Figure 1.2, the transfer curve is not linear, which means that as the valve approaches cut-off (turned off completely) or saturation (turned on completely) the characteristics change, and distortion is introduced. A (very) rough estimation of maximum rms output voltage to keep distortion below 1% is about 0.1 of the quiescent plate voltage. Thus, with a plate voltage of 150V at idle, the maximum output voltage will be 15V RMS. This assumes that the valve has been biased correctly in the first place. From the graph we can see that at high values of negative grid voltage the valve will cut off, while at low (or positive) grid voltage, the valve is turned on as hard as it can.

A valve can be thought of as having an infinite input impedance (although this is never realised in practice). The input impedance is approximately equal to the value of the grid resistor for audio frequencies. The output current is therefore controlled by a voltage at the grid, so the valve migh be considered a voltage controlled current source (or VCCS).

Valve Current Amplifier

Figure 1.3 shows a valve current amplifier, commonly known as a cathode follower, or common plate (because the plate circuit is common to both input and output - for AC signals only). Although this circuit can provide a useful increase of current, and an equally useful decrease in output impedance, it has a voltage gain that is less than unity. Typically, this will be about 0.8 to 0.9, so for every volt of signal applied to the input, we only get about 850mV output.

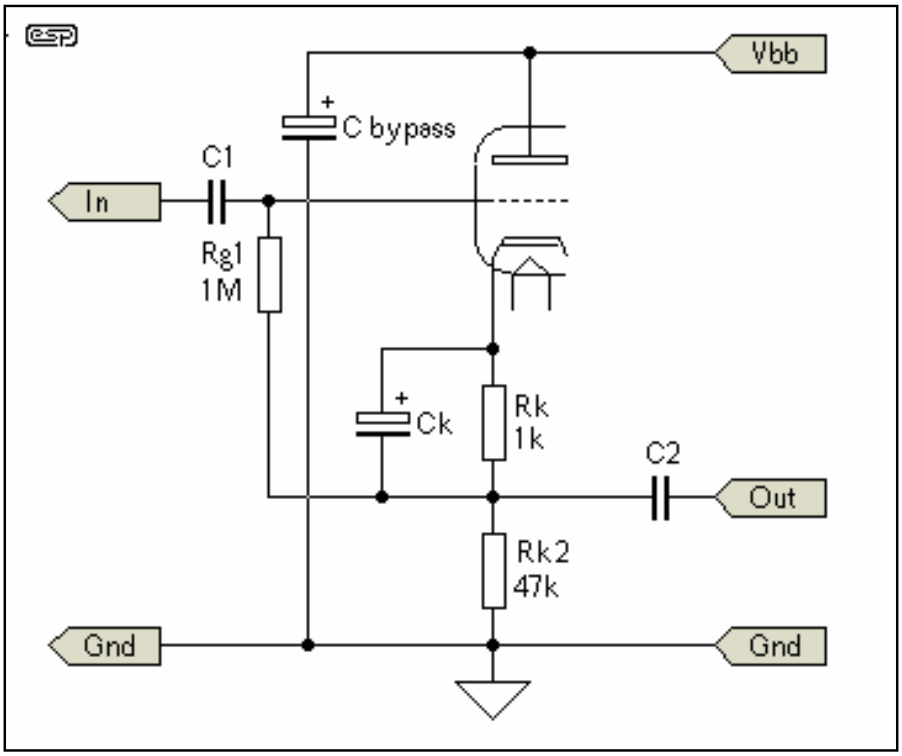


Figure 1.3 - Cathode Follower Current Amplifier

The cathode follower is typically used where a low impedance output is desired, since the output impedance of most valve circuits is rather high (equal to the value of the plate load resistor). Simply attaching a low impedance load to a voltage amplifier stage will cause the output level to be dramatically reduced, so the current amplifier (cathode follower) is a useful stage. The output impedance of the circuit of Figure 1.3 can be expected to be about 1/10th the value of the cathode resistance Rk2 - but this is highly dependent on the valve itself, and its operating current. The available current is very low, so the circuit will not be able to drive a load much less than Rk2, or 47k. Remember that output impedance and drive capability are not related.

Note that the grid must still be biased to an appropriate voltage negative with respect to the cathode. The bypassed cathode resistor is used as before, but the grid is connected to the bottom of this resistor, and not ground. If it were connected to ground, the circuit would be capable of only very small signal levels before it distorted.

Valve Power Amplifier

Finally, we can combine a voltage amplifier stage and a current amplifier stage, and get a power amplifier. Cathode followers are very unusual in the valve amplifier, and it is more common to use a "push-pull" output stage, generally using a transformer to match the high voltage and relatively high impedance of the output valves to the impedance of the speaker.

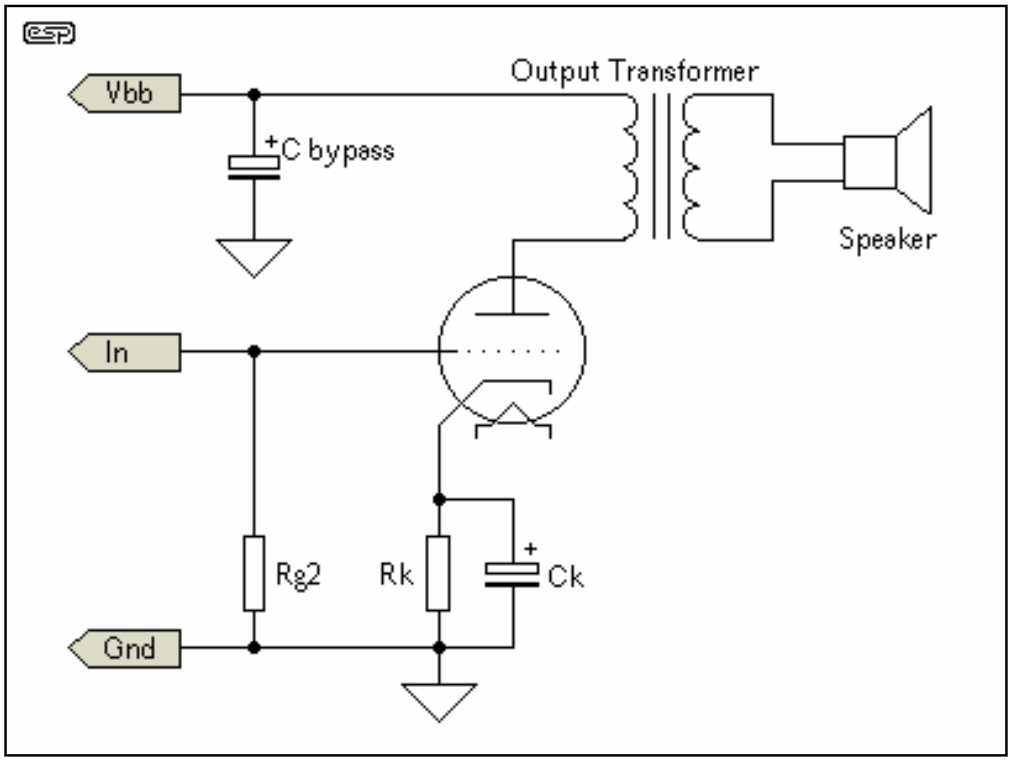


Figure 1.4 - Basic Valve Power Amplifier

Figure 1.4 shows a very basic valve power amplifier, using a triode in "single-ended" mode. The output transformer converts the high voltage, high impedance plate circuit of the valve to a low voltage, low impedance signal for the loudspeaker. Because the primary of the output transformer must carry the full DC quiescent current of the valve (which will be a large, high current unit), it needs a very large core of laminated steel.

Interestingly, these inefficient and high distortion amplifiers have made a comeback in recent years. However in the heyday of the valve, the efficiency and distortion of these circuits was such that they were replaced in nearly all installations by more efficient and lower distortion circuits, such as that shown in Figure 1.5.

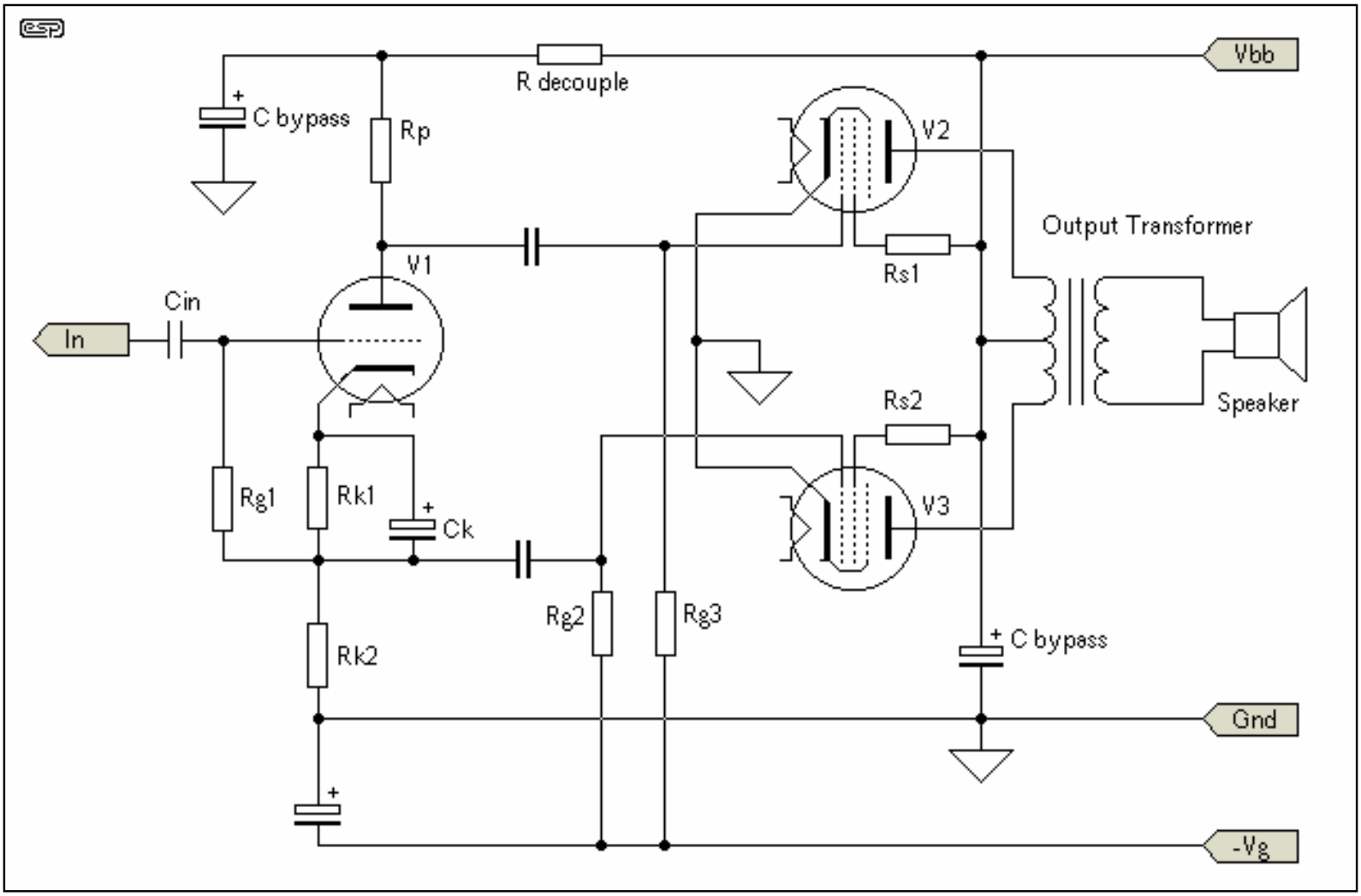


Figure 1.5 - Push-Pull Valve Power Amplifier

The valves shown for the output are called pentodes (from penta - five), having 5 electrodes instead of the three for a triode. The second grid (called the screen grid, or just screen) increases the gain of the valve dramatically, while the third grid, the suppressor, prevents what is called "secondary emission" from the plate. The screen accelerates the electron flow so much that electrons bounce off the plate, or dislodge others. The addition of the screen gives the valve some nice characteristics, such as much higher gain, but also some nasty ones (lower linearity, more distortion), which the suppressor counteracts to some degree. The suppressor grid is almost always connected internally to the cathode. It is not uncommon for designers to connect pentodes as triodes, by connecting the screen and plate together.

The first stage of the circuit is interesting, and is called a phase splitter. It is a combination of a voltage amplifier and a current amplifier, having equal values of resistance in each circuit (i.e. Rp = Rk2). Because all valves have the same "polarity", they cannot be used like transistors or MOSFETs, but must be driven with their own signal of the correct polarity.

The incoming signal is therefore sent "as is" to one valve (from the cathode circuit), and is inverted for the other - hence the term push-pull. As one valve "pushes" the anode current lower, the other simultaneously "pulls" it higher. In a properly designed circuit, the two output

valves will pass the signal between them with little disturbance. Any disturbance in this region is called crossover distortion, because it happens as the signal crosses over from one valve to the other.

Notice something else quite different. The cathodes of the output valves are connected directly to earth, and the grids are supplied with a negative bias from a separate -ve power supply. This is the most common method of biasing output valves in high power circuits, having a much greater efficiency than cathode biasing.

For many large output valves, it is not even considered a good idea to use cathode biasing, because the amount of negative grid voltage required is too high. Voltages of up to -60V are not uncommon with high power pentodes or another common type, beam power tetrodes (I will not cover these in more detail, but there is much information to be found on the web). Using cathode bias for this sort of voltage is inefficient and reduces the output power dramatically.

Valves - A Summary

The above is but a very small offering from the world of the vacuum tube. As I said in the introduction, this is not designed to be a complete electronics training course. The circuits presented are basic only, which is to say that they will all work, but are not optimised in any way.

For further reading, the most highly recommended work is the rather old (but still considered *the* reference manual) "Radiotron Designer's Handbook", by F. Langford-Smith and originally published by Amalgamated Wireless Valve Co. Pty. Ltd. in Australia. My copy is dated 1957, but it has recently been republished (although I think it is quite expensive, unfortunately).

Overall, the valve is still an almost mystical thing, but in all honesty, modern amplifiers using transistors or MOSFETs are so vastly superior in terms of fidelity, efficiency and reliability, that I really don't see what all the fuss is about. Having said this, I was using a valve preamplifier on my own system until recently.

There is no doubt that valves do have some very nice characteristics, and for guitar amplifiers there are few guitarists who would argue otherwise. A soft overload behaviour means that a valve amp does not sound as harsh as a transistor amp when it is overdriven - which is great for guitar, but a hi-fi should never be overdriven anyway, so the point is moot.

The problems that befall valves are many, and include

- **Fragile** - The glass envelopes are very thin, and are easily broken.
- **Limited Life** - Even if a valve is operated well within its ratings, it still has a finite life. The main causes of valve failure are cathode emission degradation (happening all the time),

and gas, when small amounts of air "break" the full vacuum.

- **Microphonics** - All valves tend to be slightly microphonic, which is to say that they act as a microphone. This can cause additional colouration to the signal if the sound from speakers vibrates the amplifier.
- **High Voltage** - Having to ensure that the 600V DC typical of a high power amp does not "flash over" valve bases is a constant headache, and ensuring that these voltages are kept well away from small fingers is mandatory. High voltage capacitors are also more expensive than lower voltage ones.
- **Heaters** - Valve cathodes must be operated at the correct temperature so they emit electrons properly, and "boil off" contaminants. If the heater voltage is too low, the cathode will become poisoned, and the valve is useless. The heater power used is all wasted, in that none of it is turned into sound.
- **Output Transformers** - The output transformer for a valve amp is expensive, bulky and heavy. It introduces its own distortion components, which are difficult (or impossible) to eliminate completely.
- **Heat** - All valve amps run hot. The valve will not work unless it is hot, and the heat causes problems for other components, shortening their life. The heat is all wasted energy.
- **Damping** - Valve amplifiers nearly all have a low damping factor, caused by a relatively high output impedance. Speakers must be very well damped indeed to work well with any valve amp, or the bass will become poorly defined, and crossover networks (which rely on a very low amp impedance) may not work as well as intended.

On the positive side, valve amplifiers have a "warm" sound, partly because of the low order harmonic distortion introduced. A good valve amp will also have a very wide bandwidth, and will have an easy job driving loads that cause solid-state equipment to have severe heartburn (or just blow up on the spot).

At low levels, valve equipment has vanishingly small distortion levels, and when all is said and done, there is something nice about little glass tubes, with little lights inside, making your music.

Previous Next

Articles Index

Main Index

Amp Basics Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999-2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Elliott Sound Products **Amplifier Basics - How Amps Work**

Copyright (c) 1999 - Rod Elliott (ESP)
Page Last Updated 06 Apr 2005

Articles Index

Main Index

Amp Basics Index

- Part 2 - Bipolar Transistors
  - Transistor Characteristics
  - Transistor Current Amplifier
  - Transistor Power Amplifiers
  - Summary

Part 2 - Bipolar Transistors

Since it was invented, the transistor (from "transfer resistor") has come a long way. Early transistors were made from germanuim, which was "doped" with other materials to give the desired properties required for a semiconductor. In the beginnings of the transistor era, nearly all devices were PNP (Positive Negative Positive), and it was very difficult to make the opposite (NPN) polarity. The NPN transistors that were available at that time were low power and did not work as well as their PNP counterparts.

When silicon was first used, the opposite was the case, and for quite some time the only really high power devices available were all of silicon NPN construction. More recently, it has become possible to make NPN and PNP transistors that are almost identical in performance. Germanium is rarely used, although some examples are still available.

All transistors have three "elements":

- Emitter - analogous to the cathode of a valve. The emitter only emits electrons in an NPN device, but this is of no consequence
- Base - the controlling terminal. A current at the base controls the current through the transistor
- Collector - basically, collects the emitted electrons. Somewhat analogous to the plate of a valve

A transistor can be represented as two diodes, with a junction in the middle. This is shown for both polarities in Figure 2.1. This is only an analogy, and connecting two discrete diodes in this manner will not produce a transistor, because the point where they meet must be a common junction on the same piece of silicon (or germanium) - hence (in part) the term Bipolar Junction Transistor. The "bipolar" term means that transistors use "charge carriers" of both polarities - positive and negative, or minority and majority.

Since the base to collector junction is reverse biased in normal operation, there will be no current flow. It is the action of injecting current into the base that causes current flow in the collector circuit. I do not intend to explain the exact conduction mechanism, as it is somewhat outside the scope of this article.

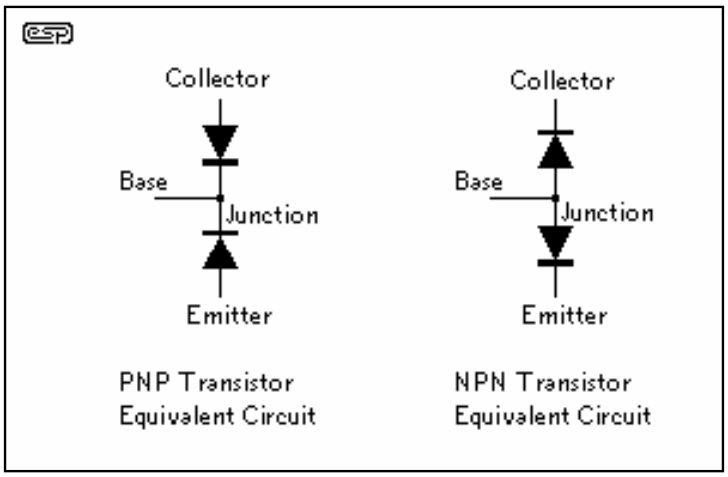


Figure 2.1 - Analogous Depiction of Transistors

This is very convenient, because it gives us an easy way to check if a transistor is likely to be good or bad, simply by measuring the "diodes". Early PNP germanium devices would actually work equally well if the emitter and collector were reversed, but devices are now optimised to maximum performance, so this trick is not as successful (it does still work, but the device gain is *much* lower when the terminals are reversed).

To make the transistor actually do something useful, it is necessary to bias it correctly. This is done (having selected a suitable collector resistance) simply by applying enough base current to ensure that the collector is at 1/2 the supply voltage. In the same way that the plate load resistor determines the output impedance of a valve amplifier, the collector resistor determines the output impedance of a transistor amplifier. Unlike a valve, the transistor is not said to have a "collector resistance" as in the equivalent resistance between emitter and collector, because this is not relevant to the operation of a transistor.

Figure 2.2 shows three methods of biasing a transistor. Of these Figure 2.2a is the least usable, because there is no mechanism to ensure that the circuit will be repeatable with different devices or with temperature. Variations caused by temperature are (and always have been) a real problem, and it is necessary always to ensure that the circuit has some feedback mechanism for DC operating conditions to ensure stability. Different transistors (of the same type and even from the same manufacturing run) will have different gains, and this, too, must be compensated for.

Assume that the gain of the transistor is 100 (exactly). This means that for 1mA of base current, 100mA of collector current will flow. The emitter current is the sum of the base and collector currents. To bias the transistor we need only meet this criterion (in theory), and everything will be well. With a Supply voltage (Vcc) of 20V, we want to have 10V at the collector, to allow maximum voltage swing. This will allow the voltage to go to +20V or to 0V, however the signal will be badly distorted by then.

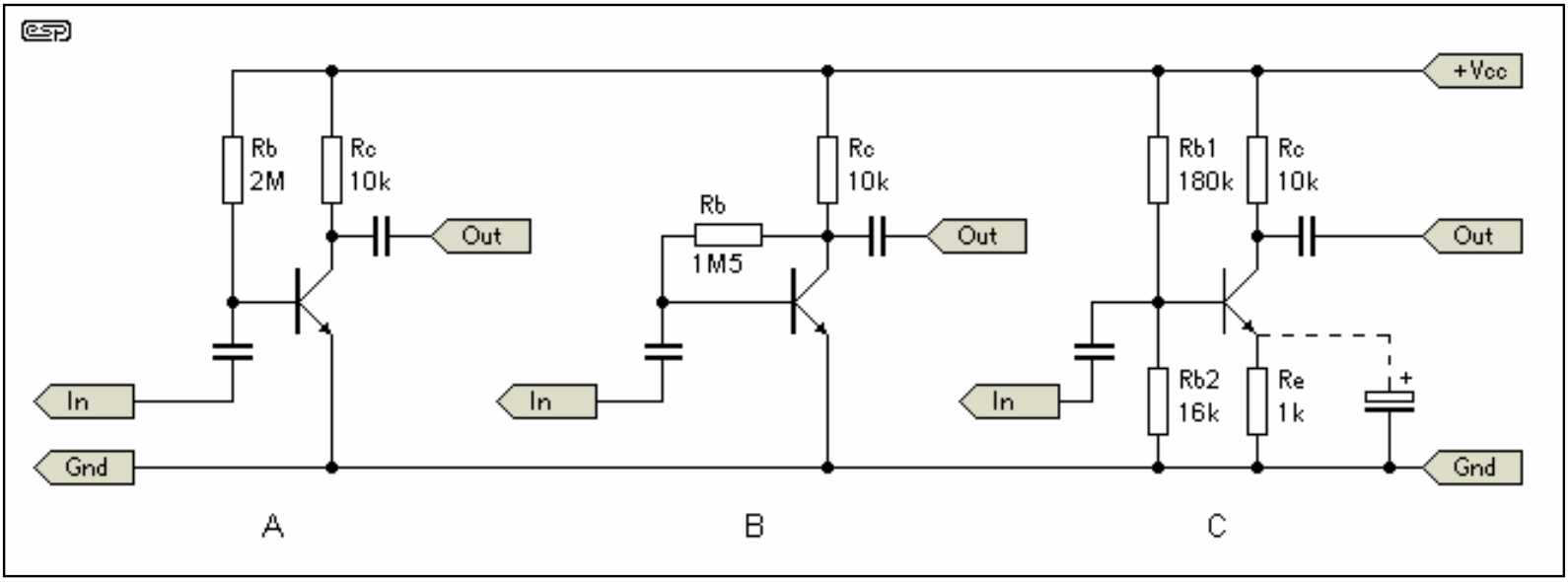


Figure 2.2 - Biasing a Transistor Voltage Amplifier, Three Methods

Figure 2.2b is a simple way to achieve this, but has some drawbacks. Because the bias resistor (Rb) is supplied from the collector circuit, it will have some of the collector current flowing in it. This will introduce negative current feedback, which at DC stabilises the circuit, but with the AC signal makes the input impedance very low, as well as reducing gain for any finite value of source impedance. This is not necessarily a drawback, however, as the feedback also reduces the distortion components.

This problem is overcome with the circuit in Figure 2.2c, with a bias divider providing a fixed voltage reference, and the emitter resistor (Re) providing stabilising feedback as we saw with the valve voltage amplifier. In the same way as with a valve, this also provides feedback, increasing linearity and reducing gain. With a transistor we get one additional effect - the input impedance is increased (more on this subject later). Again, to achieve maximum gain, it is common to place a capacitor in parallel with Re to defeat the feedback for AC signals, allowing maximum gain.

To bias a PNP device, we use exactly the same circuitry, but the supply polarity is reversed, so the collector (and base) will have a negative voltage with respect to the emitter.

One of the major differences between valves and transistors is that once we have decided on a suitable biasing circuit (or specified a gain from the amplifier), we can make device substitutions with little or no change in performance, provided the transistors have similar basic parameters. Often the same circuit will work just as well with perhaps 10 or 20 different devices, all from different manufacturers.

Transistor Characteristics

I shall only discuss the basic characteristics of transistors (as with valves), and there is really only one variable parameter and two fixed parameters (which are the same for every silicon transistor) to deal with. With transistors, the parameters are not as interactive as with valves, and the circuit gain is not as reliant on the device gain as with valves. In the same way as with valves, there are small signal devices (low power), working all the way up to power devices, which can have collector current ratings of 50 to over 100A for some of the very large power transistors.

- Gain - (AKA DC Current Gain, h_{FE}, beta) - the measure of the base current versus the collector current. This figure is only slightly dependent on the collector voltage, but can be affected considerably by collector current.
- re - Internal emitter resistance. With silicon transistors this is **26 / Ie** (in mA). Ie is the emitter current.
- Thermal coefficient - the diode junction voltage of a silicon transistor is nominally 0.65V, and falls by 2.0mV / degree C (this applies equally to diodes and transistors)

As stated before, the gain of a transistor is dependent on collector current, but will normally be applicable over a fairly wide range. The gain normally falls at very low currents (compared to the device maximum), and again at high current (approaching the maximum rated collector current for a given device).

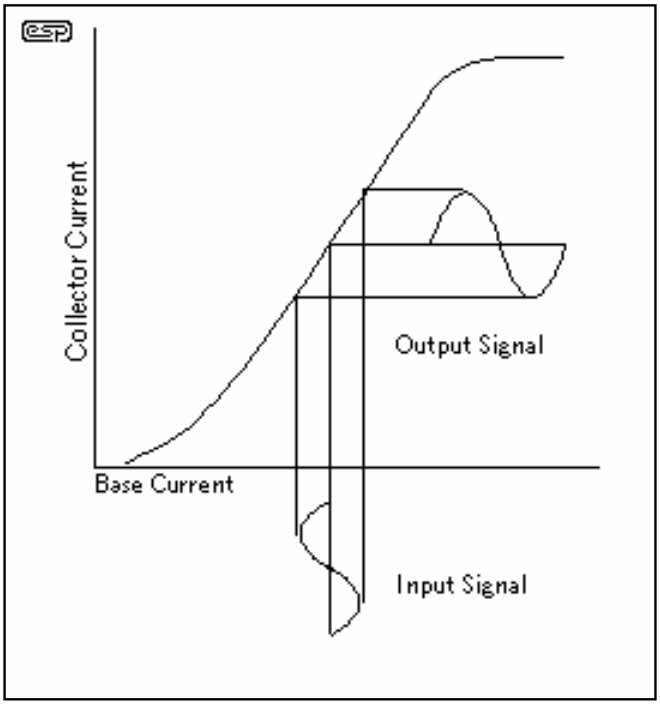


Figure 2.3 - Transistor Transfer Characteristics

The signal transfer curve is similar to that of a valve, and is shown in Figure 2.3. There is generally less distortion in the linear part of the curve, but because of the lower operating voltage, a transistor amp must work closer to the supply rail and earth, so distortion is often much higher with simple circuits such as those in Figure 2.2 than with an equivalent valve amplifier.

The major cause of distortion in small signal transistor amplifiers is the variation in the internal emitter resistance (re). Because transistors can tolerate a wider range of supply voltage and operating current than valves, it was common (when transistors were new and frightfully expensive) to try to extract as much voltage gain as possible from each device. This is no longer an issue, but the problem is still there, and it is necessary to take steps to prevent distortion. Very high gain circuits with global feedback are extremely common with transistor circuits, which renders the circuit immune from almost any variation of the device parameters, whether intrinsic (internal fixed) or manufacture dependent.

The gain of a transistor stage is approximately equal to the collector resistance divided by the emitter resistance (including the internal resistance re). So for the circuit of Figure 2.2c, the gain will be 9.75 without the emitter bypass capacitor, or about 384 with it installed. The distortion will be much higher with the emitter bypass in place, and it is uncommon to see these circuits any more.

The input impedance of a transistor voltage amplifier is low, and the output impedance is determined by the collector resistance (ignoring any feedback that may be applied from collector to base).

The input impedance is essentially determined by the gain of the device, and the value of emitter resistance (including the internal resistance), and in theory (that word again) is approximately equal to the emitter resistance multiplied by gain. The circuit of Fig 2.2a will therefore have an input impedance in the order of 2600 Ohms, Fig 2.2b will be very low because of the feedback, and 2.2c

(without the bypass capacitor) will have an input impedance of 100k - but as this is shunted by the bias resistors, the impedance will actually only be about 12k.

A transistor can be thought of as a current controlled current source (CCCS).

Transistor Current Amplifier

The current amplifier is much more common in transistor circuits than with valves, and is called an emitter follower (or occasionally common-collector). The emitter follower (like the cathode follower) has a voltage gain of less than 1 (or unity), but the difference is much less. Typically, the gain of an emitter follower circuit will be about 0.95 to 0.99 - depending on the operating current. The use of feedback to lower this further is very common, and output impedances of less than 1 Ohm are quite possible.

Figure 2.4 shows a standard configuration for an emitter follower current amplifier stage. It is common to bias the base to exactly 1/2 the supply voltage, using equal value resistors. I say "a" standard because there are many different configurations that can be (and are) used, including direct coupling, which is very common with transistor circuits.

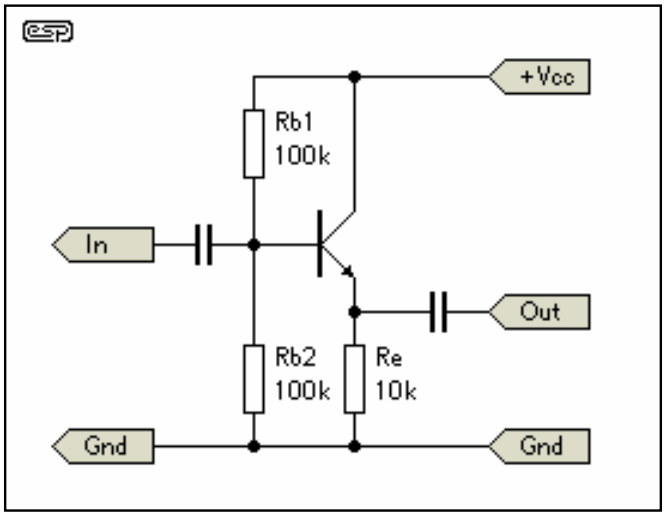


Figure 2.4 - Transistor Current Amplifier

One of the great attractions of transistors is their flexibility, which is considerably enhanced by having two polarities of device to work with. Because of this, circuits such as that shown in Figure 2.5 are common (or they were before the advent of opamps). Indeed, opamps themselves use the flexibility of transistors to the full, as can be seen if you have a look at the "simplified equivalent circuit" often published as part of the specification sheet for many opamps.

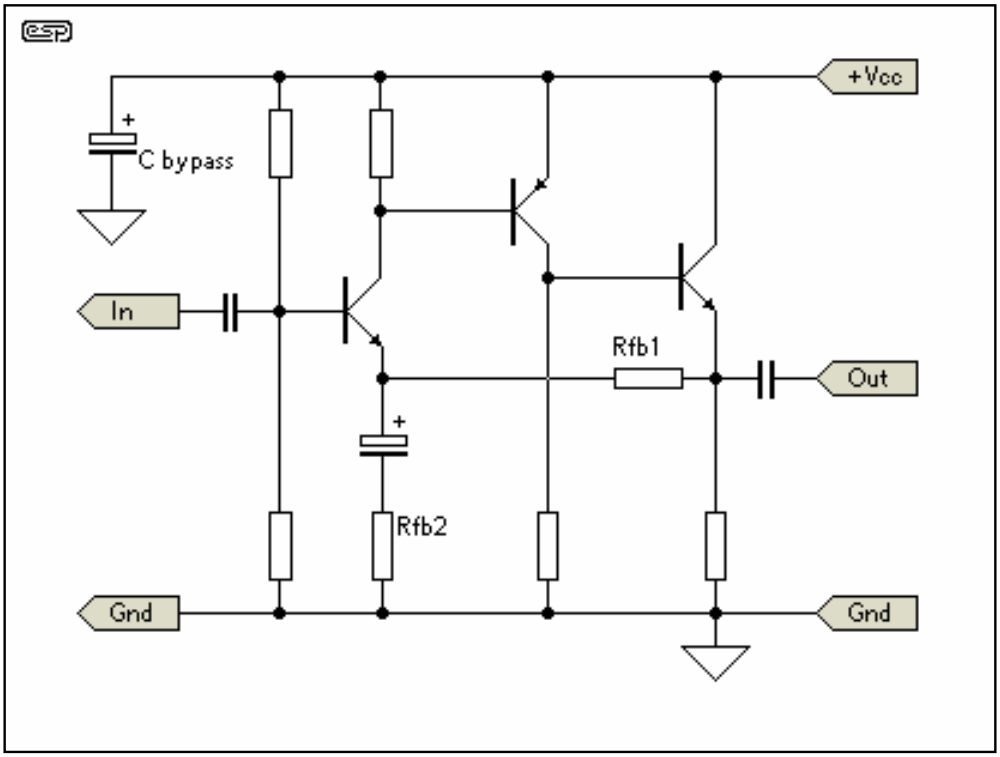


Figure 2.5 - A Typical Direct Coupled Transistor Amplifier

As can be seen, this amplifier uses an emitter follower for the output, is direct coupled within the circuit itself, uses both NPN and PNP devices, and has feedback to set a gain which is dependent only on the ratio of the two resistors Rfb1 and Rfb2. It is this sort of circuit that the opamp came from in the beginning, and there are still ICs (and small power amplifiers) that use similar circuitry internally. Regular readers may even recognise the basic circuit from the Projects Pages - essentially this is a discrete opamp, and will have a very high gain, which is brought back to something sensible by the feedback.

The actual gain is almost entirely dependent on the resistor values (for gains less than about 50 or so), and may be calculated by

Av = (Rfb1 + Rfb2) / Rfb2 where Av is voltage gain (Amplification, voltage)

So to obtain a gain of 20, Rfb1 would be 22k, and Rfb2 1k2 - this is actually a gain of 19.33, representing an error of 0.3dB. This gain is so stable that a completely different set of transistors from a different manufacturer would make no difference to measured gain performance. Other factors, such as noise or distortion must vary with the quality of the active devices, but the changes will generally be very subtle, and may not be noticeable at all, depending on the similarity of the transistors.

Transistor Power Amplifiers

A transistor power amplifier uses (typically) another configuration for the input stage. This is called a "long tailed pair", (LTP) and acts as both the input stage and error amplifier (Q1 and Q2). This

circuit operates in current mode, so there is little output voltage to be seen from its output.

The second stage (Q3) is a Class-A amplifier, and is responsible for a large proportion of the overall gain of the circuit. Notice the current sources that are typically used for the LTP and Class-A amp sections. These are commonly made using transistors and maintain a constant current regardless of the voltage at the collector. If the current were truly constant, this implies that the impedance is infinite (which means that the gain of the transistor stage is also infinite!), and although this is not the case in reality, it will still be remarkably high.

For more information on how current sources are constructed, see Section 5.1

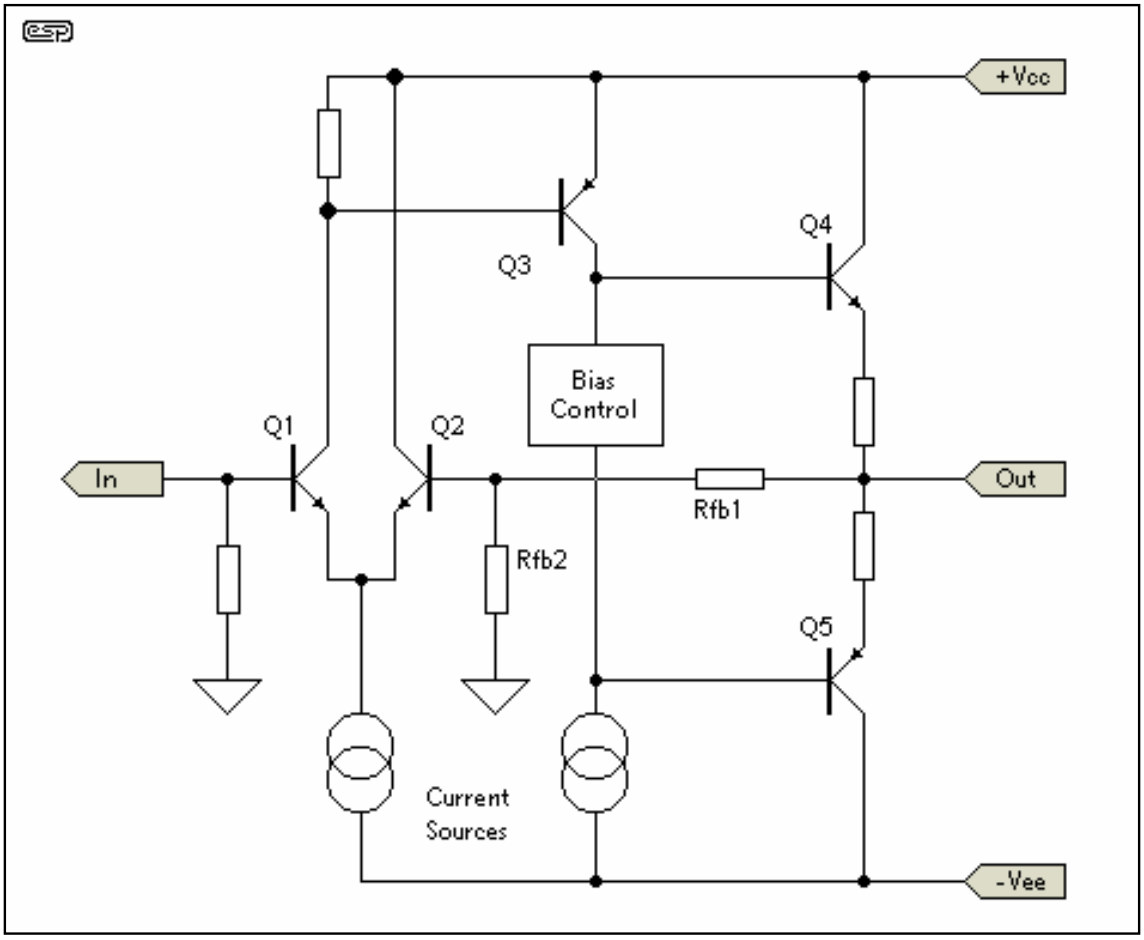


Figure 2.6 - Transistor Power Amplifier

The output stage (Q4 and Q5) typically is a pair of complementary emitter followers, which must be correctly biased to ensure that as the signal passes from one transistor to another, there is no discontinuity. This form of operation is known as Class-AB, since the amp operates in Class-A for very low level signals, then changes to Class-B at higher levels. Any discontinuity while passing the signal from one transistor to the other is the cause of crossover distortion, and for many years gave transistor amplifiers a bad name in the audio world. With proper biasing, and properly applied feedback, the crossover distortion can be made to go away - although never completely, but amplifiers with distortion levels of well below 0.01% are common.

The resistors at the emitters of the output transistors help to maintain a stable bias, and also

introduce some local feedback to linearise the output stage. This is a simplified circuit, and in reality the output stage will usually consist of multiple transistors, commonly a driver transistor followed by the output transistor itself. This does not change the operation of the circuit, but simply gives the output stage more gain, so it does not load the Class-A driver too heavily (this will result in greatly increased distortion).

Like the previous example, the gain is entirely dependent on the ratio of Rfb1 and Rfb2. As shown, the amp in Figure 2.6 is DC coupled, meaning that it will amplify any voltage from DC up to its maximum bandwidth. Not shown on this circuit are the various components needed to stabilise the circuit to prevent oscillation at high frequencies - often in the MHz range. Such oscillation is a disaster for the sound, and will quickly overheat and destroy the output transistors.

There are also transistor amplifiers that operate in Class-A, which means that the output transistors conduct all the time, and are never turned off. This can produce distortion levels that are almost impossible to measure, but this is at the expense of efficiency, and Class-A amplifiers will get very hot while doing nothing. Unlike the more common Class-AB amplifier, they will actually get slightly cooler as they reproduce a signal, since some of the input power is then diverted to the loudspeaker.

Transistors - A Summary

Just as with valve amplifiers, I have only scratched the surface. Entire books are written on the subject, and range from basic texts used in technical schools, to very advanced tomes intended for university students. Since transistors are easy to work with (and safe), there is much to be gained by experimentation, and you will have the satisfaction of having designed and built a functioning amplifier.

Transistors also have their fair share of problems, and there are some things that they are just not very good at. Some of the major failings include:

- **Low Impedance** - Bipolar transistors are inherently low impedance, and additional circuitry is needed to make them work satisfactorily in high impedance circuitry. Noise is also a problem when high impedance sources are used.
- **Heat** - Transistors dislike heat, and if it is not removed, they will destroy themselves. Most transistors can operate with junction temperatures up to about 125 degrees C, but at that temperature, can do no work at all. The life of a transistor is severely shortened by operating at high temperatures.
- **Thermal Stability** - Transistors are subject to some major changes in operation, depending on their temperature. This can make the design of high quality amplifiers difficult, because the transistor has a tendency towards "thermal runaway". This means that as the device gets hotter, it will draw more current, which makes it get hotter still. This continues until the maximum operating temperature is exceeded, and the transistor(s) fail.
- **Second Breakdown** - This is a version of thermal runaway, but at a molecular level. Parts of the internal structure become hotter than others, causing the hottest part to do the most

work. This makes it hotter still until the transistor fails. Second breakdown is the most common cause of output transistor failure in power amps. It also happens very fast, and without warning - transistors can fail from second breakdown even when at ambient temperature.

- **Short-term Overload** - Largely due to second breakdown effects, transistors do not tolerate short term overloads, and in many cases even a momentary short circuit will cause instantaneous failure. Compared to valves, transistor circuits are much less capable with difficult loads, and usually must be over-engineered to sometimes extreme levels to prevent failures.
- **Hard Overload** - when a transistor amp goes into overload, it does so with startling clarity. The harmonics produced are such that tweeters can be destroyed very easily, and the sound is altogether unpleasant

Again, there are many advantages as well. Transistor amplifiers are very reliable, and can be counted on to give many years of life without requiring even a basic service (most of the time anyway).

They are also very quiet (generally much quieter than valve amps) and do not suffer from microphony, so room vibrations are not re-introduced into the music. Efficiency is much higher, with lower voltages and no heaters (its a pity they don't look really nice, though).

Output impedances of 0.01 Ohm are achievable, so loudspeaker damping can be very high. Because transistor amps are very mechanically rugged, they can be installed in speaker boxes, so speaker lead lengths can be very short.

Typical transistor amplifiers have much wider bandwidth than valve amps, because there is no transformer, this is especially noticeable at the lowest frequencies - a transistor amp can reproduce 5Hz as easily as 500Hz.

Previous Next

Articles Index

Main Index

Amp Basics Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999-2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Elliott Sound Products **Amplifier Basics - How Amps Work (Part 3)**

Copyright (c) 1999 - Rod Elliott (ESP)
Page Last Updated 06 Apr 2005

Articles Index

Main Index

Amp Basics Index

Contents

- Part 3 - Field Effect Transistors and MOSFETs
  - FET Characteristics
  - Junction FETs
  - MOSFETs
- FET Current Amplifier
- FET Power Amplifiers
- Summary
  - Junction FETs
  - MOSFETs

Part 3 - Field Effect Transistors and MOSFETs

Now on to FETs and MOSFETs. FET stands for Field Effect Transistor, and MOSFET means Metal Oxide Semiconductor Field Effect Transistor. This topic is something of a can of worms, not because of some deficiency in the devices, but because of the huge array of different types. The basic FET types are

- N-Channel Junction FETs
- P-Channel Junction FETs
- N-Channel Enhancement Mode MOSFETs
- P-Channel Enhancement Mode MOSFETs
- N-Channel Depletion Mode MOSFETs
- P-Channel Depletion Mode MOSFETs

There are a couple of major sub-classes of MOSFET - lateral and vertical. Lateral MOSFETs are particularly suited to audio applications, as they are far more linear than their vertical brethren, although their gain is generally lower. Vertical MOSFETs are ideally suited to switching applications, and this includes Pulse Width Modulated (PWM) amplifiers.

The terms 'lateral' and 'vertical' refer to internal fabrication methods, so many others you may come across (such as HEXFETs ®) are essentially variations of the vertical process. This is still not all the possibilities, because there are additional sub-classes as well, particularly with switching MOSFETs. However, for the purpose of a general article on their characteristics and how they work, I will concentrate on the most commonly used versions. This narrows the field, and we are left with both polarities of junction FETs, and both polarities of enhancement mode MOSFETs. With these, we cover the major proportion of current designs, so even 'though I will be leaving out a lot, the stuff I leave out is not all that common (he says hopefully).

FETs are "unipolar" devices, in that they use only one polarity of carrier, in contrast to bipolar transistors, which use both majority and minority charge carriers (electrons or "holes", depending on the polarity). FETs are far more resistant to the effects of temperature, X-Rays and cosmic radiation - any of these can cause the production of minority carriers in bipolar transistors).

I shall concentrate only on three terminal FETS, and the terminals are

- Source - The electron "source" (for N-channel devices), and is the equivalent of the cathode of a valve or the emitter of a transistor
- Gate - Control terminal - (more or less) equivalent to the grid of a valve or base of a transistor
- Drain - The terminal from which current is "drained" - equivalent to the plate of a valve or collector of a transistor

There is no simple equivalent circuit for FETs (as there is for transistors), but this is of no consequence. The gate is the controlling element, and affects the electron flow not by amplifying a current (as in the transistor), but by the application of a voltage. The input impedance of junction FETs is very high at all usable frequencies, but MOSFETs are different. They have an almost infinite input resistance, but appreciable capacitance between the gate and the rest of the device. This can make MOSFETs hard to drive, because the capacitive loading makes most amplifier devices unhappy.

The junction FET is common in the inputs of high performance opamps, and offers extremely high input impedance. Indeed this is the case for discrete FETs as well, and a simple voltage amplifier using a junction FET and a power MOSFET are both shown in Figure 3.1. Both devices are N-Channel, and note that the arrow points in a different direction for each. The arrows point in the opposite direction for a P-Channel device, and all polarities are reversed. Vdd is +20V.

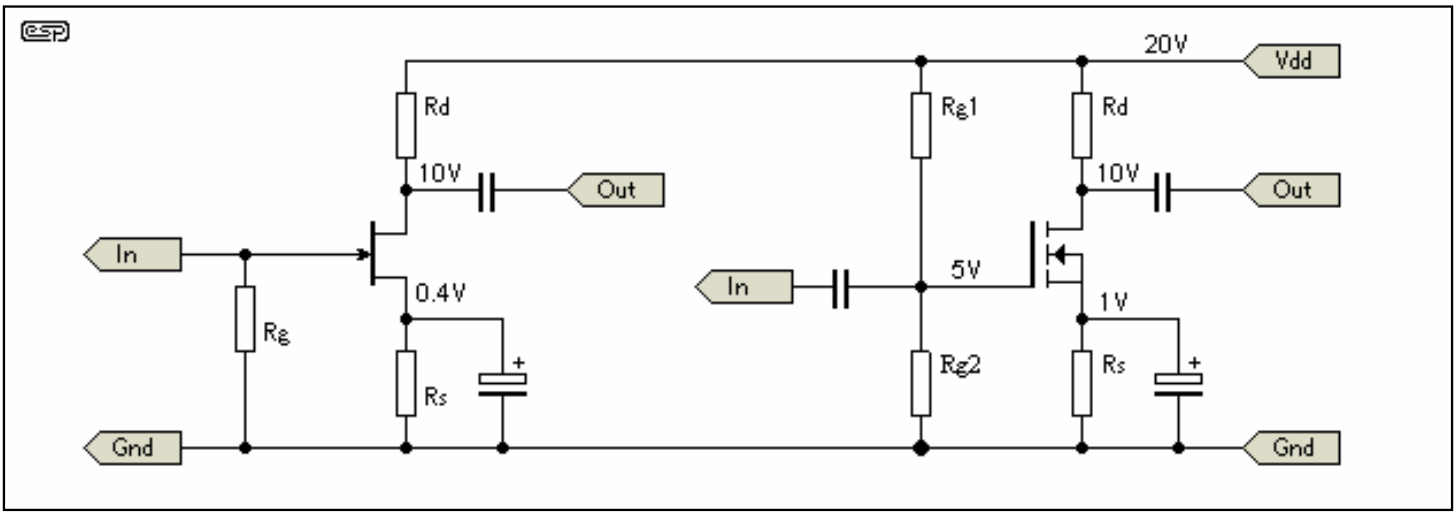


Figure 3.1 - Junction FET and Power MOSFET Voltage Amplifiers

Junction FETs are depletion mode devices, and (like all depletion mode FETs and MOSFETs) can be biased in exactly the same way as a valve. Depletion mode means that without a negative bias signal on the controlling element (the gate), there will be current flow between the drain (equivalent to plate or collector) and source (equivalent to cathode or emitter).

An enhancement mode device remains turned off until a threshold voltage is reached, after which the device conducts, passing more current as the voltage increases. Although there are MOSFETs made for low power operation, the majority (in audio, anyway) are power devices. These are almost exclusively enhancement mode, and can be capable of very high current.

In Figure 3.1, the power MOSFET is an enhancement mode device, and the junction FET is depletion mode. These are the most commonly used in audio. Enhancement mode power MOSFETs are also used in switching power supplies, and are far better than bipolar transistors in this role. They are faster, so switching losses are not as great (therefore the MOSFETs run cooler), and they are more rugged, and able to withstand abuses that would kill a bipolar transistor almost instantly.

This ruggedness (coupled with the complete freedom from second breakdown effects), means that MOSFETs are very popular as output devices for high power professional amplifiers. In this area, the MOSFET is second to none, and they are firmly entrenched as the device of choice for high power.

This is not to say that this is the only place MOSFETs are used. There are many fine audiophile power amps (and even preamps) that use power MOSFETs, and there are many claims that they are sonically superior to bipolar transistors (again, a debate that I will not discuss here).

Somewhat like valves, FETs and MOSFETs are very device dependent, and it is not normally possible to just substitute one device for a different type. Also like valves, the gain that can be expected from a voltage amplifier circuit is device dependent, and the manufacturer's data sheet (or testing) is the only way that one can be sure of obtaining the gain required in a given circuit.

FET Characteristics

The characteristics of FETs must be covered in two parts, since we are dealing with two quite different devices. The first will be the junction FET, and as with transistors, I shall only describe the N-Channel, but virtually identical P-Channel devices are available (although not as commonly used).

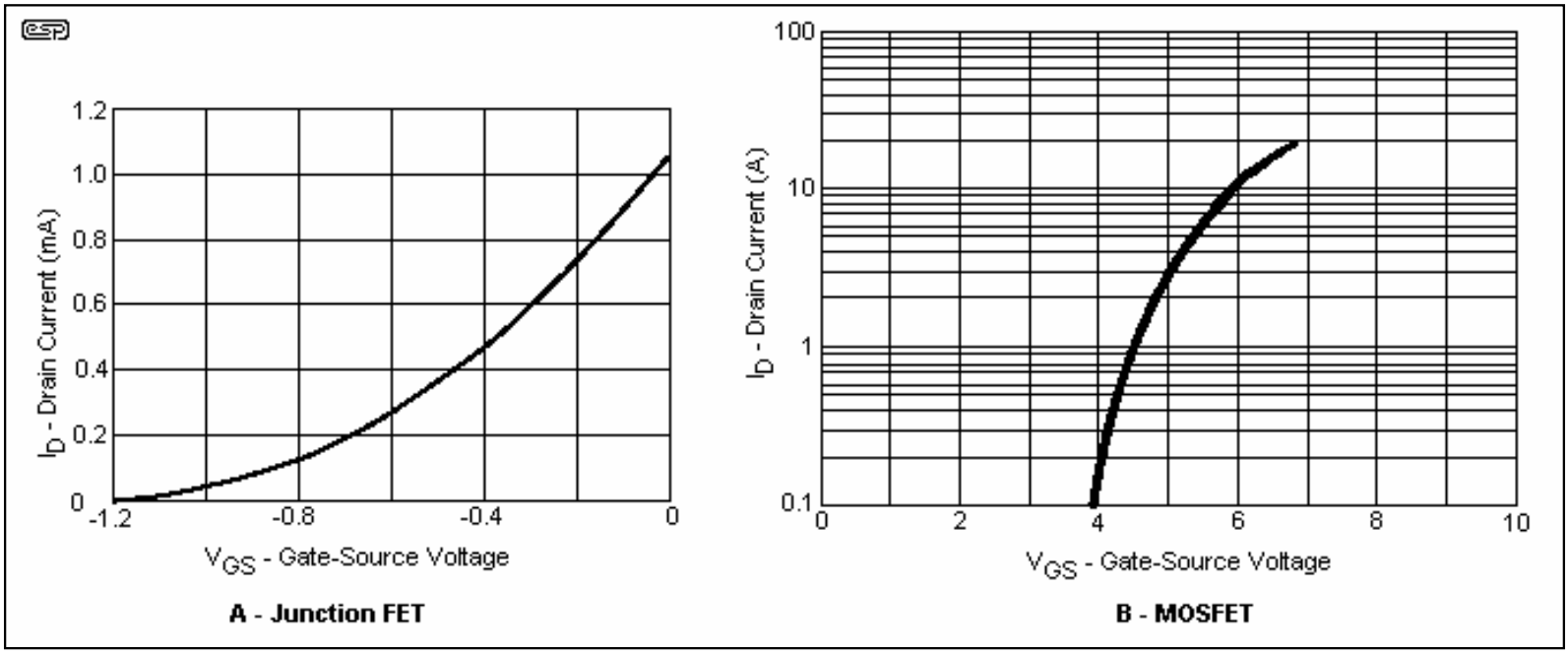
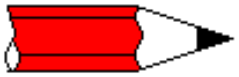


Figure 3.2 - Transfer Curves For a Junction FET and MOSFET

Initially, so the transfer characteristics of the two devices can be seen side by side for comparison, Figure 3.2 shows a fairly typical device from each "family" The junction FET data is from a 2N5457, and the MOSFET is an IRFP240 (a vertical MOSFET - more suited to switching applications).

Rather than show the input and output signals superimposed on a graph, this time I show only the graph itself. These are excerpts from manufacturers data, but with a small catch - Figure 3.2b has the drain current displayed on a logarithmic scale, so the linearity of the device cannot be seen properly. If this graph were redrawn as linear, it will show that linearity is best at higher currents (on the graph shown it looks the other way around), and the device becomes almost perfectly linear with drain currents above about 3A.

Note that because the junction FET is depletion mode, drain current is highest at 0V gate-source voltage. The MOSFET on the other hand is enhancement mode, so at 0V gate-source, there is no current. Conduction starts at 4V, and by 6V the drain current is 10A.



The term Siemens (S) is now replacing Mhos as the unit of transconductance in some literature: 1S = Mho (1uS=1 uMho. For the above graphs, it may be worked out that the junction FET has a transconductance of 1500uS, and for the MOSFET it is approx. 9,000uS (9,000 uMhos)

Junction FETs

Like valves, FET data sheets provide gain information as gm (mutual conductance - in uMhos). The junction FET shown has a gm of (typically) 1,500 uMhos (in the graph shown it is actually closer to 1425 uMhos in the linear section), which translates to about 1.5 mA/V.

The most common of the quoted parameters for junction FETs are

- Forward Transfer Admittance (Common Source) - Transconductance - essentially the gain of the device
- Input Capacitance - The effective capacitance of the gate terminal to the remainder of the FET
- Gate–Source Cutoff Voltage - The gate voltage at which the FET is turned off

The process of amplification is almost identical to that of a valve, except that the voltages are lower. The device is biased in the same way (although fixed bias can also be used). This means that the gate must be reverse biased with respect to the source, with the gate having the opposite polarity of the source-drain voltage.

FETs offer low noise, especially with high impedance inputs, and in this respect are the opposite of bipolar transistors, which are generally at their best with low source impedance.

Junction FETs are predominantly low power, although there are some high power devices available. These are uncommon in audio applications.

MOSFETs
Again, MOSFET data sheets also provide information similar to junction FETs, but there are more items of importance to the designer. The most useful of these are

- Forward Transconductance - The device gain characteristic
- Drain to Source On Resistance - The minimum resistance when the MOSFET is fully on
- Gate Threshold Voltage - The gate voltage at which the MOSFET will start conduction
- Gate to Source Voltage - The maximum voltage (of either polarity) that may be applied between source and gate. (This is typically in the order of +/-20V)
- Input Capacitance - The value of capacitive loading that is placed on the driving circuit

Enhancement mode MOSFETs pass virtually no current when there is no gate voltage present. To conduct, a voltage must be applied between source and gate (of the same polarity as the drain voltage). Once the threshold has been reached, the device will start to conduct between drain and source.

At increasing gate voltages, the drain current increases until either a) the maximum permissible drain current or total dissipation limit is reached, or b) the drain voltage falls to its lowest possible value. In this instance, since the source-drain channel is now fully conducting, the value of $R_{DS(on)}$ determines the voltage.

Typical power MOSFETs offer extremely low on resistance, with values of 0.2 Ohm being fairly typical. There are some devices with much lower values, but this is only important in switching circuits. In an audio amp, the MOSFETs should never be turned completely on, since this means the amplifier is clipping.

Another area that must be addressed with MOSFETs is the voltage between gate and source. Because the gate is insulated from the channel by a thin layer of metal oxide, it is susceptible to damage by

static discharge or other excessive voltage. It is common to include a zener diode between source and gate to ensure that the maximum voltage cannot be exceeded. Voltage spikes in excess of the breakdown voltage of the insulating layer will cause instantaneous failure of the device.

FET Current Amplifier

Again, I have shown both a junction FET and a MOSFET in Figure 3.3, both common-drain or source-follower circuits. As can be seen, the junction FET is biased almost identically to a valve, but all voltages are much lower. The MOSFET requires a positive voltage, and this must be greater than the source voltage, by an amount that takes the characteristics of the MOSFET into consideration. For the device characteristics shown in Figure 3.2 this means that at a current of 100mA, the gate must be 4V higher than the source.

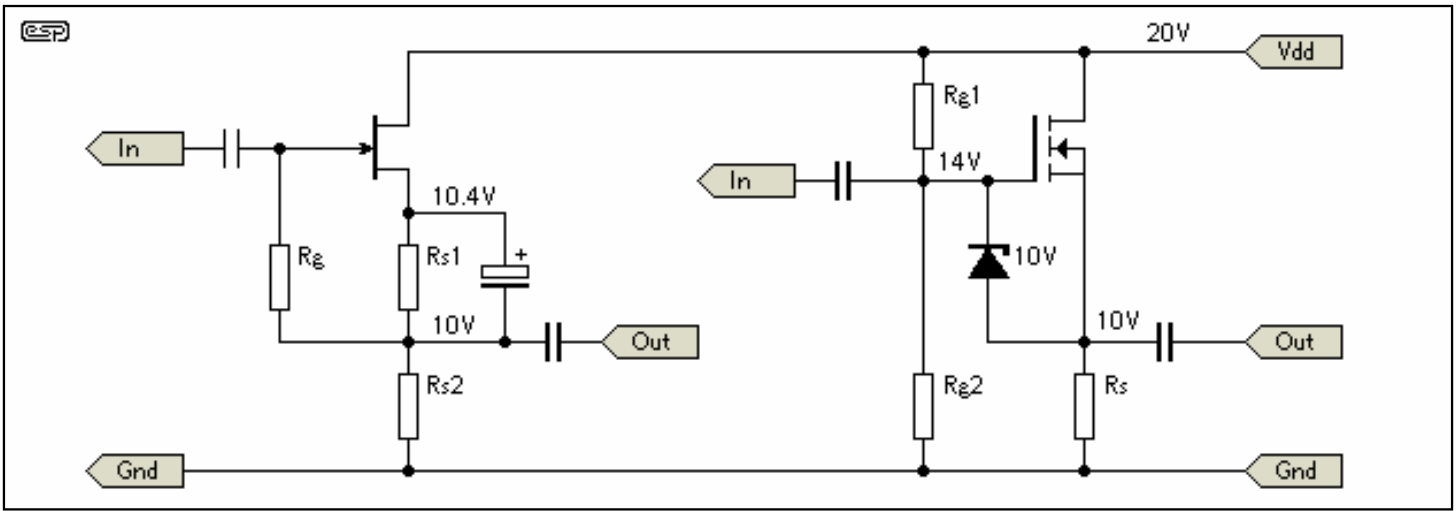


Figure 3.3 - FET Current Amplifiers

Included in the MOSFET version is a zener for protection of the gate insulation. A 10V zener is used, as this gives good protection and is still able to let the maximum possible MOSFET current flow. A 6V zener could have been used, and this would still allow current up to 10A, which is far more than can be achieved from this simple circuit.

FET Power Amplifiers

In exactly the same way as a power valve can be used in single-ended Class-A, so too can a MOSFET. A simple circuit is shown in Figure 3.4 which will provide about 10W of audio. Using a constant current source as a load (as shown) gives much better efficiency than a resistor, and improves linearity. The distortion from a circuit such as that shown will be roughly the same as that from a single ended triode valve circuit. Overall efficiency will be higher, since there is no cathode bias resistor needed, and no heaters as with a valve.

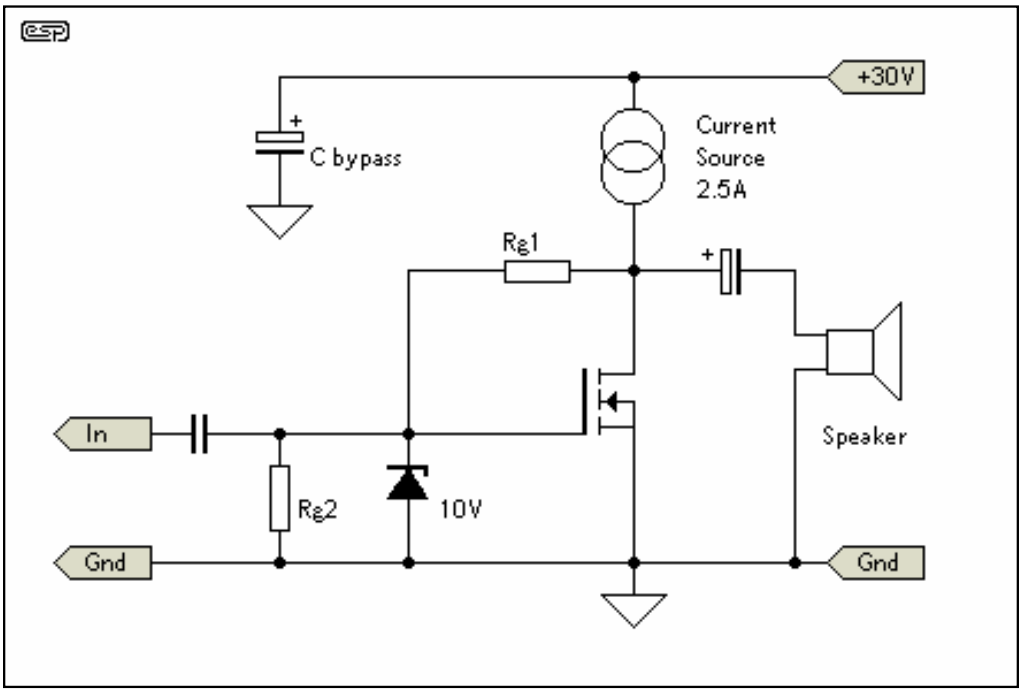


Figure 3.4 - Single Ended MOSFET Class-A Amplifier

Although there are a few, all MOSFET power amplifiers are uncommon. Most use a combination of bipolar transistors (for the input and gain stages), and MOSFETs for the output devices. This seems to be the most popular circuit arrangement, so I will concentrate on this. Figure 3.5 shows a fairly typical arrangement (in simplified form), and the operation of this is almost identical to that of an amplifier using bipolar transistors in the output. Note that emitter followers are needed to be able to provide the low impedance drive that MOSFETs need, although in some circuits they are not used. Instead, the Class-A driver stage (Q3) is operated at a higher than normal current to allow it to drive the MOSFETs properly.

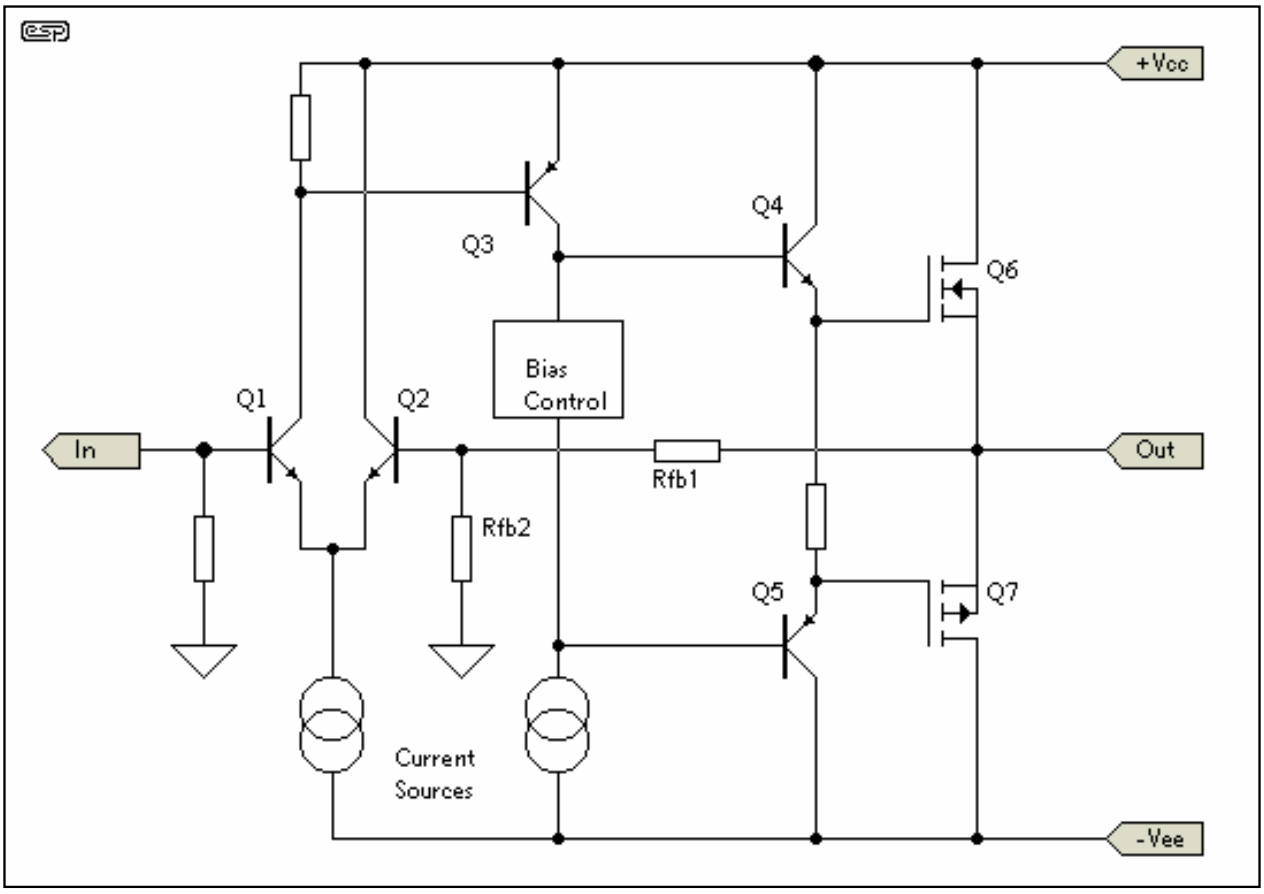


Figure 3.5 - MOSFET Output Power Amplifier

One problem with this arrangement is that the drain to source voltage represents a circuit loss, so the power supply voltage needs to be typically +/-6V higher than the required peak output voltage to the load. Although this is not a major problem, it does increase dissipation in the output stage, and the loss increases with lower impedance loads.

Some (especially very high power) amps get around this by using a low current (but higher voltage) secondary power supply for the drive circuit, and the main high current supply for the MOSFETs. In an amp using +/-50V at 20 Amp main supplies, the secondary supply might be +/-60V, but capable of perhaps 1A maximum.

As with the bipolar amp (did you notice how similar they are?), I have not included components for stability. These are typically the same as for a standard bipolar transistor amp, but will usually include "stopper" resistors in series with the gates of the MOSFETs, and sometimes additional capacitance to prevent parasitic oscillation - the need for these varies from one device type to the next.

FETs - A Summary

Junction FETs

The surface is again, only barely scratched. The junction FET (A.K.A. JFET) is ideally suited to circuits where high impedances are expected, and will give the lowest noise. They are an invaluable electronic building block when used where they excel - providing extremely high input impedance.

Like all devices so far, JFETs have their limitations ...

- **Gain** - JFETs do not have the high gain of bipolar transistors
- **High Frequency Response** - Generally, JFETs have a high frequency performance that is not as good as bipolar transistors
- **Linearity** - The linearity of JFETs is not as good as bipolar transistors (so distortion is greater), but can be improved by using current source loading or feedback.

There is generally an ideal (or close to ideal) amplifying device for every application, and when used properly, the JFET is extremely versatile and at its best when high impedances are needed. If you have a need to send an amplifier into space, then JFETs are preferred due to their greater "radiation hardness".

JFETs (in fact all FETs) are more sensible than bipolar transistors when heated, and problems of thermal runaway are not encountered with these devices.

MOSFETs
The MOSFET is one of the most powerful of all the current range of amplifying device, with extraordinary current capability. Ideally suited to very high power amplifiers, where extremes of operating conditions are regularly encountered, the MOSFET has no equal.

... And, as always, there are limitations ...

- **Gain** - Like JFETs, MOSFETs have a lower gain than bipolar transistors, which usually means that additional gain must be applied to the driving circuit to ensure that the global feedback is sufficient to maintain low distortion at low levels.
- **Gate Capacitance** - The capacitance of the gate to source can be as high as 2nF (although more typically around 1.2nF). This is not a lot at low frequencies, but makes the drive circuit work very hard at high frequencies.
- **Static Damage** - Until installed in a circuit with full protection, the MOSFET is susceptible to damage from static discharge. The voltage and current needed to destroy the device are generally below the threshold of feeling for humans. Some devices have (limited) protection built in.
- **Linearity** - Most MOSFETs are not very linear at low currents, so for low distortion higher quiescent is needed to ensure that crossover distortion is minimised.

To some extent, all the above can be forgiven when you really need the capabilities of a MOSFET. The complete freedom from second breakdown and the massive current capabilities of MOSFETs are unmatched by any other active device. With a properly designed drive circuit, MOSFETs are also very fast, capable of performance that is generally superior to that of bipolar transistors. This is not very helpful in audio, but is essential for switching circuits.

Coupled with a positive temperature coefficient that completely stops thermal runaway in a linear circuit, the MOSFET is almost indestructible, provided that precautions are taken to ensure the gate voltage is kept below the breakdown voltage.

The positive temperature coefficient is a great help in audio circuits, although it can be a problem in switching power supplies, since the "on" resistance also increases with temperature, and in a switch-mode power supply this can cause thermal runaway (exactly the reverse of bipolar transistors in this application).

Previous Next

Articles Index

Main Index

Amp Basics Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999-2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Elliott Sound Products **Amplifier Basics - How Amps Work (Part 4)**

Copyright (c) 1999 - Rod Elliott (ESP)
Page Last Updated 06 Apr 2005

Articles Index

Main Index

Amp Basics Index

Part 4 - Operational Amplifiers (Opamps)

No discussion of amplifying devices would be complete without a discussion of opamps. Although not a single device, the opamp is considered to be a building block, just like a valve or any transistor.

The format I used for the other discussions is not appropriate for this topic, so will be changed to suit this most versatile of components. I shall not be covering esoteric or special purpose types, only the basic variety, as there are too many variations to cover.

The operational amplifier was originally used for analogue computers, although at that time they were made using discrete components. Modern (good) opamps are so good, that it is difficult or impossible to achieve results even close with discrete transistors or FETs. However, there are still some instances where opamps are just not suitable, such as when high supply voltages are needed for large voltage swings.

The majority of power amplifiers (whether bipolar or MOSFET) are in fact discrete opamps, with a +ve input and a -ve input. You tend not to see this, but have a look at Figure 3.5 again. The signal is applied to the +ve input at the base of Q1. The base of Q2 is the -ve input, and is used for the feedback signal, exactly the same as you will see in Figure 4.1a below.

Unlike the other devices, opamps are primarily designed as voltage amplifiers, and their versatility comes from their input circuitry. Opamps have two inputs, designated as the non-inverting and inverting (or simply + and -).

When wired into a conventional amplifier circuit, the opamp has one major goal in its little life ...

Make both inputs the same voltage

If, because some swine of a designer has made this impossible (very common with a lot of circuits), the opamp then takes another approach ...

Make the output the same polarity as the most positive input

The latter condition needs a small explanation. If the +ve input is most positive, then the output will swing to the positive supply rail (or as close as it can get). Should the -ve input be more positive, then the output will swing to the negative supply rail. The difference between the two inputs may be less than 1mV! Simple as that.

I call these "The First and Second Laws of Opamps". These two statements describe everything an opamp does, and just by knowing this, makes the task of working out what most common circuits do a simple process. There is actually nothing especially complex about opamps, unless you look at the "simplified" circuit diagram often included in data sheets. Don't do this, as it is too depressing. (By the way, the first statement is not strictly true of real-life devices, which will always have some error, however without very specialised equipment you will be unable to measure it.)

Modern opamps (the good ones, anyway) are as close as anyone has ever got to the ideal amplifier. The bandwidth is very wide indeed, with very low distortion (0.00003% for one of the Burr Brown devices), and low noise. Although it is quite possible to obtain an output impedance of far less than 10 Ohms, the current output is usually limited to about +/-20mA or so. Supply voltage of most opamps is limited to a maximum of about +/-18V, although there are some that will take more, and others less.

Depending on the opamp used, gains of 100 with a frequency response up to 100kHz are easily achieved, with noise levels being only very marginally worse that a dedicated discrete design using all the noise reducing tricks known. The circuits shown below have frequency response down to DC, with the upper frequency limit determined by device type and gain.

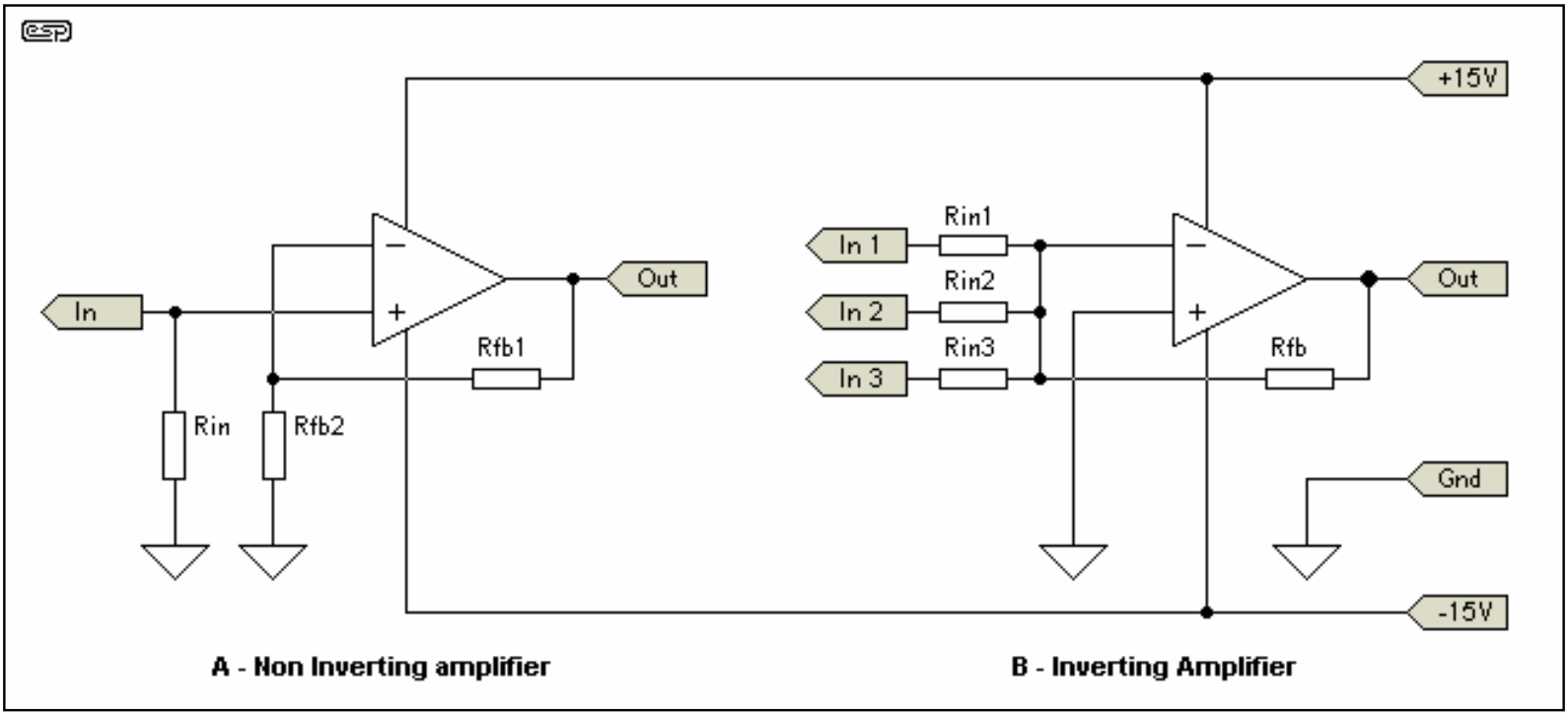


Figure 4.1 - Standard Opamp Configurations

Figure 4.1 shows the two most common opamp amplifier circuits. The first (4.1a) is non-inverting, and is the better connection for minimum noise. The voltage fed back through Rfb1 will cause a voltage to be developed across Rfb2. The output will correct itself until these two voltages are equal at any instant in time. It does not matter if the signal is a sinewave, square wave, or music, the opamp will keep up (provided you stay within its capabilities). Once the speed of the opamp is not significantly higher than the rate of change of the input (generally a factor of 10 is sufficient - i. e. the opamp needs to be 10 times faster than the highest frequency signal it is expected to amplify), the output will become distorted. At voltage gains of 10 or less, almost any opamp will be able to keep up with typical audio signals, but (and be warned) this is no guarantee that they will sound any good.

Input impedance is equal to Rin, and voltage gain (Av) is calculated from

$$Av = (Rfb1 + Rfb2) / Rfb2$$

The second circuit (4.1b) is an inverting amplifier, and is commonly used as a "summing" amplifier - the output is the negative sum of the three (or more) inputs. It is also called a "virtual earth" mixer, because the -ve input is a virtual earth (remember my "First law of opamps"). If the +ve input is earthed (grounded), then the opamp must try to keep the -ve input at the same voltage - namely 0V.

It does this by adjusting its output until the current flowing through Rfb is exactly the same (but of the opposite polarity) as the current flowing into the inputs from each Rin. They must all sum to 0V, as they are equal and opposite. This is done with amazing speed, and good opamps will continue to succeed in fulfilling the First Law up to over 100kHz or more (depending on gain). Lesser devices will start to have trouble, and the appearance of a measurable voltage at the -ve

input is an indication that the opamp can no longer keep up with the signal.

Input impedance is equal to RinX (where X is the number of the input), and voltage gain is calculated from

$$Av = Rfb / RinX$$

Multiple inputs can all have different gains (and input impedances). There are two catches to this circuit. The first is that if the source does not have an output impedance significantly lower than Rin, then the gain will be lower than expected. The other, not always realised, is that if the circuit is configured for a gain of 1 (actually it is technically correct to refer to it as -1), Rin1, Rin2 etc. will all be equal to Rfb. If the circuit has 10 inputs, then from the opamp's perspective it has a gain of 10, and its frequency response and noise will reflect this.

There are literally hundreds of different opamp circuit configurations. Feedback circuits with frequency dependent components (capacitors or inductors) make the opamp into a filter, or a phono equaliser, or almost anything else.

Power Opamps

Opamps even come in power versions, using a TO-220 (or other specialised) case, and are typically capable of around 25W into an 8 Ohm speaker load. These devices, while not exactly to audiophile standards, are still very capable, and have been used by many domestic appliance manufacturers in such things as high-end TV sets and standard hi-fi equipment. Some of the more advanced devices are capable of output power up to 80W.

They typically have distortion figures below 0.1%, and can be used anywhere a small, convenient and cheap power amp is required. The circuit looks almost identical to that of a small signal opamp, except that a Zobel stabilisation network is used on the output to prevent oscillation.

Previous Next

Articles Index

Main Index

Amp Basics Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999-2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Elliott Sound Products **Amplifier Basics - How Amps Work (Part 5)**

Copyright (c) 1999 - Rod Elliott (ESP)
Page Last Updated 06 Apr 2005

Articles Index

Main Index

Amp Basics Index

Contents

- Part 5 - Some Basic Linear Circuit Building Blocks
  - Current Sources and Sinks
  - Current Mirror
  - Long Tailed Pair
  - Grounded Grid (Gate or Base) Circuits
  - Cascode

Part 5 - Some Basic Linear Circuit Building Blocks

There are some circuits in the world of electronics that are just too useful. While some have been around for many years, others only became practical with the advent of the transistor. The circuits described are a mixture, some are very old, and others much newer.

I shall not go into the history (this is an electronics tutorial, not a history lesson), but will show the various stages in their basic form for each type of circuit.

On the topic of current sources, sinks and mirrors, click here to see the full article describing how they are most commonly used in audio circuits (and why).

Current Sources And Sinks

The constant current source (or sink) is one of the most versatile and widely used of the circuits shown in this section. The ideal current source provides a current into a load that is independent of the

resistance (or impedance) of the load, from zero to infinity. As always, the ideal does not exist, but within the capabilities of the power supply voltage, it is quite simple to do, and surprisingly accurate.

There is no real difference between the two circuits - one sources current (or sinks electrons) or vice versa. Sometimes it might help to consider the circuit "upside down" to see that there is no real difference, only one of terminology.

As an example, if we wanted to supply a current that were fixed at 1A into any load impedance, then we might use a circuit similar to that in Figure 5.1 - a basic transistor current source. As shown, this will supply 1A into any resistance from zero Ohms up to a little under 50 Ohms. The power supply is the limiting factor - to be able to supply the same current into 1M Ohm would need a 1,000,000V power supply, which is an unrealistic expectation.

The current source or sink can be imagined as a device with infinite impedance - this must be the case if the current remains unchanged even as the load resistance is varied over a wide range. Naturally, the impedance of actual current sources is not infinite, but can easily reach values of many megohms, even in a simple circuit.

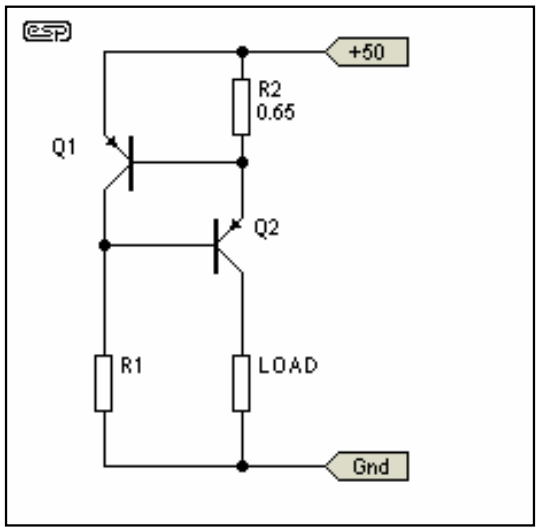


Figure 5.1 - A Basic Transistor Current Source / Sink

The operation of the circuit is simple. If the voltage across the emitter resistor of Q2 attempts to exceed 0.65V (the base turn-on voltage for a silicon transistor), then Q1 will turn on, and short out all base current to Q2 except for exactly that amount required to maintain the specified current of 1A. If the collector current of Q2 falls, then the voltage across the emitter resistor also falls. This turns off Q1 until the current is again stable at the preset value. (This is only one way to make a current source - there are many others.)

Thermal stability is not good. The emitter-base potential falls at 2mV / degree C, so as the temperature increases, the current will fall from the nominal 1A. At low temperatures, the opposite will occur. A precision voltage reference can be used, or an opamp can monitor the voltage across the resistor, resulting in a much more stable current. Fortunately, in most circuits, it is not that critical, so the circuit of Figure 5.1 is very common.

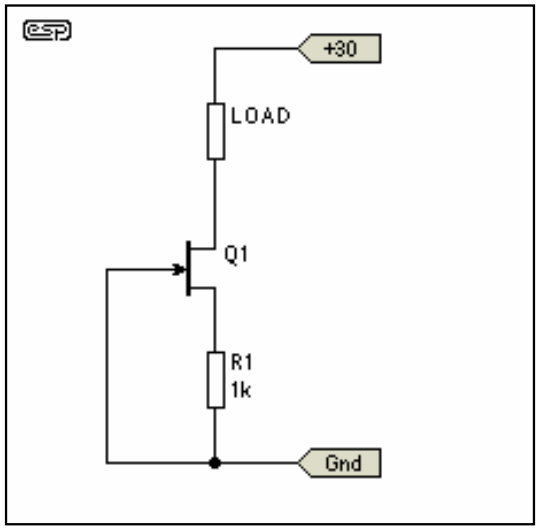


Figure 5.2 - A JFET Current Source / Sink

Junction FETs, being a depletion mode device, can be used as a current source very easily, as shown in Figure 5.2. Because JFETs are mainly low current devices, the useful range is from about 0.1mA up to 10mA or so. This is ideal for many of the circuits that need a current source. The actual current is dependent on the FET's characteristics, but is sufficiently stable for many non-critical applications. Based on the FET curve shown in Figure 3.2a, this current source will supply a current of about 0.4mA into a load from zero to 72k Ohms. The voltage is also lower, because of the lower voltage rating of most FETs.

Current Mirror

The current mirror is one of the "new" circuits, and works well with bipolar transistors. It is unusual to see this circuit implemented with valves or FETs, and I will not change this (i.e. I will show transistors only). Figure 5.3 shows a simple current mirror (this version is not very accurate, but is still extremely effective and commonly used).

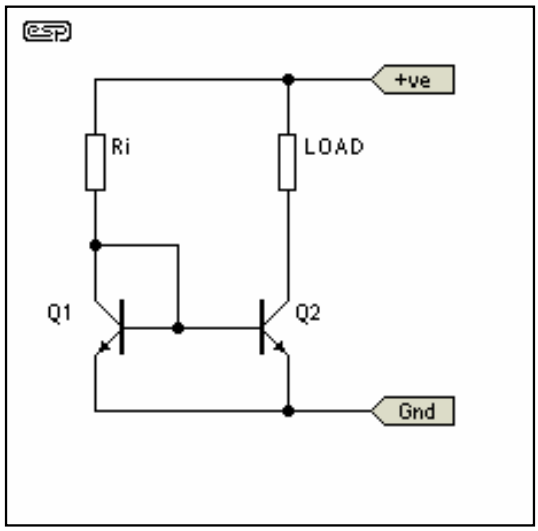


Figure 5.3 - A Simple Current Mirror

Any current injected into the collector/base circuit of Q1 (via Ri) will be "mirrored" by Q2, which will

draw the same current through its load resistor (within the capability of the transistor and power supply). Current mirrors are sometimes used as current sources (one less resistor), and are not as dependent on temperature, since both transistors will ideally be at the same temperature. It is not uncommon to use dual transistors (or thermal bonding) to ensure stability.

Long Tailed Pair

The long tailed (or differential) pair is an old circuit, and is used with valves, FETs and Bats. It was originally designed in the valve era, and provides a means for the comparison of two voltages. The long tailed pair (LTP) is used as the input stage of most opamps, and many (if not most) modern power amplifiers.

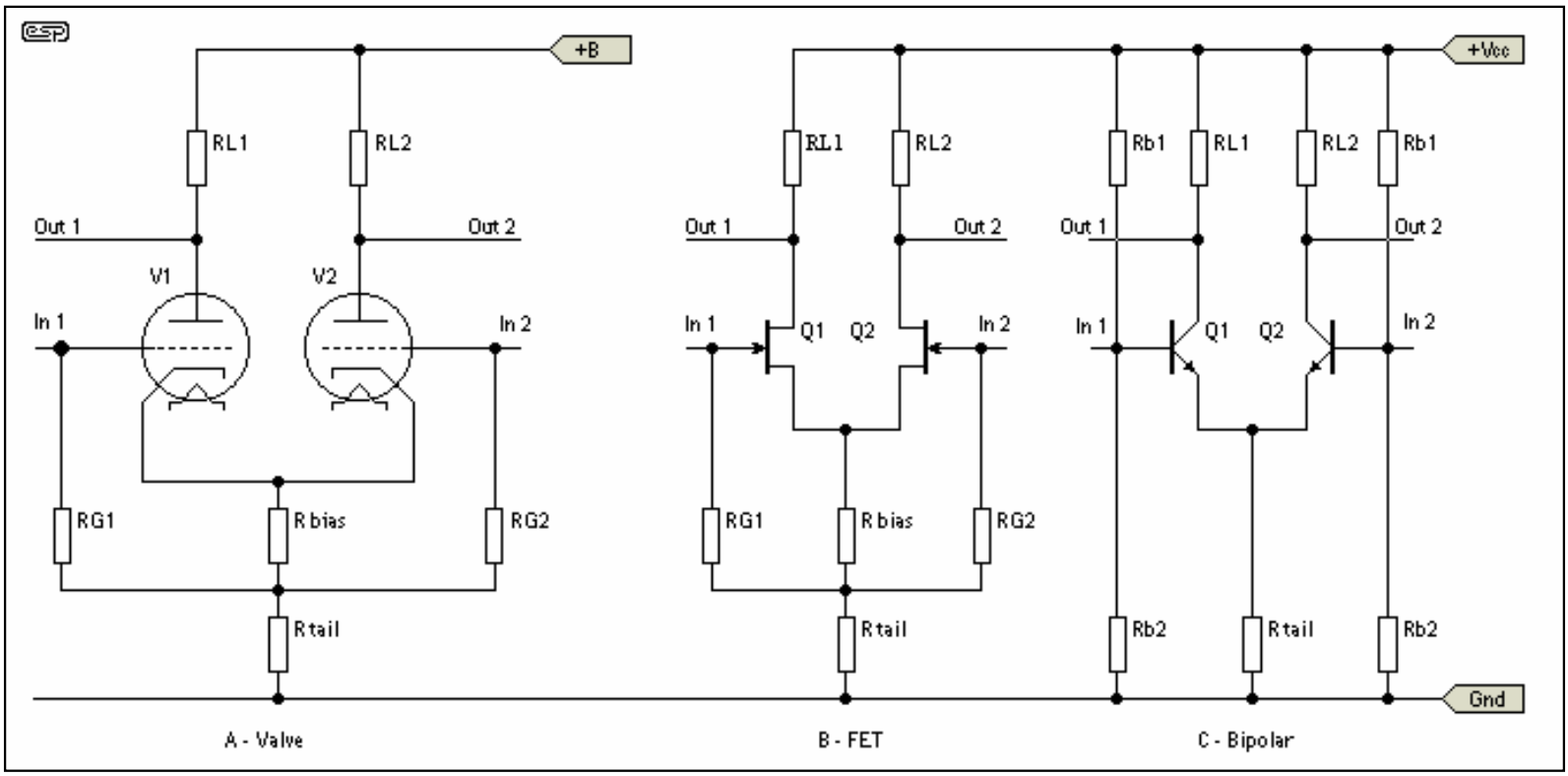


Figure 5.4 - The Long Tailed Pair - All Common Devices

As can be seen in Figure 5.4, the LTP can be made using valves (A), JFETs (B) or bipolar transistors (C). Valves and JFETs can be self biased as shown, but BJT circuits must have external bias resistors. A pair of MOSFETs could be used, but at the typical currents used (less than 5mA), the gain and linearity would be very poor. Although each circuit is shown using a resistor as the "tail", in FET and bipolar circuits this is most commonly a current source (or sink if you prefer).

The use of a current source stabilises the overall current, so the device input current is not affected by supply voltage changes, or variations in the input bias voltages.

In each case, the circuit has an inverting and a non-inverting input, and an inverted and non-inverted output. Application of the same voltage and polarity to both inputs at once results in (theoretically) zero output - this is called the common mode signal, and is commonly quoted for opamps as the common mode rejection ratio.

The valve and FET versions only require capacitive coupling, as they are self biasing as shown. The bipolar circuit cannot be self biased, and requires the biasing resistors Rb1 and Rb2 for each input.

The output of each version may be taken from either or both outputs, and may be capacitively or direct coupled. Direct coupling is very common with LTP circuits, especially in opamps and audio power amplifiers, where it is the rule, rather than the exception.

When used as the input stage of an amplifier, the LTP uses one input as the signal input, and the other is used for the application of feedback, in the same way as in an opamp.

Grounded Grid (Gate or Base) Circuits

Sometimes, it is desirable to have an extremely low input impedance, for example where the output impedance of the source is very low. One way to achieve this is to use the control element (grid, gate or base) as the reference, and apply the signal to the cathode, source or emitter (as appropriate). Figure 5.5 shows an example of grounded (or common) grid (A), gate (B) and emitter (C).

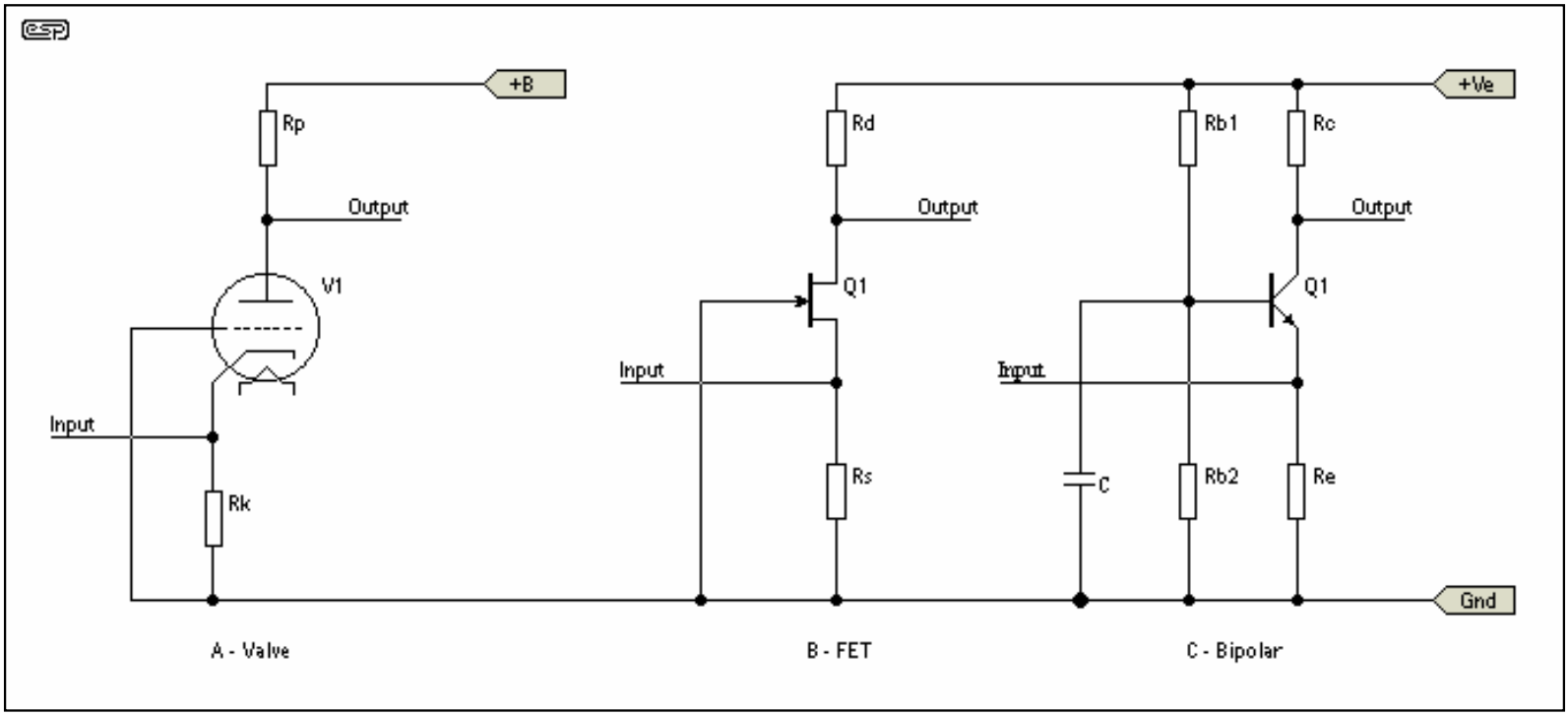


Figure 5.5 - Common Grid, Gate and Base Amplifiers

Apart from having an extremely low input impedance, this class of amplifier has an additional advantage. The normal capacitance from output to input is bypassed to earth, and no longer acts as a feedback path. Such circuits are therefore capable of a much better high frequency response than when used in the "conventional" way, and are common in radio frequency circuits. The bypass path is direct for a valve or FET, and is via a capacitance for the bipolar circuit.

All inputs and outputs must be capacitively coupled, unless the preceding circuit is to be direct coupled (unusual) or the output is direct coupled to a follower (quite common).

Because the input impedance is so low, there are few applications in audio, except where this circuit is used in conjunction with a "normal" amplifier stage. This forms a new circuit, called cascode.

Cascode

This circuit was developed in the valve era, primarily to obtain better response at high radio frequencies. Valves have capacitance between the plate and grid, and this acts as a feedback path at high frequencies, causing a drop in gain as the frequency is increased. This is the so-called "Miller" effect. Operating in cascode allows the circuit to have a high input impedance (via the normal grid input), and the grounded grid amplifier (the signal is applied to the cathode) means that there is no feedback from plate to grid, and the grid acts as a shield to prevent feedback to the cathode. The lower half of the stage contributes a relatively small amount of gain, and is not subject to the feedback effect since it is operating as a current amplifier (there is very little voltage swing on the plate, so there is little or no signal to feed back).

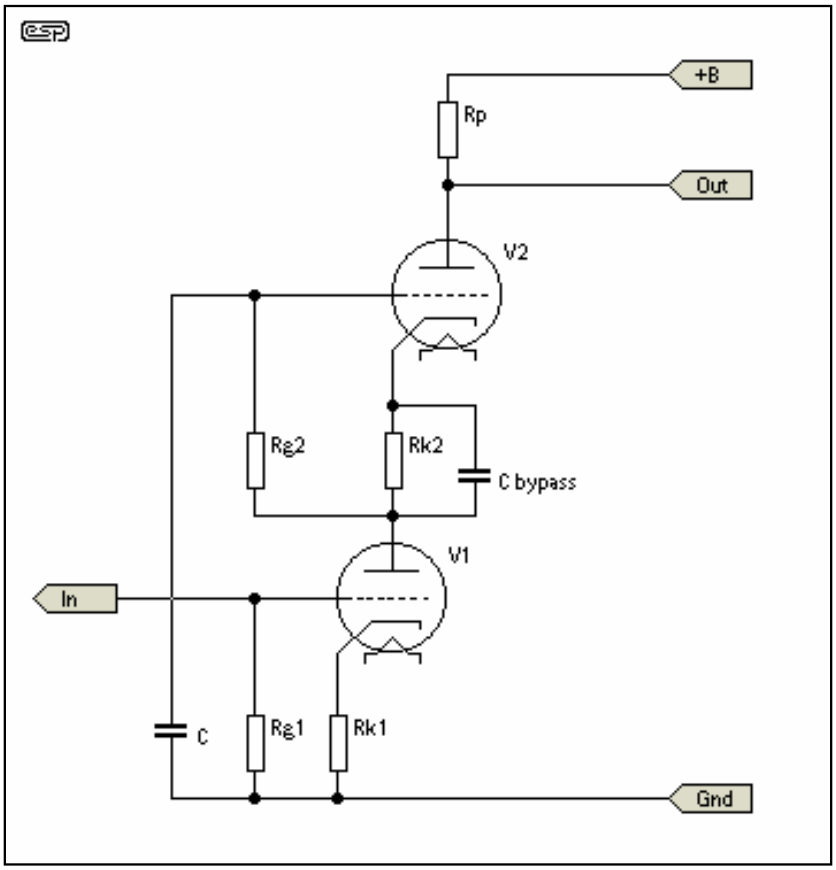


Figure 5.6 - Valve Cascode Amplifier

As can be seen, the grid of V2 is earthed via the capacitor for all signal frequencies, allowing V2 to operate as common grid. The capacitor (C bypass) is used to ensure that there is no gain lost due to cathode degeneration (local feedback). The cathode of V1 will also be bypassed in many cases, especially where low noise is a primary goal.

The same principles can be applied to FETs or Bats, and has similar advantages. The capacitance between Drain and gate (or collector and base) is isolated in the same way as with a valve circuit, with the signal being coupled to the source or emitter by the first device. Figure 5.7 shows a composite

JFET / BJT cascode circuit, which will have better linearity than a conventional amplifier, and a much better high frequency response.

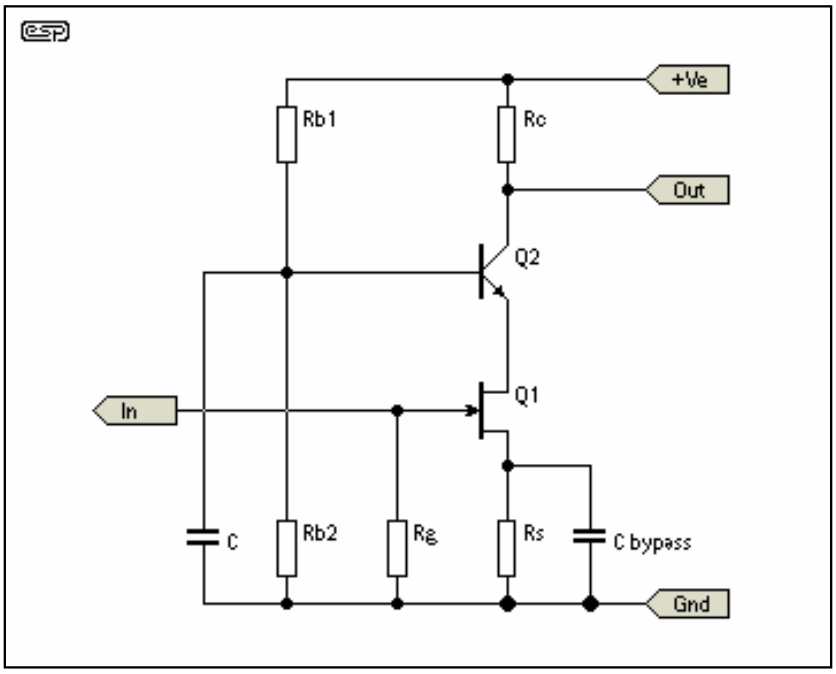


Figure 5.7 - Composite FET / BJT Cascode Amplifier

This type of circuit is not uncommon in high performance opamps, where very wide bandwidth and good linearity before feedback are required. Cascode circuits are also sometimes seen in solid state power amplifier circuits, where the designer is trying to obtain the maximum possible bandwidth from the amp.

The base of Q2 is grounded to all signal frequencies, so the stage operates as a common base circuit. Using a JFET as the input element means that the circuit has a high input impedance, while the BJT ensures maximum gain. To obtain even more gain, Rc might be replaced by a current source, in which case the gain from this single stage can exceed 1000 times, with wide bandwidth and excellent linearity.

Previous Next

Articles Index

Main Index

Amp Basics Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999-2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Elliott Sound Products **Amplifier Basics - How Amps Work (Part 6)**

Copyright (c) 1999 - Rod Elliott (ESP)
Page Last Updated 06 Apr 2005

Articles Index

Main Index

Amp Basics Index

Contents

- Conclusions
- References
- Copyright Info
- Update Info

Conclusion

Section 5 finalises this series - at least until such time as I find (or someone points out) a mistake or major omission that I will then have to fix, there will be no further updates.

The articles in this series describe the essential building blocks of nearly all circuits in common use today. There are others (of course) but they are most often combinations of the above - for example, a LTP stage can be built using two cascode circuits, a current source and a current mirror. The resulting circuit looks complex, but is simply a combination of common circuits such as those shown.

Other circuits are modification of the basic stages to exploit what might otherwise be

seen as a deficiency - for example circuits that deliberately exploit the temperature dependency of a BJT can be used as high gain thermal sensors, or to stabilise the quiescent current in a power amplifier.

There are also some bizarre combinations possible. A valve and BJT operating in cascode would be interesting, and would no doubt have some desirable characteristics (I haven't seen this particular combination, but there is no reason that it would not work). Likewise, a valve with a transistor current source instead of the load resistor would have far better linearity and more gain than a simple resistor loaded version.

In many cases, ICs are available to accomplish many of the functions described. Opamps are an obvious one, but there are also IC current sources, transistor arrays (ideal for current mirror applications because of the excellent thermal tracking), plus quite a few others.

I hope that I have shed some light on the subject, and that you get some benefit from my musings.

References

1. Philips "Miniwatt" Technical Data, 7th Edition, 1972
2. RCA Receiving Tube Manual, 1968
3. Basic Electronics - Grob, McGraw Hill, 1971
4. Radiotron Designer's Handbook - Langford-Smith, AWV Pty. Ltd, 1957
5. Analysis and Design of Electronic Circuits - P.M. Chirlean, McGraw Hill, 1965
6. Data Sheets, various

Previous

Articles Index

Main Index

Amp Basics Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999-2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 20 Dec 1999
Updates: 06 Apr 05 - layout update, divided article into sections
13 Feb 03 - Corrected error (1 Siemens = 1 Mho, not 1000 uMho as I stated in the note)
29 Jan 03 - Some corrections, and cleared up a couple of misleading/ambiguous statements
10 Jan 2000 - added section 5
22 Dec 1999 - found and fixed a few typos, and other small errors. 21 Dec 1999 - added new sections

The No. 1 online resource for
starting in electronics soldering. Now referenced by the
Edinburgh Engineering Virtual Library.

Bookmark this site!

The Basic Electronics Soldering & Desoldering Guide

by Alan Winstanley
EPE Magazine Online Editor

How to solder - types of iron - desoldering - troubleshooting.

Types of Iron?
The Art of Soldering
NEW "Cold Soldering" - better than a traditional iron?
How to De-solder
Desoldering Photo Gallery & Black Museum of Bad Soldering
Electronics Soldering Guide Photo Gallery

New! "Cold Soldering" - a review of the Coldheat Classic Iron, tested on a commercial electronic kit. What are the benefits and drawbacks? More details....

This written guide will help beginners and novices to obtain effective results when soldering **electronic components**. If you have little or no experience of using a soldering iron, then EPE *(Everyday Practical Electronics* magazine) recommends that you practice your soldering technique on some fresh surplus components and clean stripboard (protoboard), before experimenting with a proper constructional project. This will help you to avoid the risk of disappointment when you start to assemble your first prototypes. If you've never soldered before, then read on!

Everyday Practical Electronics magazine contains the widest variety of interesting projects and information for beginners, trainees, hobbyists and professionals in electronics. This *Basic Soldering Guide* was condensed from our fully-illustrated series *Build Your Own Projects*, written by Alan Winstanley and published in Everyday Practical Electronics magazine, from November 1996 to March 1997. EPE is read in more than sixty countries: to go to our Subscriptions Page, please click here. For EPE Back Issue information, please click here.

Please refer to the Copyright Notice appearing at the end.

Soldering irons

Topics in this section include:

Voltage

Wattage

Temperature Control

Soldering Stations

Anti Static Protection

Bits (Tips)

Spare Parts

Gas-Powered irons

The most fundamental skill needed to assemble any electronic project is that of ***soldering***. It takes some practice to make the perfect joint, but, like riding a bicycle, once learned is never forgotten! The idea is simple: to join electrical parts together to form an electrical connection, using a molten mixture of lead and tin (solder*) with a soldering iron. A large range of soldering irons is available - which one is suitable for you depends on your budget and how serious your interest in electronics is.

[*Note: the use of lead in solder is now increasingly prohibited in many countries. "Lead free" solder is now statutory instead.]

Electronics catalogues often include a selection of well-known brands of soldering iron. Excellent British-made ones include the universally popular Antex, Adcola and Litesold makes. Other popular brands include those made by Weller and Ungar. A very basic mains electric soldering iron can cost from under £5 (US$ 8), but expect a reasonable model to be approximately £10-£12 (US$ 16 - 20) - though it's possible to spend into three figures on a soldering iron "station" if you're really serious! Check some suppliers' catalogues for some typical types. Certain factors you need to bear in mind include:-

Voltage: most irons run from the mains at 240V. However, low voltage types (e.g. 12V or 24V) generally form part of a "soldering station" and are designed to be used with a special controller made by the same manufacturer.

Wattage: Typically, they may have a power rating of between 15-25 watts or so, which is fine for most work. A higher wattage does ***not*** mean that the iron runs hotter - it simply means that there is more power in reserve for coping with larger joints. This also depends partly on the design of the "bit" (the tip of the iron). Consider a higher wattage iron simply as being more "unstoppable" when it comes to heavier-duty work, because it won't cool down so quickly.

Temperature Control: the simplest and cheapest types don't have any form of temperature regulation. Simply plug them in and switch them on! Thermal regulation is "designed in" (by physics, not electronics!): they may be described as "thermally balanced" so that they have some degree of temperature "matching" but their output will otherwise not be controlled. Unregulated irons form an ideal general purpose iron for most users, and they generally cope well with printed circuit board soldering and general interwiring. Most of these "miniature" types of iron will be of little use when attempting to solder large joints (e.g. very large terminals or very thick wires) because the component being soldered will "sink" heat away from the tip of the iron, cooling it down too much. (This is where a higher wattage comes in useful.)

A proper temperature-controlled iron will be quite a lot more expensive - retailing at say £40 (US$ 60) or more - and will have some form of built-in thermostatic control, to ensure that the temperature of the bit (the tip of the iron) is maintained at a fixed level (within limits). This is desirable especially during more frequent use, since it helps to ensure that the temperature does not "overshoot" in between times, and also guarantees that the output will be relatively stable. Some irons have a bimetallic strip thermostat built into the handle which gives an audible "click" in use: other types use all-electronic controllers, and some may be adjustable using a screwdriver.

Yet more expensive still, ***soldering stations*** cost from £70 (US$ 115) upwards (the iron may be sold separately, so you can pick the type you prefer), and consist of a complete bench-top control unit into which a special *low-voltage* soldering iron is plugged. Some versions might have a built-in digital temperature readout, and will have a control knob to enable you to vary the setting. The temperature could be boosted for soldering larger joints, for example, or for using higher melting-point solders (e.g. silver solder). These are designed for the most discerning users, or for continuous production line/ professional use. The best stations have irons which are well balanced, with comfort-grip handles which remain cool all day. A thermocouple will be built into the tip or shaft, which monitors temperature.

Anti-static protection: if you're interested in soldering a lot of static-sensitive parts (e.g. CMOS chips or MOSFET transistors), more advanced and expensive soldering iron stations use static-dissipative materials in their construction to ensure that static does not build up on the iron itself. You may see these listed as "ESD safe" *(electrostatic discharge proof).* The cheapest irons won't necessarily be ESD-safe but never the less will still probably perform perfectly well in most hobby or educational applications, if you take the usual anti-static precautions when handling the components. The tip would need to be well earthed (grounded) in these circumstances.

Bits: it's useful to have a small selection of manufacturer's bits (soldering iron tips) available with different diameters or shapes, which can be changed depending on the type of work in hand. You'll probably find that you become accustomed to, and work best with, a particular shape of tip. Often, tips are *iron-coated* to preserve their life, or they may be bright-plated instead. Copper tips are seldom seen these days.

Spare parts: it's nice to know that spare parts may be available, so if the element blows, you don't need to replace the entire iron. This is especially so with expensive irons. Check through some of the larger mail-order catalogues.

You will occasionally see ***gas-powered*** soldering irons which use butane rather than the mains electrical supply to operate. They have a catalytic element which, once warmed up, continues to glow hot when gas passes over them. Service engineers use them for working on repairs where there may be no power available, or where a joint is tricky to reach with a normal iron, so they are really for occasional "on the spot" use for quick repairs, rather than for mainstream construction or assembly work. A ***solder gun*** is a pistol-shaped iron, typically running

at 100W or more, and is completely unsuitable for soldering modern electronic components: they're too hot, heavy and unwieldy for micro-electronics use. Plumbing, maybe..!

Soldering irons are best used along with a heat-resistant ***bench-type holder***, so that the hot iron can be safely parked in between use. Soldering stations already have this feature, otherwise a separate soldering iron stand is essential, preferably one with a holder for tip-cleaning sponges. Now let's look at how to use soldering irons properly, and how to put things right when a joint goes wrong.

Soldering Guide Photo Gallery

How to solder

Quick Summary Guide

Cleanliness of Components

Temperature

Time

Amount

Turning to the actual techniques of soldering, firstly it's best to *secure the work* somehow so that it doesn't move during soldering and affect your accuracy. In the case of a printed circuit board, various holding frames are fairly popular especially with densely populated boards: the idea is to insert all the parts on one side ("stuffing the board"), hold them in place with a special foam pad to prevent them falling out, turn the board over and then snip off the wires with cutters before making the joints. The frame saves an awful lot of turning the board over and over, especially with large boards. Other parts could be held firm in a modeller's small vice, for example.

Solder joints may need to possess some degree of mechanical strength in some cases, especially with wires soldered to, say, potentiometer or switch tags, and this means that the wire should be looped through the tag and secured before solder is applied. The down side is that it is more difficult to *de-solder* the joint (see later) and remove the wire afterwards, if required. Otherwise, in the case of an ordinary circuit board, components' wires are bent to fit through the board, inserted flush against the board's surface, splayed outwards a little so that the part grips the board, and then soldered.

In my view - opinions vary - it's generally better to snip the surplus wires leads off *first*, to make the joint more accessible and avoid applying a mechanical shock to the p.c.b. joint. However, in the case of semiconductors, I often tend to leave the snipping until *after* the joint has been made, since the excess wire will help to sink away some of the heat from the semiconductor junction. Integrated circuits can either be soldered directly into place if you are confident enough, or better, use a dual-in-line socket to prevent heat damage. The chip can then be swapped out if needed.

Parts which become hot in operation (e.g. some resistors), are raised above the board slightly to allow air to circulate. Some components, especially large electrolytic capacitors, may require a mounting clip to be screwed down to the board first, otherwise the part may eventually break off due to vibration.

The perfectly soldered joint will be nice and shiny looking, and will prove reliable in service. I would say that:

- **cleanliness**
- **temperature**
- **time**
- **adequate solder coverage**

are the key factors affecting the quality of the joint. A little effort spent now in soldering the perfect joint may save you - or somebody else - a considerable amount of time in troubleshooting a defective joint in the future. The basic principles are as follows.

Really Clean

Firstly, and without exception, all parts - including the iron tip itself - must be **clean** and **free from contamination**. Solder just will not "take" to dirty parts! Old components or copper board can be notoriously difficult to solder, because of the layer of oxidation which builds up on the surface of the leads. This repels the molten solder and this will soon be evident because the solder will "bead" into globules, going everywhere except where you need it. *Dirt is the enemy of a good quality soldered joint!*

Hence, it is an absolute necessity to ensure that parts are free from grease, oxidation and other contamination. In the case of old resistors or capacitors, for example, where the leads have started to oxidise, use a small hand-held file or perhaps scrape a knife blade or rub a fine emery cloth over them to reveal fresh metal underneath. Stripboard and copper printed circuit board will generally oxidise after a few months, especially if it has been fingerprinted, and the copper strips can be cleaned using an abrasive rubber block, like an aggressive eraser, to reveal fresh shiny copper underneath.

Also available is a fibre-glass filament brush, which is used propelling-pencil-like to remove any surface contamination. These tend to produce tiny particles which are highly irritating to skin, so avoid accidental contact with any debris. Afterwards, a wipe with a rag soaked in cleaning solvent will remove most grease marks and fingerprints. After preparing the surfaces, avoid touching the parts afterwards if at all possible.

Another side effect of having dirty surfaces is the tendency for people to want to apply *more heat* in an attempt to "force the solder to take". This will often do more harm than good because it may not be possible to burn off any contaminants anyway, and the component may be overheated. In the case of semiconductors, temperature is quite critical and they may be harmed by applying such excessive heat.

Before using the iron to make a joint, it should be "tinned" (coated with solder) by applying a few millimetres of solder, then wiped on a damp sponge preparing it for use: you should always do this immediately with a new bit, anyway. Personally, I always re-apply a very small amount of solder again, mainly to improve the thermal contact between the iron and the joint, so that the solder will flow more quickly and easily. It's sometimes better to tin larger parts as well before making the joint itself, but it isn't generally necessary with p.c.b. work. (All *EPE* printed circuit boards are "roller-tinned" to preserve their quality and to help with soldering.) A worthwhile product is Weller's *Tip Tinner & Cleaner*, a small 15 gram tinlet of paste onto which you dab a hot iron - the product cleans and tins the iron ready for use. An equivalent is Adcola *Tip-Save.*

Normal electronics grade solder is now "lead free" and typically contains Sn 97 Ag 2.5 Cu 0.5 (i.e. 97% tin, 2.5% silver and 0.5% copper). **It already contains cores of "flux"** which helps the molten solder to flow more easily over the joint. Flux removes oxides which arise during heating, and is seen as a brown fluid bubbling away on the joint. **The use of separate acid flux paste (e.g. as used by plumbers) should NEVER be necessary in normal electronics applications because electronics-grade solder already contains the correct grade of flux!** Other solders are available for specialist work, including aluminium and silver-solder. Different solder diameters are produced, too; 20-22 SWG (19-21 AWG) is 0.91-0.71mm diameter and is fine for most work. Choose 18 SWG (16 AWG) for larger joints requiring more solder.

Temperature

Another step to successful soldering is to ensure that the **temperature** of *all* the parts is raised to roughly the same level before applying solder. Imagine, for instance, trying to solder a resistor into place on a printed circuit board: it's far better to heat *both* the copper p.c.b. *and* the resistor lead at the same time before applying solder, so that the solder will flow much more readily over the joint. Heating one part but not the other is far less satisfactory joint, so strive to ensure that the iron is in contact with *all* the components first, before touching the solder to it. The melting point of most solder is in the region of 188°C (370°F) and the iron tip temperature is typically 330-350°C (626°-662°F). The latest lead-free solders typically require a higher temperature.

Now is the time

Next, the joint should be heated with the bit for just the right amount of **time** - during which a short length of solder is applied to the joint. Do **not** use the iron to carry molten solder over to the joint! Excessive time will damage the component and perhaps the circuit board copper foil too! Heat the joint with the tip of the iron, then continue heating whilst applying solder, then remove the iron and allow the joint to cool. This should take only a few seconds, with experience. The heating period depends on the temperature of your iron and size of the joint - and larger parts need more heat than smaller ones - but some parts (semiconductor diodes, transistors and i.c.s), are sensitive to heat and should not be heated for more than a few seconds. Novices sometimes buy a small clip-on heat-shunt, which resembles a pair of aluminium tweezers. In the case of, say, a transistor, the shunt is attached to one of the leads near to the transistor's body. Any excess heat then diverts up the heat shunt instead of into the transistor junction, thereby saving the device from over-heating. Beginners find them reassuring until they've gained more experience.

Solder Coverage

The final key to a successful solder joint is to apply an appropriate amount of solder. *Too much solder* is an unnecessary waste and may cause short circuits with adjacent joints. *Too little* and it may not support the component properly, or may not fully form a working joint. How much to apply, only really comes with practice. A few millimetres only, is enough for an "average" p.c.b. joint, (if there is such a thing).

Desoldering methods

A soldered joint which is improperly made will be electrically "noisy", unreliable and is likely to get worse in time. It may even not have made any electrical connection at all, or could work initially and then cause the equipment to fail at a later date! It can be hard to judge the

quality of a solder joint purely by appearances, because you cannot say how the joint actually formed on the *inside*, but by following the guidelines there is no reason why you should not obtain perfect results.

A joint which is poorly formed is often called a "dry joint". Usually it results from dirt or grease preventing the solder from melting onto the parts properly, and is often noticeable because of the tendency of the solder not to "spread" but to form beads or globules instead, perhaps partially. Alternatively, if it seems to take an inordinately long time for the solder to spread, this is another sign of possible dirt and that the joint may potentially be a dry one.

There will undoubtedly come a time when you need to *remove* the solder from a joint: possibly to replace a faulty component or fix a dry joint. The usual way is to use a ***desoldering pump*** or vacuum pump which works like a small spring-loaded bicycle pump, only in reverse! (More demanding users using CMOS devices might need a pump which is ESD safe.) A spring-loaded plunger is released at the push of a button and the molten solder is then drawn up into the pump. It may take one or two attempts to clean up a joint this way, but a small desoldering pump is an invaluable tool especially for p.c.b. work.

Sometimes, it's effective to actually *add more* solder and then desolder the whole lot with a pump, if the solder is particularly awkward to remove. Care is needed, though, to ensure that the boards and parts are not damaged by excessive heat; the pumps themselves have a P.T.F.E. nozzle which is heat proof but may need replacing occasionally.

An excellent alternative to a pump is to use *desoldering braid*, including the famous American "Soder-Wick" (sic) or Adcola "TISA-Wick" which are packaged in small dispenser reels. This product is a specially treated fine copper braid which draws molten solder up into the braid where it solidifies. The best way is to use the tip of the hot iron to press a short length of braid down onto the joint to be de-soldered. The iron will subsequently melt the solder, which will be drawn up into the braid. Take extreme care to ensure that you don't allow the solder to cool with the braid adhering to the work, or you run the risk of damaging p.c.b. copper tracks when you attempt to pull the braid off the joint. See my photo gallery for more details.

I recommend buying a small reel of de-soldering braid, especially for larger or difficult joints which would take several attempts with a pump. It is surprisingly effective, especially on difficult joints where a desoldering pump may prove a struggle.

Here's a summary of how to make the perfect solder joint.

1. All parts must be clean and free from dirt and grease.
2. Try to secure the work firmly.
3. "Tin" the iron tip with a small amount of solder. Do this immediately, with new tips being used for the first time.
4. Clean the tip of the hot soldering iron on a damp sponge.
5. Many people then add a tiny amount of fresh solder to the cleansed tip.
6. Heat all parts of the joint with the iron for under a second or so.
7. Continue heating, then apply sufficient solder only, to form an adequate joint.
8. Remove and return the iron safely to its stand.
9. It only takes two or three seconds at most, to solder the average p.c.b. joint.
10. Do not move parts until the solder has cooled.

Troubleshooting Guide

- Solder won't "take" - grease or dirt present - desolder and clean up the parts. Or, material may not be suitable for soldering with lead/tin solder (eg aluminium).
- Joint is crystalline or grainy-looking - has been moved before being allowed to cool, or joint was not heated adequately - too small an iron/ too large a joint.
- Solder joint forms a "spike" - probably overheated, burning away the flux.

First Aid

If you are unlucky enough to receive burns which require treatment, here's what to do :-

1. Immediately cool the affected area with cold running water for several minutes.
2. Remove any rings etc. before swelling starts.
3. Apply a sterile dressing to protect against infection.
4. Do not apply lotions, ointments etc., nor prick any blisters which form later.
5. Seek professional medical advice where necessary.

How to contact the author

Written by Alan Winstanley Email to alan@epemag.demon.co.uk.

Soldering Guide Photo Gallery

Bookmark this site!

Copyright Notice

Text © 1996-2006 Wimborne Publishing Limited, Wimborne, Dorset, England. Everyday Practical Electronics Magazine has provided this document as a free web resource to help constructors, trainees and students. You are welcome to download it, print it and distribute it for personal or educational use. It may not be used in any commercial publication, mirrored on any commercial site nor may it be appended to or amended, or used or distributed for any commercial reason, without the prior permission of the Publishers.

Photographs © 1996-2006 Alan Winstanley WORLD COPYRIGHT RESERVED

Every care has been taken to ensure that the information and guidance given is accurate and reliable, but since conditions of use are beyond our control no legal liability or consequential claims will be accepted for any errors herein.

The British mains voltage supply is 230V a.c. and you should amend ratings for local conditions.

Please check the Everyday Practical Electronics Web Site for details of the current issue, subscription rates, Back issue availability and contact information.

Overture™ Series High Power Solutions

National Semiconductor
Application Note 1192
John DeCelles and Troy Huebner
June 2004

Table of Contents

1.0 Introduction 2
2.0 Objective 2
3.0 Conclusion 2
4.0 Thermal Background 2
4.1 TYPICAL CHARACTERISTIC DATA 3
4.2 SINGLE-ENDED AMPLIFIER Pdmax EQUATION: 3
4.3 BRIDGED-OUTPUT AMPLIFIER Pdmax EQUATION 3
4.4 PARALLEL AMPLIFIER Pdmax EQUATION 4
4.5 BRIDGED/PARALLEL AMPLIFIER Pdmax EQUATION 4
4.6 THERMAL CONCLUSION 4
4.7 THERMAL TESTING CONDITIONS 5
5.0 BR100—100W Bridge Circuit 5
5.1 AUDIO TESTING 5
5.1.1 Linearity Tests 5
5.2 SCHEMATICS 6
5.2.1 Bridged Amplifier Schematic 6
5.2.2 Electrical Design Notes 6
6.0 PA100—100W Parallel Circuit 7
6.1 AUDIO TESTING 7
6.1.1 Linearity Test 7
6.2 SCHEMATICS 8
6.2.1 Parallel Amplifier Schematic 8
6.2.2 Electrical Design Notes 8
7.0 BPA200–200W Bridged/Parallel Circuit 9
7.1 AUDIO TESTING 9
7.1.1 Linearity Tests 9
7.1.2 Output Power Tests 9
7.1.3 Noise Floor Tests 10
7.1.4 Electrical Design Notes 11
7.2 SCHEMATICS 12
7.2.1 Detailed Bridged/Parallel Amplifier Schematic 12
7.2.2 Servo Circuits 13
7.2.3 Power Supply Circuit 14
7.2.4 Basic Bridged/Parallel Amplifier Schematic 15
8.0 Parts List and Vendors 16
8.1 BUILD OF MATERIALS FOR BR100 AMPLIFIER 16
8.2 BUILD OF MATERIALS FOR PA100 AMPLIFIER 16
8.3 BUILD OF MATERIALS FOR BPA200 AMPLIFIER 18
9.0 Heat Sink Drawings 19
9.1 BR100 AND PA100 HEAT SINK DRAWING 19
9.2 BPA200 HEAT SINK DRAWING 20

Overture® is a registered trademark of National Semiconductor Corporation.
SPiKe™ is a trademark of National Semiconductor Corporation.

© 2004 National Semiconductor Corporation

1.0 Introduction

National Semiconductor has a broad portfolio of monolithic power integrated circuits covering power levels from a few hundred milliwatts up to 60W of non-clipped continuous average power. These ICs cover most audio applications by themselves, however, for really high power applications, other methods need to be employed because IC packages have limited power dissipation capabilities.

There are many different ways of obtaining over 100W of output power. Most high-end power amplifier manufacturers utilize discrete circuits which allows them to market their amplifiers as “specially designed.” However, there is a price to be paid for discrete amplifier designs; they are complex, difficult to design, require many components, lack the comprehensive protection mechanisms of integrated circuits and are not as reliable.

Other methods of obtaining output power greater than 100W include the use of power ICs as drivers for discrete power transistors. There are a number of these types of circuits, but they too possess all of the same flaws as discrete circuits, including a lack of comprehensive output stage protection.

2.0 Objective

The objective is to provide simple high power solutions that are conservatively designed, highly reliable and have low part count. This document provides three specific, but not unique, application circuits that provide output power of 100W, 200W, and above. These circuits are the parallel, bridged, and bridged/parallel configurations.

These three circuits are simple to understand, simple to build and require very few external components compared to discrete power amplifier designs. Simplicity of design and few components make this solution much more reliable than discrete amplifiers. In addition, these circuits inherently possess the full protection of each individual IC that is very difficult and time consuming to design discretely. Finally, these circuits are well known and have been in industry for years.

3.0 Conclusion

The BR100 (100W Bridged Circuit), PA100 (100W parallel circuit), and the BPA200 (200W Bridged/Parallel Circuit) are high power solutions that can be used in many applications, but they are primarily targeted for home theater amplifier applications such as powered subwoofers, self-powered speakers, and surround sound amplifiers.

While bridged amplifier configurations are able to provide high power levels, they also consume four times more power than a conventional single-ended solution. However, it is feasible to conservatively design a 100W bridged amplifier solution, as will be shown here. The bridged solution is designed to drive an 8Ω nominal load for self-powered speaker or powered subwoofer applications.

The parallel amplifier is another configuration that can be used to obtain higher output power levels by combining two IC outputs and doubling output current drive capability. The parallel topology provides a great way of achieving higher power levels while keeping within IC power dissipation limits by driving low impedance loads, which is the case for many self-powered speaker and powered subwoofer designs. The main advantage of the parallel configuration is its ability to divide total power dissipation between ICs, since each amplifier is providing half of the load current. Another advantage of the parallel design is that unlike the bridge design, more than two ICs can be used. In fact, any number of ICs can be used in a parallel design and when configured the same will share the power dissipation equally. For example, using four ICs to drive a 1Ω load means that each IC dissipates 1/4 of the total power dissipation. In other words, the load to each IC looks like a 4Ω load (Number Of ICs in Parallel * Load Impedance = Load Impedance seen by each individual IC.) Odd numbers of ICs can also be used.

For lower impedance loads (<8Ω), the parallel circuit is a good solution for 100W power levels using just two devices. Power levels above 100W may be obtained by using more than two devices to increase output current capability and power dissipation limits along with lower impedance loads.

If the bridged and parallel configurations are combined, the outcome is a very high power amplifier solution that far exceeds the capabilities of one IC alone, while maintaining reasonable power dissipation levels within each IC. The bridged portion doubles the output voltage swing and quadruples the total power dissipation while the parallel portion halves the current between each IC set and divides the total power dissipation between each of the four ICs. The result is higher system output power with each IC not exceeding its individual power dissipation capabilities. Higher output power levels are attained, while the ICs run at a normal temperature, keeping long term reliability high. The schematic of the Bridged/Parallel Amplifier is shown in Figure 13.

The bridged/parallel circuit using four devices will produce the maximum output power (>200W) into loads with an impedance from 4Ω to 8Ω. For loads less than 4Ω, additional devices may need to be placed in parallel or the supply voltage reduced.

The data in the following sections will exemplify that the parallel, bridged, and bridged/parallel solutions using multiple power ICs can meet high fidelity specifications while providing output power from 100W up to 400W. The low noise and excellent linearity traits of the monolithic IC are transferred to the high power solution, making the circuit even more attractive. In addition, the protection mechanisms within the IC, which are not easily designed discretely, are inherently designed into the circuit.

While the data show what specs can be achieved by the configurations, as always, good design practices need to be followed to achieve the stated results. In addition to good electrical and layout design practices, the thermal design is equally critical with Overture ICs. The following section will expand on the thermal design aspects of Overture ICs. This concept of “design by power dissipation” is applicable to all types of high power solutions.

The PA100, BR100, and BPA200 schematics and test results exemplify what can be achieved with proper component selection, thermal design, and layout techniques.

4.0 Thermal Background

The voltage and current ratings of a power semiconductor are typically the first specs considered in designing high power amplifiers. The same is true for an integrated monolithic power amplifier. However, power dissipation ratings are equally important to the long term reliability of the power amplifier design. When using a monolithic IC in its intended application and within its specified capabilities, the thermal design is relatively straightforward. When an IC is used beyond is capabilities, as in high power circuits, power dissipation issues become more critical and not as straight-

4.0 Thermal Background (Continued)

forward. Therefore, the designer must understand the IC's power dissipation capabilities before using the IC in a booster configuration.

4.1 TYPICAL CHARACTERISTIC DATA

The power dissipation capabilities of a power IC are either specified in the datasheet or can be derived from its guaranteed output power specification. While the power dissipation rating for the LM3886T is 125W, this number can be misleading. Its power dissipation specification is derived from the IC's junction-to-case thermal resistance, θ_{JC} = 1˚C/W, the maximum junction temperature, T_J = 150˚C, and the ambient air, T_A = 25˚C. As stated in the datasheet, the device must be derated based on these parameters while operating at elevated temperatures. The heat sinking requirements for the application are based on these parameters so that the IC will not go into Thermal Shutdown (TSD). The real problem for Overture ICs, however, comes from the sensitivity of the output stage's unique SPiKe™ Protection which dynamically monitors the output transistor's temperature. While the thermal shutdown circuitry is enabled at T_J = 150˚C, SPiKe circuitry is enabled at T_J = 250˚C for instantaneous power spikes in the output stage transistor. As the overall temperature of the IC increases, SPiKe circuitry becomes even more sensitive causing it to turn on before the 125W limit is reached. TSD circuitry will continue to function globally for the IC in conjunction with the SPiKe circuitry. However, protection circuitry should not be activated under normal operating conditions. The question then becomes, what is the power dissipation limit for the IC such that SPiKe circuitry is not enabled? Knowing the power dissipation limit and keeping the case temperature of the IC as cool as possible will expand the output power capability without activating SPiKe Protection.

The other way to determine IC power dissipation capabilities is to analyze the output power specification in the datasheet. In the case of the LM3886T, there are two output power specification guarantees: 60W (min) into a 4Ω load using ±28V supplies and 50W(typ) into an 8Ω load from ±35V supplies. Using these two conditions and the theoretical maximum power dissipation equation shown below, results in the following maximum power dissipations:

4.2 SINGLE-ENDED AMPLIFIER Pdmax EQUATION:

$$Pdmax = V_{CCtot}^2/2\pi^2R_L$$

Non-Isolated LM3886T:

$$1.\ V_{CC} = \pm 28V,\ R_L = 4\Omega$$

$$Pdmax = V_{CCtot}^2/2\pi^2R_L = (\pm 28V)^2/2\pi^2(4\Omega) = 39.7W$$

$$2.\ V_{CC} = \pm 35V,\ R_L = 8\Omega$$

$$Pdmax = V_{CCtot}^2/2\pi^2R_L = (\pm 35V)^2/2\pi^2(8\Omega) = 31.0W$$

These results show that the IC can handle a maximum of ≈ 40W of continuous power dissipation without SPiKe Protection being turned on under continuous sinusoidal input with proper heat sinking. The same theory applies to other Overture ICs as well, like the LM3876T, which is capable of dissipating 31W with proper heat sinking. It should be noted that the results shown above are for the non-isolated power package, where the back of the package is tied to the silicon substrate, or –Vee. The isolated power package has overmolded plastic on the back keeping the package electrically isolated from the silicon substrate. This extra amount of plastic increases the package thermal resistance from 1˚C/W for the non-isolated version to ≈ 2˚C/W for the isolated version. The result of increased thermal resistance is higher die temperature under the same conditions even though the heat sink temperature will not change.

There are two major points to note:

1. The maximum power dissipation analysis was taken into account using regulated power supplies. The IC for the whole analysis is being tested at the worst case power dissipation point for a constant full-load power supply voltage. When using an unregulated power supply, the no-load voltage will be somewhat higher (15%–35%) causing the overall maximum power dissipation to be higher than expected.
2. In the real "audio" application, the average music power dissipation is much less than the maximum power dissipation created by a sinusoidal input. Therefore, the IC will run cooler than expected due to the lower power dissipation.

However, when you put these two points together, they mostly cancel out, but only for music stimulus. Product qualifications may go through worse case power dissipation scenarios which implies that sinusoids will be used with unregulated power supplies. Therefore, when doing the thermal portion of the design, the higher supply voltages will increase the IC power dissipation and must be taken into account.

4.3 BRIDGED-OUTPUT AMPLIFIER Pdmax EQUATION

To determine the Pdmax equation for a bridged amplifier solution, the single-ended Pdmax equation is used as a starting point. A bridged amplifier solution requires two amplifiers and each amplifier will see 1/2 the total impedance. Adding these factors of 2 and 1/2 into the single-ended Pdmax equation results in the total Pdmax equation for a bridged amplifier.

$$Pdmax_{BTL} = 2*[V_{CCtot}^2/2\pi^2(/½R_L)] = 4*(V_{CCtot}^2/2\pi^2R_L)$$

The bridged-output Pdmax equation represents the bridged amplifier solution. If a dual amplifier IC is used, then the total Pdmax would need to be dissipated in the single IC package. However, if two individual ICs are used, then the total power dissipation is divided between each IC.

Two Non-Isolated LM3886Ts:

$$V_{CC} = \pm 28V,\ R_L = 4\Omega$$

$$Pdmax = 4V_{CCtot}^2/2\pi^2R_L = 4(\pm 28V)^2/2\pi^2(4\Omega) = 158.8W$$

$$Pdmax = 158.8W$$

$$Pdmax/IC = 79.4W$$

Therefore, using a bridged configuration, V_{cc} would have to be equal to ±20V to keep the IC's power dissipation within

4.0 Thermal Background (Continued)

40W/IC when driving a 4Ω load! This equates to about 110W of output power in bridged-mode driving a 4Ω load. When driving an 8Ω load, and using the same bridged pdmax equation and a maximum of 40W/IC of power dissipation, the supply voltages would have to be ±28V. This equates to about 120W of output power.

4.4 PARALLEL AMPLIFIER Pdmax EQUATION

To determine the Pdmax equation for a parallel amplifier solution the single-ended Pdmax equation is used as a starting point. Since a parallel solution has the load connected the same as single-ended solution (one side to GND) just more devices driving the load, the equation does not change for total Pdmax.

$$Pdmax_{PA} = V_{CCtot}^2/2\pi^2R_L$$

The advantage of the parallel solution is total Pdmax is divided equally among each of the amplifiers in the parallel solution. By dividing up the total power dissipation among two or more ICs, lower impedance loads can be driven for much higher power solutions. National's Overture power amplifier series amplifiers will give the most output power and power dissipation will be kept within limits when each amplifier sees a load impedance of 4Ω – 8Ω.

Each amplifier in a parallel solution sees a load impedance equal to the total load impedance * the number of amplifiers used. So for a 4Ω solution using two amplifiers will result in each amplifier seeing an 8Ω load. Using National's LM3886 in a two device parallel solution driving a 4Ω load will typically provide 110W of output power. For a 2Ω solution using two LM3886 ICs will result in each IC seeing a 4Ω load and typically provide 120W of output power. Or four LM3886 ICs may be used so each IC sees an 8Ω load typically providing 200W of output power. Three ICs may also be used so that each IC sees a 6Ω load typically providing 150W of output power.

4.5 BRIDGED/PARALLEL AMPLIFIER Pdmax EQUATION

The bridged/parallel amplifier consist of two amplifiers in bridge mode then additional amplifiers in parallel to the amplifiers on each side of the bridge (see Figure 13 and 17). To find the equation for $Pdmax_{BPA}$ the $Pdmax_{BTL}$ equation is used as a starting point. As discussed above, adding devices in parallel does not change the Pdmax equation so the total Pdmax for a bridged/parallel solution is the same as for a bridged solution.

$$Pdmax_{BPA} = 4*(V_{CCtot}^2/2\pi^2R_L)$$

This total power dissipation is divided equally among all of the amplifiers in the circuit. To determine the Pdmax equation for each amplifier in a bridged/parallel circuit the $Pdmax_{BPA}$ equation is used as a starting point. The $Pdmax_{BPA}$ equation calculates the total peak power dissipation for the entire circuit. To find the power for each side of the bridge the $Pdmax_{BPA}$ equation must be multiplied by a factor of 1/2 giving:

$$Pdmax_{BPA(IC)} = 4*(V_{CCtot}^2/2\pi^2R_L) * (½) = 2*(V_{CCtot}^2/2\pi^2R_L)$$

Adding one additional amplifier in parallel to each side of the bridge as shown in Figures 13 and 17 divides the power between the two amplifiers in parallel on each side of the bridge. The equation must be multiplied by another factor of 1/2 giving:

$$Pdmax_{BPA(IC)} = 2*(V_{CCtot}^2/2\pi^2R_L) * (½) = V_{CCtot}^2/2\pi^2R_L$$

which is just the single-ended Pdmax equation. If more devices are added in parallel to each side of the bridge then instead of a factor of ½ a factor of 1/(Number of Amplifiers) would be used in the last step above.

An alternate way to arrive at the $Pdmax_{BPA(IC)}$ equation is to use the singled-ended Pdmax equation and determine the load impedance seen by each amplifier in the circuit. As discussed above, a bridge circuit means each side of the bridge will see 1/2 the load impedance and a parallel circuit results in each amplifier seeing a load impedance equal to the load * the number of amplifiers in parallel. Putting these two results together gives a general equation for $Pdmax_{BPA(IC)}$.

$$Pdmax_{BPA(IC)} = V_{CCtot}^2/2\pi^2[(\#\text{ of amps in parallel}/2)*R_L]$$

Where # of amps in parallel is the number of amplifiers in parallel on each side of the bridge and not the total number of amplifiers in parallel. As an example, if a 4Ω load is used and the number of amplifiers in parallel is three per side of the bridge (six amplifiers total) then the load seen by each amplifier is: 3/2 * 4Ω = 6Ω.

4.6 THERMAL CONCLUSION

Because of National's portfolio of products and the capabilities of the bridged/parallel circuit, the bridged solution is applicable for a power output window between 80W and 120W. Trying to exceed this power level without a rigorous thermal design will be difficult to achieve. More caution needs to be applied along with better thermal management for bridged circuit designs. The proposed bridged/parallel solution is a more robust design than the bridged circuit, allowing higher output power levels to be obtained by paralleling the two bridged sets of ICs. Table 1 below summarizes the maximum supply voltages for each type of configuration and load impedance while keeping Pdmax per LM3886 IC to less than 40W. See Figure 3, 6, and 13 for detailed information on each circuit.

TABLE 1. Maximum Power Supply Voltages

	2Ω	4Ω	6Ω	8Ω	16Ω
BR100	NR*	±20V	±24V	±28V	±37V
PA100	±28V	±37V	±37V	NA**	NA**
BPA200	±20V	±28V	±32V	±37V	NA**

*NR = Not Recommended

**NA = No Advantage compared to single-ended or other configurations

In addition to better heat sinking, the application of a small fan can substantially increase the IC's continuous power dissipation capabilities. While the air flow of the fan used to take the data is not known, its air flow seemed to be consistent with a typical computer fan. The IC maximum power dissipation data for an individual LM3886 is summarized below in *Table 2*. The data shown below should only be used as a guideline of possible IC power dissipation capability. Your electrical design parameters and thermal management may be different, changing the achievable results. As always, lab testing is recommended to verify any solution.

TABLE 2. Power Dissipation Results

Power IC	Pdmax (No Fan)	Pdmax (With Fan)
LM3886T	40W	60W
LM3886TF	30W	45W

4.0 Thermal Background (Continued)

4.7 THERMAL TESTING CONDITIONS

The data summarized in *Table 2* was obtained by using the bridged/parallel configuration and the following conditions: The system was warmed up for an hour using a power dissipation of 30W per device with a 4Ω load. Four different temperature points were measured after stabilizing, then the supply voltages were incremented while insuring that SPiKe Protection was not enabled during each test by monitoring each amplifier output. The supply voltages continued to be incremented until SPiKe protection or thermal shutdown was enabled, providing the IC's power dissipation limits under those operating conditions.

The input stimulus was a 20Hz sinewave with an amplitude corresponding to the worst case power dissipation for the given load and supply voltage. The ICs were evenly spread out along the heatsink with dimensions of: 3.25" high x 13.25" long x 1.3125" deep. The main body of the heatsink is 0.25" thick with (10) 1.0625" deep fins and the heatsink is black anodized. (See section 9.2 for detailed drawing.) Unfortunately, the fins ran horizontally, which hindered heat radiation without a fan, but helped with air flow and heat dissipation when a fan was used.

This same testing procedure can be used for any number of booster circuits, including variations of the bridged/parallel circuit. Another variation would be to add more ICs in parallel to further reduce power dissipation, allowing low impedance loads to be driven to obtain even higher output power levels.

5.0 BR100—100W Bridge Circuit

5.1 AUDIO TESTING

The following graphs represent the performance level attainable from the bridged circuit found in Figure 3 with a well designed PCB and properly heat sinked. The testing focused on maximum output power capabilities and amplifier linearity. The low THD+N plots shown in Figures 1 & 2 exemplify the high degree of linearity of the bridged circuit which directly translates into a cleaner sounding more transparent amplifier. Other bridged circuit topologies that use the output of one amplifier as the input to the second inverting amplifier inherently possess higher THD and noise that will degrade the solution's sound quality.

5.1.1 Linearity Tests

The linearity of the amplifier is represented by the low THD+N values shown in *Figures 1, 2*. *Figure 1* represents the THD+N vs Frequency for 1W, 56W, and 100W power levels. The 20kHz THD+N is less than 0.02% for 1W and about 0.008% for 56W and above. For normal listening levels, the THD+N is about 0.004% for most of the audio band. *Figure 2* represents the THD+N vs Output Power Level for 20Hz, 1kHz, and 20kHz. The THD+N between 20Hz and 1kHz is less than 0.004% from 1W to the clipping point. The 20kHz THD+N is less than 0.02% from 1W to the clipping point. The continuous clipping point power is around 105W while the power at 10% THD+N is about 140W. These THD+N graphs were obtained using relative THD units, which indicates that the noise level for the amplifier is quite low. Typically, the noise level becomes a significant THD+N contributor at low power levels and shows up as a linearly decreasing function of increasing input signal amplitude. The low power level THD+N for this amplifier is more than acceptable for home entertainment applications.

Figure 3 represents the bridged amplifier schematic. The design is extremely simple, consisting of a non-inverting power op amp configuration and an inverting power op amp configuration. The input to the amplifier solution goes to each individual configuration. While closed-loop gain matching is not critical, it is recommended to have fairly close values. The main functional point to note about this solution is that for a positive going input signal, amplifier U1 will have a positive changing output signal while U2 will have a negative changing output signal. The final voltage across the load is two times the peak amplitude of each individual amplifier output. Since output power is based on the square of the output voltage, the output power is theoretically quadrupled. This document will not go further into the functionality of the circuit as it is widely known in industry.

BR100 THD+N vs Frequency, R_L = 8Ω, V_{CC} ±25.5V, BW <80kHz, P_O = 1W, 56W, 100W

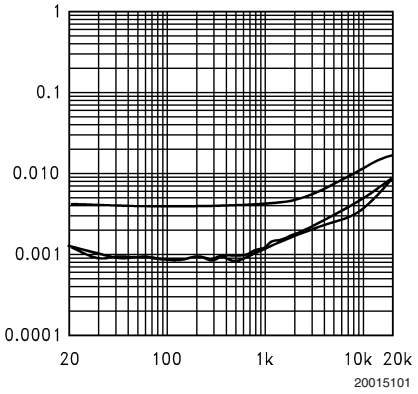


FIGURE 1. THD+N vs Frequency

BR100 THD+N vs Output Power f = 20Hz, 1 kHz, 20kHz, R_L = 8Ω, V_{CC} = ±25.5V, BW <80kHz

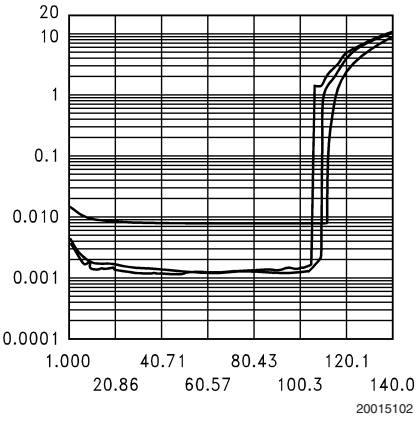


FIGURE 2. THD+N vs Output Power

5.0 BR100—100W Bridge Circuit

(Continued)

5.2 SCHEMATICS

5.2.1 Bridged Amplifier Schematic

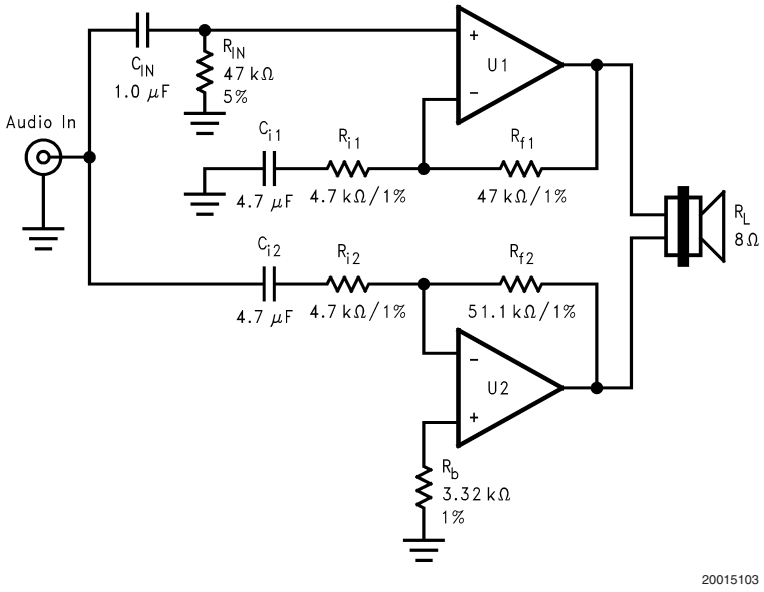


FIGURE 3. Bridged Amplifier Schematic

5.2.2 Electrical Design Notes

The following electrical design notes will aid in making the bridged amplifier design go more smoothly while also helping to achieve the highest level of performance.

- The input impedance of the inverting amplifier is essentially resistor, R_i. The value of this resistance affects the gain setting of the amplifier as well as the low frequency rolloff in conjunction with C_i. There is a tradeoff between having a low frequency rolloff, a high input impedance and a small capacitor size and value. It is critical to have a flat band response down to 20Hz while it is equally important to have a high enough input impedance so that heavy loading does not occur from the preamp stage. Using large valued low-cost capacitors implies the use of leaky electrolytics which affect the output offset voltage. Electrolytic capacitors are also less linear than other premium caps and should not be used in the signal path when not necessary. This tradeoff issue is the toughest portion of the design. The amplifier gain setting is just as one would expect for an inverting op amp. Of course, the input impedance issue can be quickly resolved by using a voltage follower as an input buffer, but it was omitted from this design to minimize cost and simplify the design. The values provided in the bridged schematic are at a good tradeoff point. There is sufficient input impedance for practically all audio op amps, the closed-loop gain setting is 11 for each amplifier, (gain of 22 overall) while the capacitor value of 4.7µF sets the low frequency –3dB rolloff at about 7Hz.
- The non-inverting input resistance, R_b, is used to create a voltage drop at the non-inverting terminal to offset the voltage at the inverting input terminal due to the input bias current flowing from the output to the inverting input. Generally, the value of this resistor equals the value of the feedback resistor so that the output offset voltage will be minimized close to zero. However, if this value is too large, noise can easily be picked up which will be amplified and seriously affect the THD+N performance. If the resistor is eliminated and the terminal is grounded, the THD+N performance will be much better, but it will not necessarily be optimized. By connecting the non-inverting input directly to a ground reference, any noise on that ground will be directly injected into the amplifier, amplified and thus will also affect the THD+N performance. The best solution is to use a value of resistance that is not too large that it picks up stray noise and not too small as to be affected by ground noise fluctuations. The value used in the previous plots was a 3.32kΩ resistor. It should be noted that this is not necessarily the optimized value and can change with varying circuit layouts.
- Low leakage signal path capacitors should be used where possible to reduce output offset voltages. This is not too big of an issue since each gain stage has only unity gain at DC. This is another reason why 1% resistor tolerances are not necessarily required. To obtain the highest quality amplifier, polypropylene capacitors should be employed in the signal path and supply bypassing.

5.0 BR100—100W Bridge Circuit

(Continued)

- As always, the better the supply bypassing, the better the noise rejection and hence higher performance.

6.0 PA100—100W Parallel Circuit

6.1 AUDIO TESTING

The following graphs are the same format as those presented in the BR100 analysis, namely THD+N verses Frequency at 1W, 56W and 100W and THD verses Output Power with plots at 20Hz, 1kHz, and 20kHz.

6.1.1 Linearity Test

The linearity of the amplifier is represented by the low THD+N values shown in Figures 4, 5. Figure 4 represents the THD+N vs Frequency for 1W, 56W and 100W power levels. The 20kHz THD+N is less than 0.05% for all power levels. Figure 5 represents the THD+N vs Output Power Level for 20Hz, 1kHz, and 20kHz. The THD+N between 20Hz and 1kHz is less than 0.01% for power levels above 1W up to the clipping point. The 20kHz THD+N is 0.04% from 0.1W to the clipping point. The 1% THD+N power point is around 110W while the 10% THD+N power point is near 150W. These THD+N graphs were obtained using relative THD+N units, which indicates that the noise level for the amplifier is very low. Typically, the noise level becomes a significant THD+N contributor at lower power levels and shows up as a linearly decreasing function of increasing input signal amplitude. The low power level THD+N for this amplifier configuration is more than acceptable for home entertainment applications.

Figure 6 represents the parallel amplifier schematic. The design is extremely simple, consisting of two power op amps configured identically and tied in parallel to the load each through a 0.1Ω/3W resistor. The closer matched the gain of each IC the more equal the current sharing between them as well as the temperature of each IC due to power dissipation being near equal. This document will not go further into the functionality of the circuit as it is well known in industry.

PA100 THD+N vs Frequency
R_L = 4Ω, V_{CC} = ±35V,
BW < 80kHz, P_O = 1W, 56W, 100W

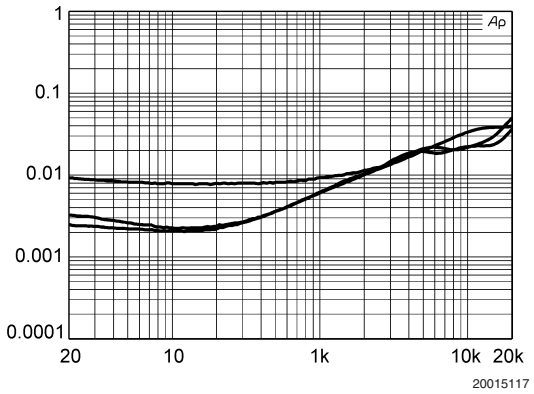


FIGURE 4. THD+N vs Frequency

PA100 THD+N vs Output Power
f = 20Hz, 1kHz, 20kHz,
R_L = 4Ω, V_{CC} = ±35V, BW < 80kHz

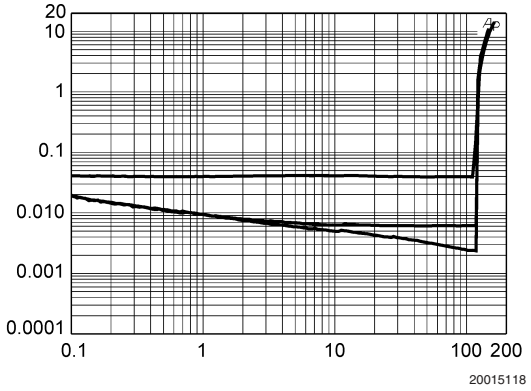


FIGURE 5. THD+N vs Output Power

6.0 PA100—100W Parallel Circuit

(Continued)

6.2 SCHEMATICS

6.2.1 Parallel Amplifier Schematic

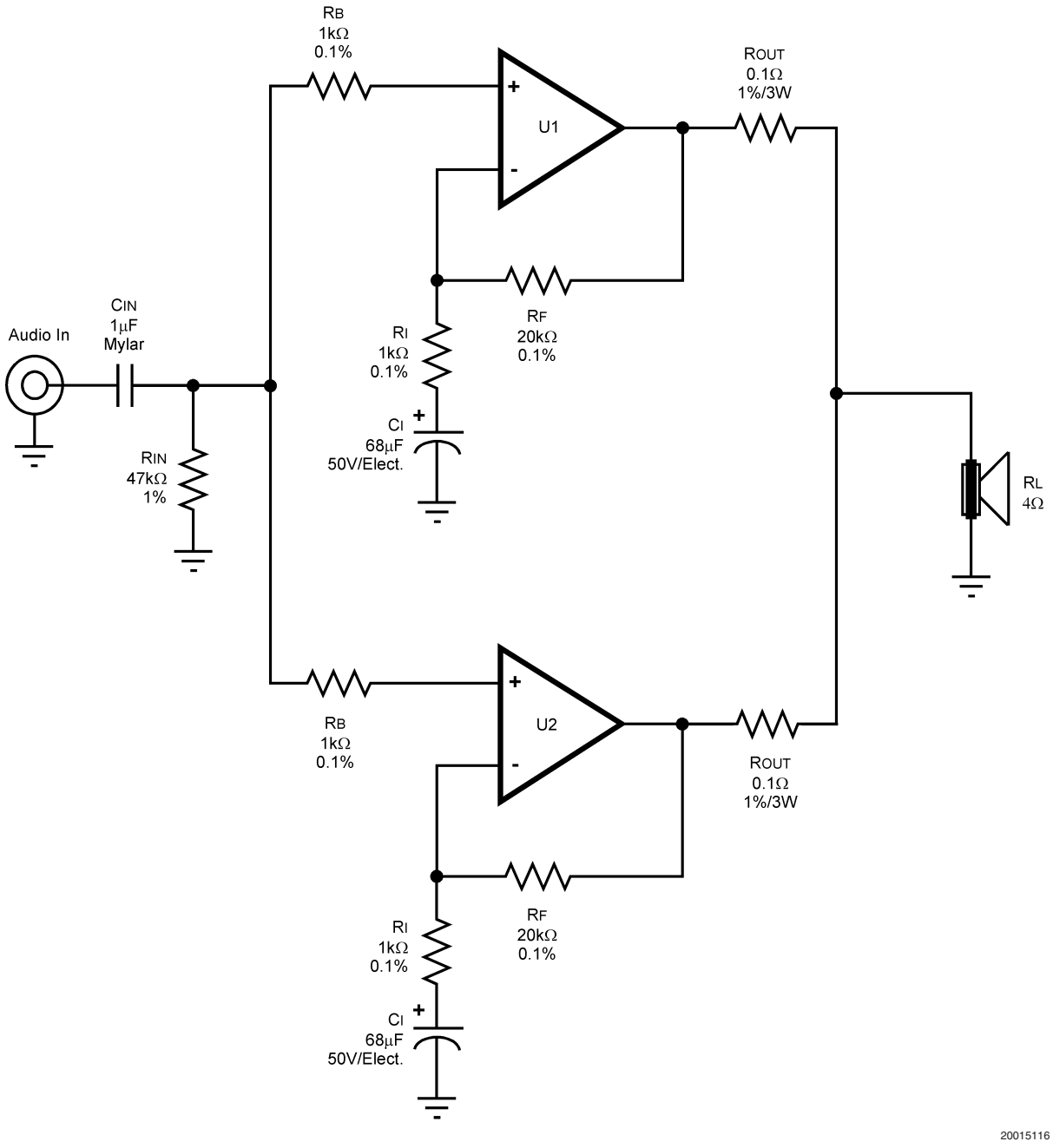


FIGURE 6. Parallel Amplifier Schematic

6.2.2 Electrical Design Notes

The following electrical design notes will aid in making the parallel amplifier design go more smoothly while also helping to achieve the highest level of performance.

6.0 PA100—100W Parallel Circuit

(Continued)

- The input resistance is equal to R_{IN}.The value of R_{IN} should be high enough to eliminate any loading placed on the previous stage (i.e. pre-amplifier). The DC blocking input capacitor value should be calculated on the value of Rin to be sure the correct size is used so low frequency signals will be coupled in without severe attenuation. $f_{IN} = 1/(2\pi R_{IN}C_{IN})$.
- 1% gain setting resistors (R_i and R_f) will give good results but it is recommended 0.1% tolerance resistors be used for setting the gain of each op amp for closer matched gain and equal output current and power dissipation.
- The output resistors, R_{OUT}, wattage rating is based on the load impedance and the output current or maximum output power. As the load impedance is increased or reduced the output current is reduced or increased, respectively. The wattage rating of R_{OUT} should increase as output current increases and decrease as output current decreases. A very conservative design will use peak output current to calculate the needed wattage rating of R_{OUT} ($P = I^2R$).
- As always, the better the supply bypassing, the better the noise rejection and hence higher performance.

7.0 BPA200–200W Bridged/Parallel Circuit

7.1 AUDIO TESTING

The following graphs represent the performance level attainable from the bridge/parallel circuit found in Figure 13 with a well designed PCB and properly heat sinked. The testing focused on maximum output power capabilities, amplifier linearity and noise level.

7.1.1 Linearity Tests

The linearity of the amplifier is represented by the low THD+N values shown in *Figures 7, 8*. *Figure 7* represents the THD+N vs Frequency for 1W, 56W, and 200W power levels. *Figure 8* represents the THD+N vs Output Power Level for 20Hz, 1kHz, and 20kHz. The THD+N between 20Hz and 1 kHz is less than 0.004% from 1W to the clipping point. The 20kHz THD+N is less than 0.02% from 1W to the clipping point, The continuous clipping point power is around 210W while the power at 10% THD+N is 300W. These THD+N graphs were obtained using relative THD+N units, which indicates that the noise level for the amplifier is quite low. Typically, the noise level becomes a significant THD+N contributor at low power levels and shows up as a linearly decreasing function of increasing input signal amplitude. In *Figure 8*, the THD+N decreases from 0.004% to 0.001% from 1W to the clipping point for frequencies between 20Hz and 1kHz. The THD+N with a 20kHz input decreases from 0.02% to 0.009% from 1W to 50W and rises thereafter up to about 0.015%.

BPA200 THD+N vs Frequency
P_O = 1W, 56W, 200W
R_L = 8Ω, BW < 80 kHz, 9/16/97

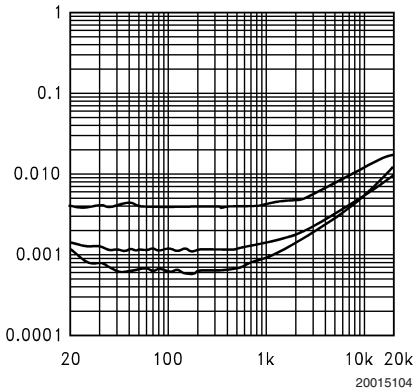


FIGURE 7. THD+N vs Frequency

BPA200 THD+N vs Output Power
f = 20Hz, 1kHz, 20kHz,
R_L = 8Ω, BW < 80kHz, 9/16/97

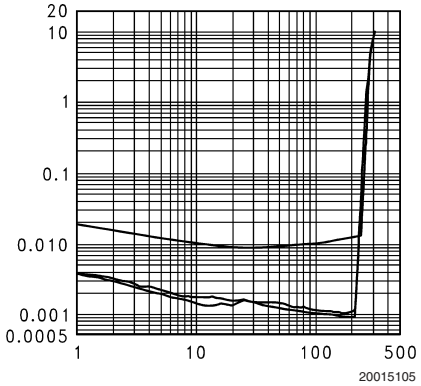


FIGURE 8. THD+N vs Output Power

7.1.2 Output Power Tests

Although the amplifier was deigned based on thermal dissipation capabilities using continuous sinusoidal inputs, the output power levels attainable are significantly greater with pulsed waveforms that more accurately reflect music material. The continuous clipping point power and burst power levels are shown in *Table 3* below:

TABLE 3. BPA200 Maximum Output Power Levels

Load Impedance	Continuous Clipping Point Power	Burst Clipping Point Power
8Ω	225W	295W
4Ω	335W	450W

The burst power levels were obtained using a 20Hz sinewave with two cycles on and twenty cycles off. The output power capability of the BPA200 is further substantiated by the power bandwidth measurement. The amplifier is capable of producing 200W continuously into an 8Ω load up to f = 90.5kHz with little change in THD+N. The graph in *Figure 9* shows the power bandwidth measurement. Also notice that the low frequency power in the graph is not rolled

7.0 BPA200–200W Bridged/Parallel Circuit (Continued)

off as would normally occur with a DC blocking capacitor. The servo circuits allow the low frequency power to remain constant down to DC without high output offset voltage.

BPA200 Power Bandwidth
P_o = 200W, R_L = 8Ω, BW > 500kHz, 9/16/97

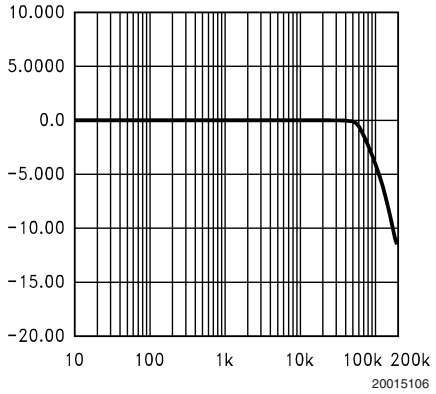


FIGURE 9. Power Bandwidth

7.1.3 Noise Floor Tests

The following plots exemplify the low-noise aspects of the BPA200. *Figure 10* was obtained using an 8k FFT relative to 1dBV with a measurement bandwidth of 22kHz. *Figure 11* is the same measurement as *Figure 10*, but shown in a logarithmic scale.

BPA200 Spot Noise Floor (dBV)
R_L = 8Ω, BW < 22kHz, 9/16/97

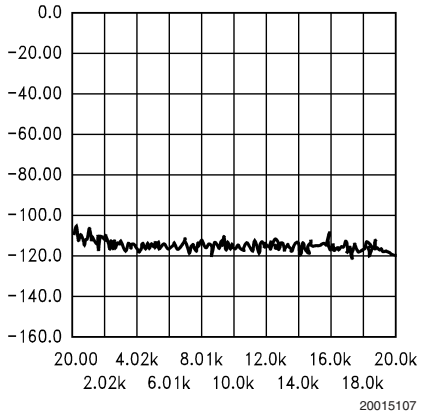


FIGURE 10. Linear-Scale Noise Floor

BPA200 Noise Floor (dBV)
R_L = 8Ω, BW < 22 kHz, 9/16/97

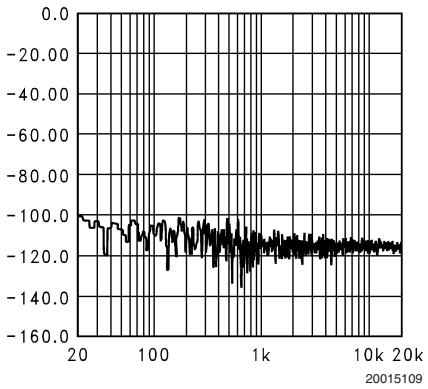


FIGURE 11. Log-Scale Noise Floor

An FFT analyzer is extremely handy in determining the noise culprit when debugging a new circuit and its layout, as well as evaluating the coupling effects of the 60Hz component and its harmonics. As shown in *Figure 12*, the noise level is quite low and the influence of the power supply is relatively small. The highest 60Hz components reach −105dBV, while the noise floor sits around −120dBV.

BPA200 Noise Floor (dBV)
R_L = 8Ω, BW < 22 kHz, 9/16/97

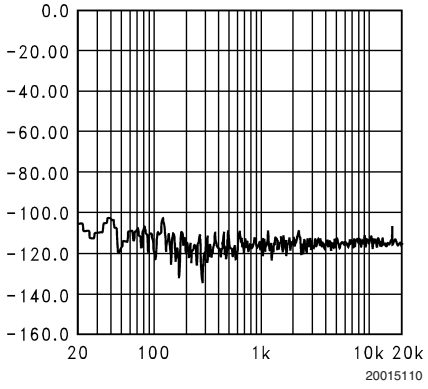


FIGURE 12. Log-Scale 60 Hz Noise Floor

Even with the limited number of graphs shown, the quality of this amplifier from a measurement perspective is quite good. However, with all audio equipment, nothing is really better than doing a listening test. It is recommended that listening test be done to confirm the audio quality of the different configurations presented.

7.0 BPA200–200W Bridged/Parallel Circuit (Continued)

7.1.4 Electrical Design Notes

The following electrical design notes will aid in obtaining a high performance amplifier solution.

- Input resistor values should be equal and have a 0.1% tolerance to minimize amplifier noise. These resistor pairs are R_{b1} & R_{i1}, R_{b2} & R_{i2}, R_{i3} & R_{c3}, and R_{i4} & R_{d3} as shown in Figure 13.
- 0.1% tolerance resistors for close gain matching should be used to minimize offset voltages when not using servo circuits.
- 1% tolerance high wattage ballast resistors are required when paralleling outputs to keep differently biased outputs from fighting each other. The wattage rating is dependent upon the amount of current expected to flow from each output and the resistance of the ballast resistor; $P = I^2R$. A good resistance value is 0.1Ω. Amplifier efficiency is lost if the resistor value is too large like 1Ω since 10A times 1Ω is 10V of output voltage drop.
- Low output offset voltage servo op amp is required to minimize solution output offset voltage.
- An input buffer is required because of the low input impedance from paralleling of the inputs. High valued gain setting resistors could be used at the risk of increasing noise susceptibility. The proposed solution with an input buffer (see Figures 13 & 17) provides the easiest way around gain resistor matching and providing a high input impedance.

7.0 BPA200–200W Bridged/Parallel Circuit (Continued)

7.2 SCHEMATICS

7.2.1 Detailed Bridged/Parallel Amplifier Schematic

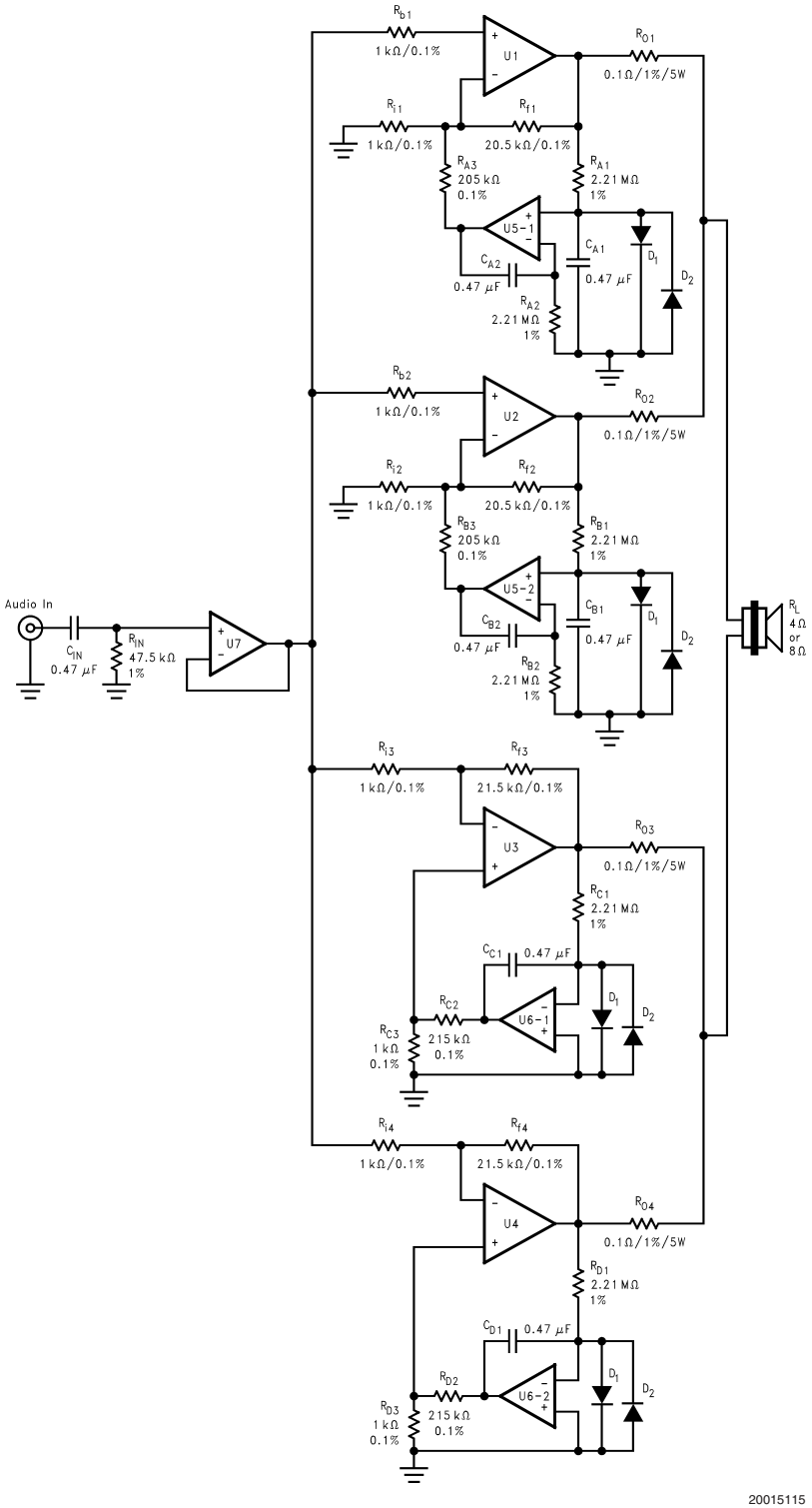


FIGURE 13. Detailed Bridged/Parallel Amplifier Schematic

7.0 BPA200–200W Bridged/Parallel Circuit (Continued)

7.2.2 Servo Circuits

While output ballast resistors in the basic bridge/parallel circuit work well to keep separately biased IC outputs from fighting each other, the addition of servo circuits will minimize output offset voltages that cause output voltage inequalities. Different output offset voltages cause a constant current to flow between outputs that increases IC power dissipation. By minimizing output offset voltages, all of the ICs will run cooler, expanding the IC's long term reliability and output power capability without activating sensitive protection circuits.

Typically, offset voltages are compensated for by using input and output coupling capacitors. Power amplifiers used in a single-supply configuration, utilize large value, large size electrolytic or polypropylene capacitors. This is because the load impedance is 4Ω or 8Ω and the RC combination creates a highpass filter that can rolloff audio frequencies. Since these output coupling capacitors have nonlinearities and are quite large, many designers choose to employ split power supplies. While split power supplies don't use these capacitors, a DC blocking capacitor is needed somewhere in the circuit to protect speakers. This capacitor is typically, C_{i1}, C_{i2}, C_{i3}, and C_{i4} as shown in *Figure 13*. With the application of a servo circuit, this capacitor can also be eliminated as shown in *Figures 14, 15*.

Servo circuits are essentially integrator op amp circuits that integrate offset voltage changes from the power op amp's output and feed back the integrated voltage to the opposite input of the power op amp. A servo circuit is required at each IC output of the bridge/parallel circuit to keep currents from flowing between IC outputs. Without each output compensated, one offset voltage will cause current to flow between ICs increasing power dissipation. If gain setting resistors are 0.1% and closely matched, the servo circuit may be left out, but DC blocking capacitors will be required.

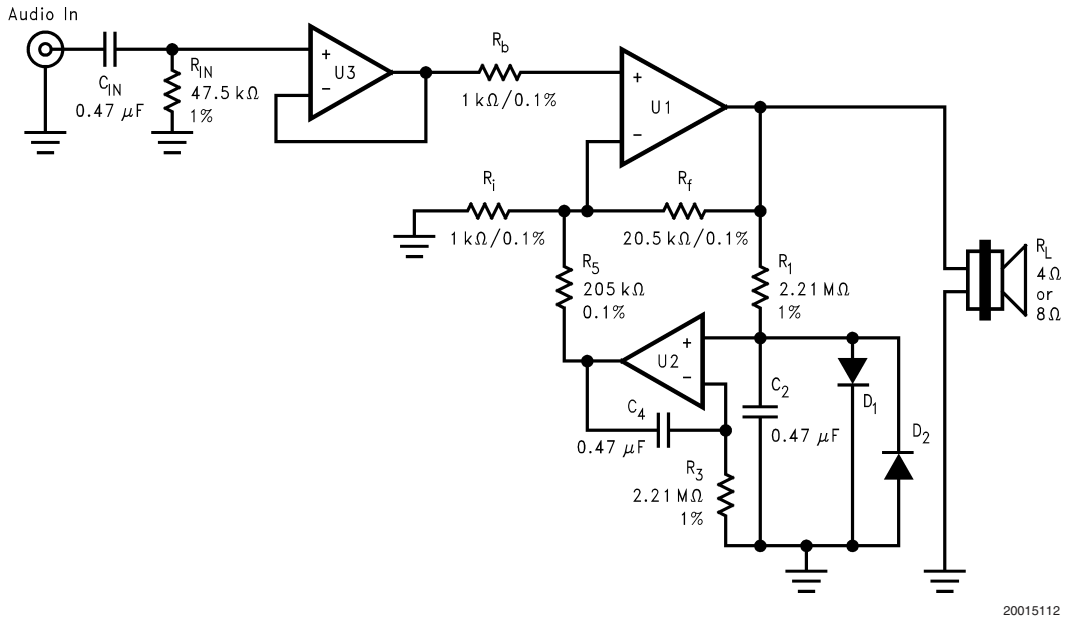


FIGURE 14. Non-Inverting Servo Amplifier Schematic

The inverting type servo amplifier applied to the inverting amplifier portion of the bridge/parallel circuit is shown in *Figure 15*. The non-inverting type servo circuit could be applied to the inverting input of U1 to achieve the same result, however, it uses an extra RC network that can be eliminated with the inverting type servo.

If a different power amplifier gain is desired, other component values can be used under the following conditions: In *Figure 14*, resistor R5 should be about 10 times the value of R_f, while R_i and R_b should be equal. In , *Figure 15*, resistor R3 should be about 10 times the value of R_f, while R_4 and R_i should be equal. For both the Non-Inverting and Inverting Servo solutions, the input clamping diodes should be low-leakage, with low-leakage film capacitors having a high-quality dielectric such as polypropylene or polystyrene (mylar), and metal-film resistors.

7.0 BPA200–200W Bridged/Parallel Circuit (Continued)

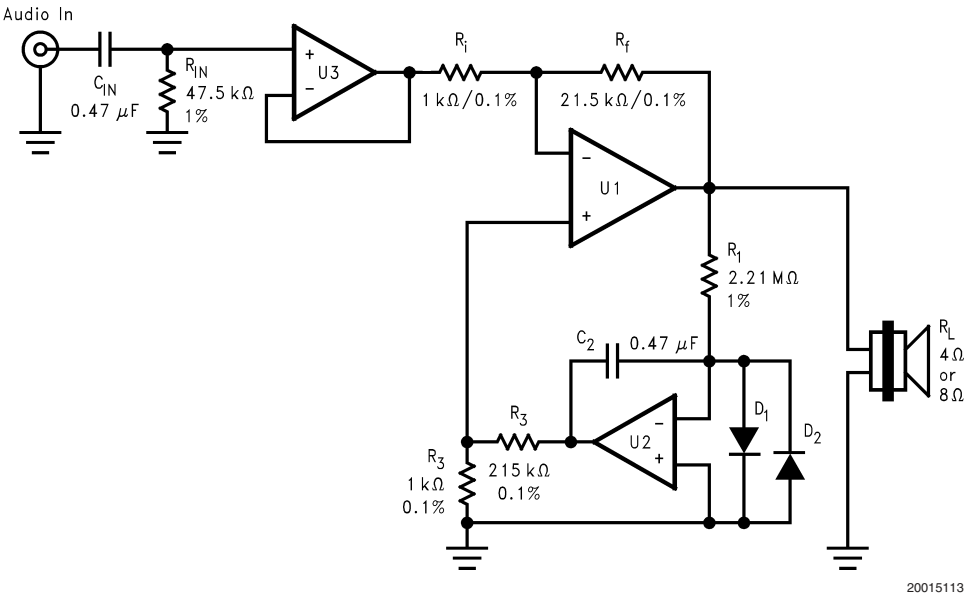


FIGURE 15. Inverting Servo Amplifier Schematic

7.2.3 Power Supply Circuit

The power supply portion of the amplifier is made up of a typical unregulated bipolar power supply. The supply is comprised of an input AC line filter, surge protecting MOVs, a separate 385VA toroidal transformer for each channel, and 40,000µF of supply reservoir capacitance for each supply voltage rail.

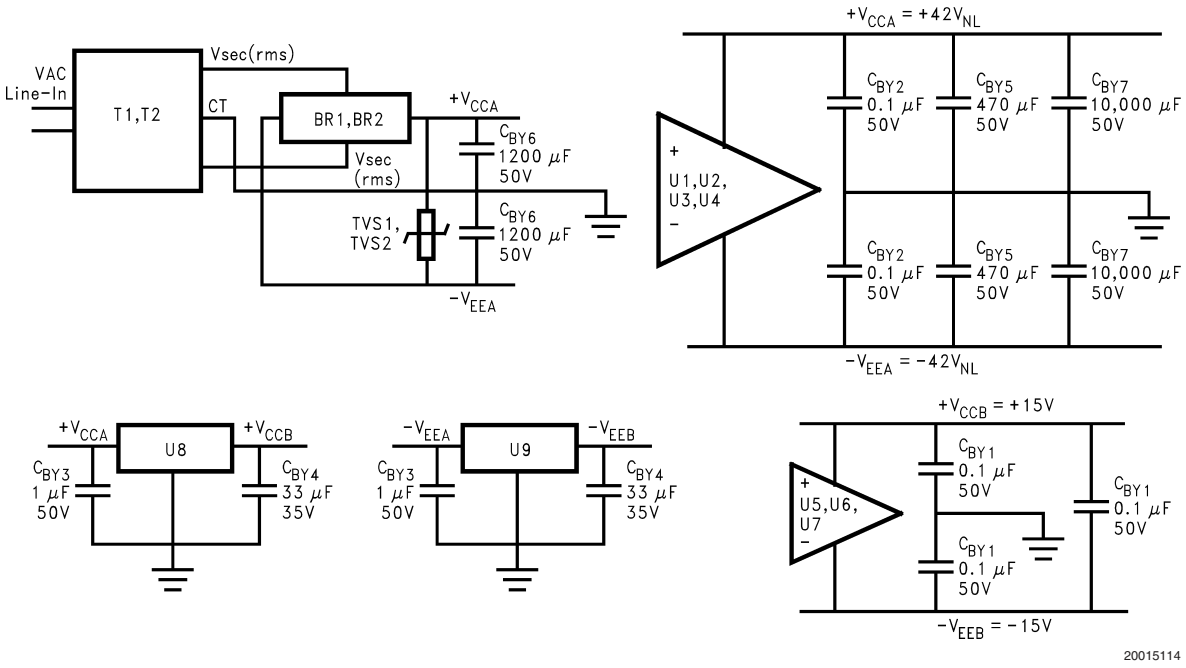


FIGURE 16. Power Supply Schematic

7.0 BPA200–200W Bridged/Parallel Circuit (Continued)

7.2.4 Basic Bridged/Parallel Amplifier Schematic

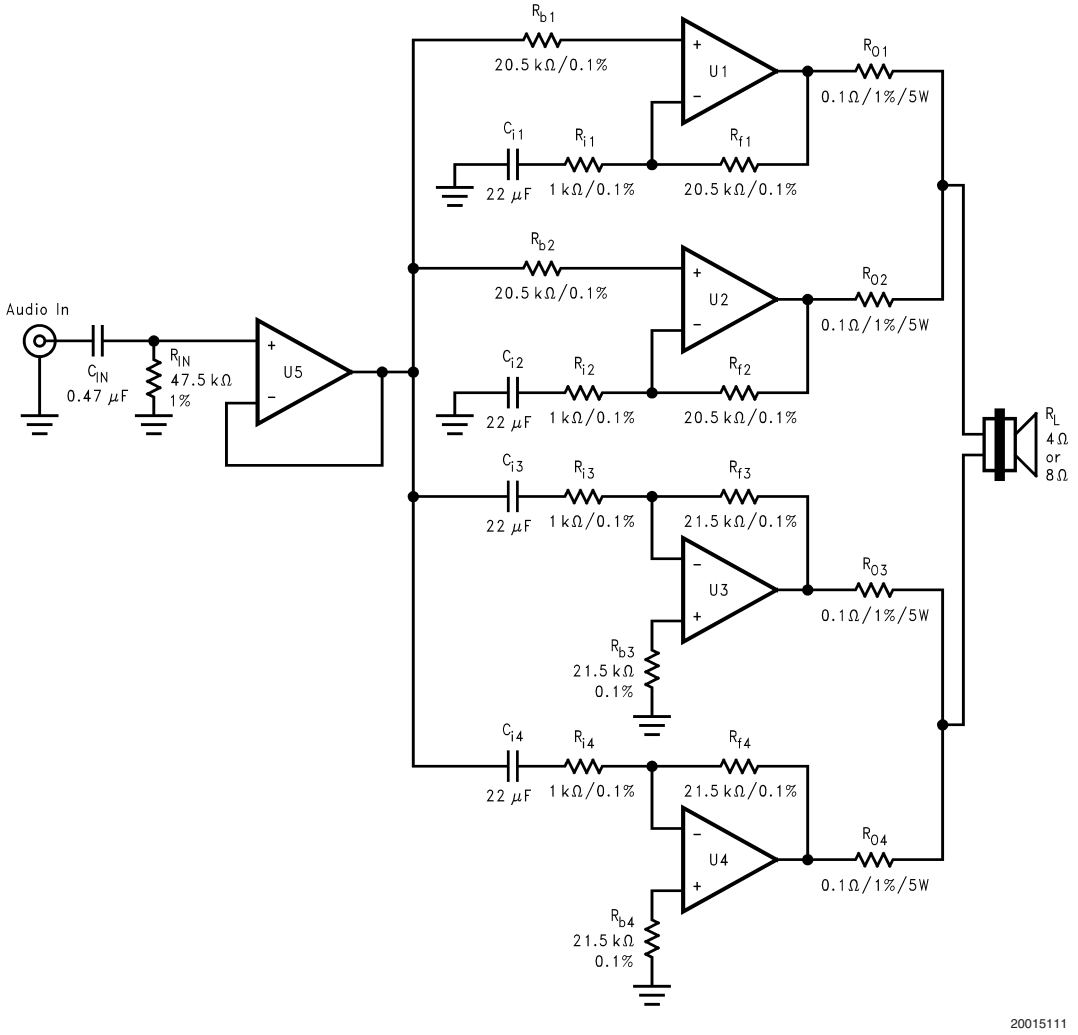


FIGURE 17. Basic Bridged/Parallel Amplifier Schematic

8.0 Parts List and Vendors

8.1 BUILD OF MATERIALS FOR BR100 AMPLIFIER

(See Figure 3 Bridged Amplifier Schematic)

Description	Designator	Manufacturer's or Example Part Number
1.0µF/100V Metallized Polyester Fim Capacitor	C_{IN}	Panasonic,ECQ-E1105KF
4.7µF/35V/Electrolytic Capacitor	C_{i1}, C_{i2}	Panasonic, ECE-A1VN4R7U
47kΩ/1/4W/5% Resistor	R_{IN}	
4.7kΩ/1/4W/1% Resistor	R_{i1}, R_{i2}	
46.4kΩ/1/4W/1% Resistor	R_{f1}	
51.1kΩ/1/4W/1% Resistor	R_{f2}	
3.32kΩ/1/4W/1% Resistor	R_B	
Additional Externals on Demo Board Not Shown on Schematic		
Description	**Designator**	**Functional Description and Example Part Number**
0.1µF/50V/Monolithic Ceramic Capacitor	C_{B1}	IC Supply Bypass Capacitor
47µF/50V/Electrolytic Capacitor	C_{B2}	IC Supply Bypass Capacitor Panasonic, EEU-FC1H470
4,700µF/50V/Electrolytic Capacitor	C_{B3}	IC Supply Bypass Capacitor Panasonic,ECO-S1HP472BA
10µF/35V/Electrolytic Capacitor	C_M	Turn on Mute
15kΩ/1/4W/5% Resistor	R_{M1}	Turn on Mute
8.2kΩ/1/4W/5% Resistor	R_{M2}	Turn on Mute
2.7Ω/1/4W/5% Resistor	R_G	Signal GND to Power GND

8.2 BUILD OF MATERIALS FOR PA100 AMPLIFIER

(See Figure 6 Parallel Amplifier Schematic)

Description	Designator	Manufacturer's or Example Part Number
1.0µF/100V Metallized Polyester Film Capacitor	C_{IN}	Panasonic, ECQ-E1105KF
68µF/50V Electrolytic Capacitor	C_i	Panasonic, EEU-FC1H680
47kΩ/1/4W/1% Resistor	R_{IN}	
1.0kΩ/1/4W/0.1% Resistor	R_i	
20.0kΩ/1/4W/0.1% Resistor	R_F	
1.0kΩ/1/4W/1% Resistor	R_B	
0.1Ω/1/3W/1% Resistor	R_{OUT}	
Additional Externals on Demo Board Not Shown on Schematic		
Description	**Designator**	**Functional Description and Example Part Number**
0.1µF/50V/Monolithic Ceramic Capacitor	C_1	IC Supply Bypass Capacitor
47µF/50V/Electrolytic Capacitor	C_2	IC Supply Bypass Capacitor Panasonic, EEU-FC1H470
2,200µF/50V/Electrolytic Capacitor	C_3	IC Supply Bypass Capacitor Panasonic,EEU-FC1H222
0.1µF/50V/Monolithic Ceramic Capacitor	C_{SN}	Snubber Network on Output
2.7Ω/1/4W/5% Resistor	R_{SN}	Snubber Network on Output
LM340T5, Fixed +5V Regulator	LM340L-5	5V PCB Supply
10µF/25V/Electrolytic Capacitor	C_{M1}	5V Output Bypass Capacitor
0.1µF/50V/Monolithic Ceramic Capacitor	C_{M2}	5V Input Bypass Capacitor

8.0 Parts List and Vendors (Continued)

Description	Designator	Manufacturer's or Example Part Number
470Ω/1/2W/5% Resistor	R_{REG}	Voltage Reducer for LM340 Input Voltage
20kΩ/1/4W/5% Resistor	R_{M1}	Mute Circuit
1MΩ/1/4W/5% Resistor	R_{M2}	Mute Circuit
120Ω/1/4W/5% Resistor	R_{LED}	Current Limit for LED Indicators
2.7Ω/1/4W/5% Resistor	R_G	Signal GND to Power GND
Ultra Bright LED Lamp, T-1 ¾ Standard Size, Green	LED	Indicator LED

8.0 Parts List and Vendors (Continued)

8.3 BUILD OF MATERIALS FOR BPA200 AMPLIFIER

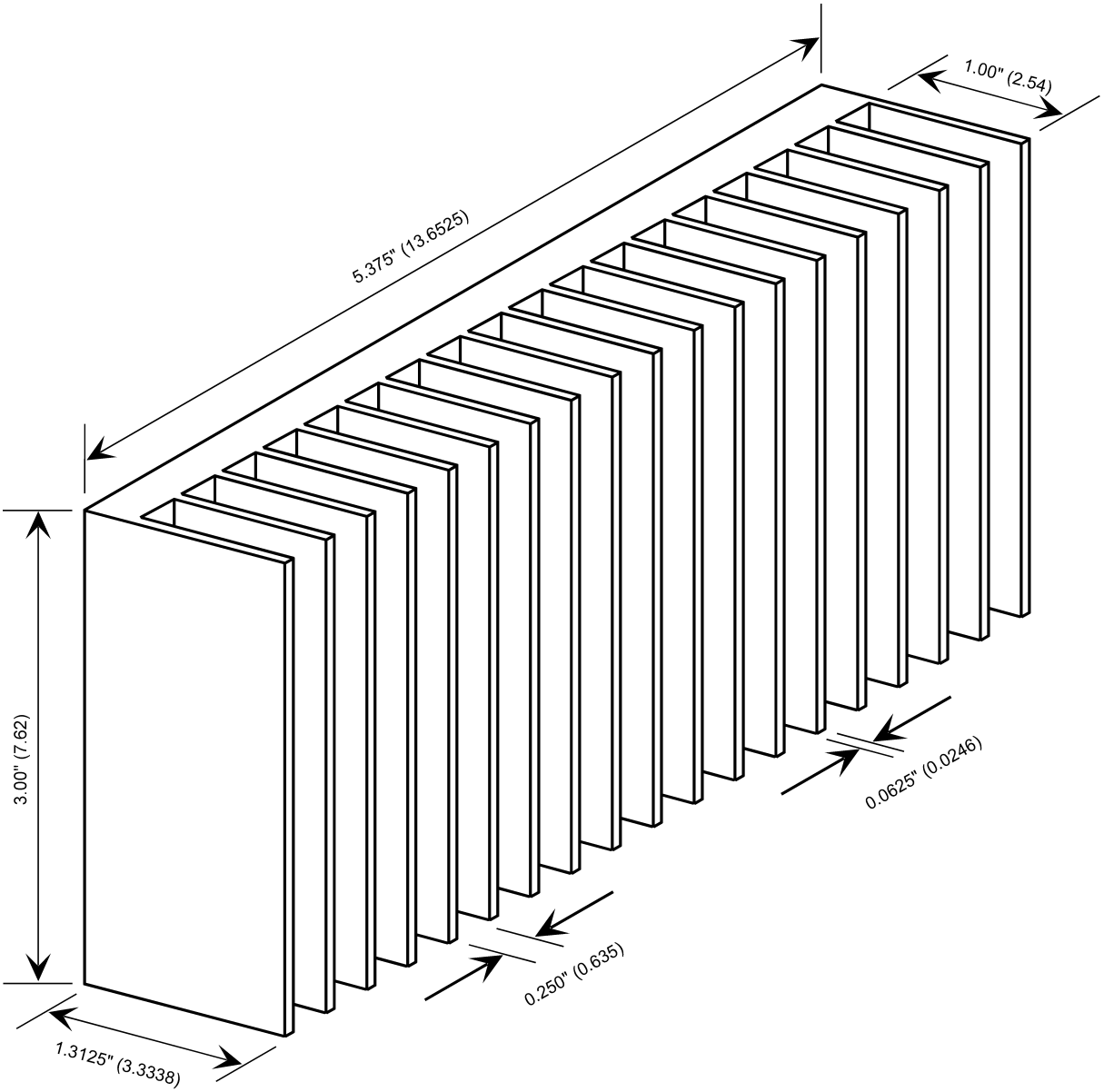
(See Figures 13 and 16, Detailed Bridged/Paralleled Amplifier Schematic and Power Supply Schematic)

Description	Designator	Manufacturer's Part Number
PASSIVE COMPONENTS		
0.47µF/100V Mylar Capacitor	C_{A1}, C_{A2}, C_{B1}, C_{B2}, C_{C1}, C_{D1}, C_{IN}	Electrocube, 230B-0.47µF-100V-JB
0.1Ω/5W/1% Power Ballast Resistor	R_{O1}, R_{O2}, R_{O3}, R_{O4}	Dale, RS-5-0.1-1%
1kΩ/0.1% Metal Film Resistor	R_{b1}, R_{b2}, R_{i1}, R_{i2}, R_{i3}, R_{i4}, R_{C3}, R_{D3}	Dale, RN-55D-1000-B
47kΩ/1% Metal Film Resistor	R_{in}	Dale, RN-55D-4702
20.5kΩ/0.1% Metal Film Resistor	R_{f1}, R_{f2}	Dale, CMF-55-20.5k-.1%-T2
21.5kΩ/0.1% Metal Film Resistor	R_{f3}, R_{f4}	Dale, MF-55-21.5k-.1%-T2
205kΩ/0.1% Metal Film Resistor	R_{A3}, R_{B3}	Dale, CMF-55-205k-.1%-T2
215kΩ/0.1% Metal Film Resistor	R_{C2}, R_{D2}	Dale, CMF-55-215k-.1%-T2
2.21MΩ/1% Metal Film Resistor	R_{A1}, R_{A2}, R_{B1}, R_{B2}, R_{C1}, R_{D1}	Dale, CMF-55-2.21M-.5%-T9
1N456A Low Leakage Diodes	D_1, D_2	National Semiconductor (NSC) 1N456A
INTEGRATED COMPONENTS		
LM3886T, 50W Monolithic Power IC	U1, U2, U3, U4	NSC, LM3886T
LF412ACN, Dual JFET Input Op Amp	U5, U6	NSC, LF412ACN
LF411ACN, JFET Input Op Amp	U7	NSC, LF411ACN
LM78L15ACZ, +15V Linear Regulator	U8	NSC, LM78L15ACZ
LM79L15ACZ, −15V Linear Regulator	U9	NSC, LM79L15ACZ
POWER SUPPLY COMPONENTS		
385V A, 60 V_{rms} Sec. Transformer	T1, T2	Toroid Corp. of Maryland, #738.302
Bridge Rectifier	B_{R1}, B_{R2}	General Instrument, KBU8B
100V, 1.5kΩ Metal Oxide Varistor (Transient Voltage Surpressor)	TVS1, TVS2	Digikey, 1.5KE100CACT-ND
0.1µF/50V Ceramic Capacitor	C_{BY1}	Sprague, 1C25Z5U104M050B
0.1µF/50V Polypropylene	C_{BY2}	Panasonic, ECQ-P1H104GZ
1µF/50V Electrolytic Capacitor	C_{BY3}	
33µF/35V Electrolytic Capacitor	C_{BY4}	
470µF/50V Electrolytic Capacitor	C_{BY5}	Mallory, SKR471M1HJ21V
1200µF/50V Electrolytic Capacitor	C_{BY6}	Mallory, LP122M050A1P3
10,000µF/50V Electrolytic Capacitor	C_{BY7}	Panasonic, ECE-S1HU103U
AC Line Connector		Schurter, 34.3124
Power Switch (Bowden Cable)		Schurter, 886.0101
MISCELLANEOUS HARDWARE		
IC 11-Pin Sockets		Yamaichi, SMT-15420

9.0 Heat Sink Drawings

9.1 BR100 AND PA100 HEAT SINK DRAWING

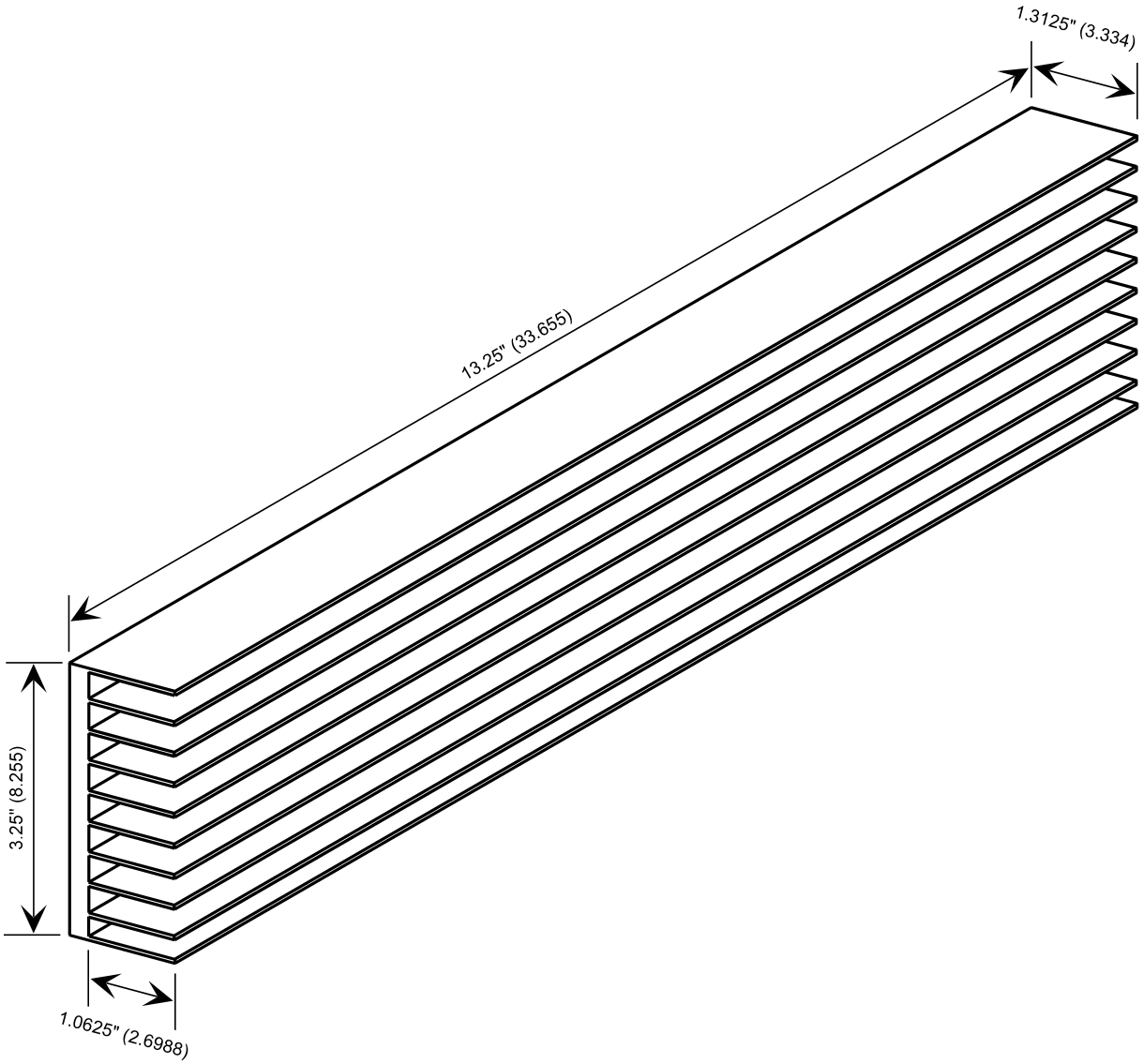
Inches (centimeters) unless otherwise noted



9.0 Heat Sink Drawings (Continued)

9.2 BPA200 HEAT SINK DRAWING

Inches (centimeters) unless otherwise noted



LIFE SUPPORT POLICY

NATIONAL'S PRODUCTS ARE NOT AUTHORIZED FOR USE AS CRITICAL COMPONENTS IN LIFE SUPPORT DEVICES OR SYSTEMS WITHOUT THE EXPRESS WRITTEN APPROVAL OF THE PRESIDENT AND GENERAL COUNSEL OF NATIONAL SEMICONDUCTOR CORPORATION. As used herein:

1. Life support devices or systems are devices or systems which, (a) are intended for surgical implant into the body, or (b) support or sustain life, and whose failure to perform when properly used in accordance with instructions for use provided in the labeling, can be reasonably expected to result in a significant injury to the user.
2. A critical component is any component of a life support device or system whose failure to perform can be reasonably expected to cause the failure of the life support device or system, or to affect its safety or effectiveness.

BANNED SUBSTANCE COMPLIANCE

National Semiconductor certifies that the products and packing materials meet the provisions of the Customer Products Stewardship Specification (CSP-9-111C2) and the Banned Substances and Materials of Interest Specification (CSP-9-111S2) and contain no ''Banned Substances'' as defined in CSP-9-111S2.

National Semiconductor
Americas Customer
Support Center
Email: new.feedback@nsc.com
Tel: 1-800-272-9959

www.national.com

National Semiconductor
Europe Customer Support Center
Fax: +49 (0) 180-530 85 86
Email: europe.support@nsc.com
Deutsch Tel: +49 (0) 69 9508 6208
English Tel: +44 (0) 870 24 0 2171
Français Tel: +33 (0) 1 41 91 8790

National Semiconductor
Asia Pacific Customer
Support Center
Email: ap.support@nsc.com

National Semiconductor
Japan Customer Support Center
Fax: 81-3-5639-7507
Email: jpn.feedback@nsc.com
Tel: 81-3-5639-7560

National does not assume any responsibility for use of any circuitry described, no circuit patent licenses are implied and National reserves the right at any time without notice to change said circuitry and specifications.

Embedded Secure Document

The file *http://sound.westhost.com/an166.pdf* is a secure document that has been embedded in this document. Double click the pushpin to view.

Texas Instruments

Op Amps For Everyone

Ron Mancini, Editor in Chief

Design Reference

August 2002 **Advanced Analog Products**

SLOD006B

IMPORTANT NOTICE

Texas Instruments Incorporated and its subsidiaries (TI) reserve the right to make corrections, modifications, enhancements, improvements, and other changes to its products and services at any time and to discontinue any product or service without notice. Customers should obtain the latest relevant information before placing orders and should verify that such information is current and complete. All products are sold subject to TI's terms and conditions of sale supplied at the time of order acknowledgment.

TI warrants performance of its hardware products to the specifications applicable at the time of sale in accordance with TI's standard warranty. Testing and other quality control techniques are used to the extent TI deems necessary to support this warranty. Except where mandated by government requirements, testing of all parameters of each product is not necessarily performed.

TI assumes no liability for applications assistance or customer product design. Customers are responsible for their products and applications using TI components. To minimize the risks associated with customer products and applications, customers should provide adequate design and operating safeguards.

TI does not warrant or represent that any license, either express or implied, is granted under any TI patent right, copyright, mask work right, or other TI intellectual property right relating to any combination, machine, or process in which TI products or services are used. Information published by TI regarding third party products or services does not constitute a license from TI to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property of that third party, or a license from TI under the patents or other intellectual property of TI.

Reproduction of information in TI data books or data sheets is permissible only if reproduction is without alteration and is accompanied by all associated warranties, conditions, limitations, and notices. Reproduction of this information with alteration is an unfair and deceptive business practice. TI is not responsible or liable for such altered documentation.

Resale of TI products or services with statements different from or beyond the parameters stated by TI for that product or service voids all express and any implied warranties for the associated TI product or service and is an unfair and deceptive business practice. TI is not responsible or liable for any such statements.

Mailing Address:

Texas Instruments
Post Office Box 655303
Dallas, Texas 75265

Copyright © 2002, Texas Instruments Incorporated

Forward

Everyone interested in analog electronics should find some value in this book, and an effort has been made to make the material understandable to the relative novice while not too boring for the practicing engineer. Special effort has been taken to ensure that each chapter can stand alone for the reader with the proper background. Of course, this causes redundancy that some people might find boring, but it's worth the price to enable the satisfaction of a diversified audience.

Start at Chapter 1 if you are a novice, and read through until completion of Chapter 9. After Chapter 9 is completed, the reader can jump to any chapter and be confident that they are prepared for the material. More experienced people such as electronic technicians, digital engineers, and non-electronic engineers can start at Chapter 3 and read through Chapter 9. Senior electronic technicians, electronic engineers, and fledgling analog engineers can start anywhere they feel comfortable and read through Chapter 9. Experienced analog engineers should jump to the subject that interests them. Analog gurus should send their additions, corrections, and complaints to me, and if they see something that looks familiar, they should feel complimented that others appreciate their contributions.

Chapter 1 is a history and story chapter. It is not required reading for anyone, but it defines the op amp's place in the world of analog electronics. Chapter 2 reviews some basic physics and develops the fundamental circuit equations that are used throughout the book. Similar equations have been developed in other books, but the presentation here emphasizes material required for speedy op amp design. The ideal op amp equations are developed in Chapter 3, and this chapter enables the reader to rapidly compute op amp transfer equations including ac response. The emphasis on single power supply systems forces the designer to bias circuits when the inputs are referenced to ground, and Chapter 4 gives a detailed procedure that quickly yields a working solution every time.

Op amps can't exist without feedback, and feedback has inherent stability problems, so feedback and stability are covered in Chapter 5. Chapters 6 and 7 develop the voltage feedback op amp equations, and they teach the concept of relative stability and compensation of potentially unstable op amps. Chapter 8 develops the current feedback op amp equations and discusses current feedback stability. Chapter 9 compares current feedback and voltage feedback op amps. The meat of this book is Chapters 12, 13, and 14 where the reader is shown how design the converter to transducer/actuator interface with the aid of op amps.

The remaining chapters give support material for Chapters 12, 13, and 14. Chapter 18 was a late addition. Portable applications are expanding rapidly and they emphasize the need for low-voltage/low-power design techniques. Chapter 18 defines some parameters in a new way so they lend themselves to low voltage design, and it takes the reader through several low voltage designs.

Thanks to editor James Karki for his contribution. We never gave him enough time to do detailed editing, so if you find errors or typos, direct them to my attention. Thanks to Ted Thomas, a marketing manager with courage enough to support a book, and big thanks for Alun Roberts who paid for this effort. Thomas Kugelstadt, applications manager, thanks for your support and help.

Also many thanks to the contributing authors, James Karki, Richard Palmer, Thomas Kugelstadt, Perry Miller, Bruce Carter, and Richard Cesari who gave generously of their time.

Regards,

Ron Mancini
Chief Editor

Contents

1 The Op Amp's Place In The World 1-1

2 Review of Circuit Theory 2-1
2.1 Introduction 2-1
2.2 Laws of Physics 2-1
2.3 Voltage Divider Rule 2-3
2.4 Current Divider Rule 2-4
2.5 Thevenin's Theorem 2-5
2.6 Superposition 2-8
2.7 Calculation of a Saturated Transistor Circuit 2-9
2.8 Transistor Amplifier 2-10

3 Development of the Ideal Op Amp Equations 3-1
3.1 Ideal Op Amp Assumptions 3-1
3.2 The Noninverting Op Amp 3-3
3.3 The Inverting Op Amp 3-4
3.4 The Adder 3-5
3.5 The Differential Amplifier 3-6
3.6 Complex Feedback Networks 3-7
3.7 Video Amplifiers 3-9
3.8 Capacitors 3-9
3.9 Summary 3-11

4 Single Supply Op Amp Design Techniques 4-1
4.1 Single Supply versus Dual Supply 4-1
4.2 Circuit Analysis 4-3
4.3 Simultaneous Equations 4-8
4.3.1 Case 1: VOUT = +mVIN+b 4-9
4.3.2 Case 2: VOUT = +mVIN – b 4-13
4.3.3 Case 3: VOUT = –mVIN + b 4-16
4.3.4 Case 4: VOUT = –mVIN – b 4-19
4.4 Summary 4-22

5 Feedback and Stability Theory 5-1
5.1 Why Study Feedback Theory? 5-1
5.2 Block Diagram Math and Manipulations 5-1
5.3 Feedback Equation and Stability 5-6

5.4 Bode Analysis of Feedback Circuits 5-7
5.5 Loop Gain Plots are the Key to Understanding Stability 5-12
5.6 The Second Order Equation and Ringing/Overshoot Predictions 5-15
5.7 References 5-16

6 Development of the Non Ideal Op Amp Equations 6-1
6.1 Introduction 6-1
6.2 Review of the Canonical Equations 6-2
6.3 Noninverting Op Amps 6-5
6.4 Inverting Op Amps 6-6
6.5 Differential Op Amps 6-8

7 Voltage-Feedback Op Amp Compensation 7-1
7.1 Introduction 7-1
7.2 Internal Compensation 7-2
7.3 External Compensation, Stability, and Performance 7-8
7.4 Dominant-Pole Compensation 7-9
7.5 Gain Compensation 7-12
7.6 Lead Compensation 7-13
7.7 Compensated Attenuator Applied to Op Amp 7-16
7.8 Lead-Lag Compensation 7-18
7.9 Comparison of Compensation Schemes 7-20
7.10 Conclusions 7-21

8 Current-Feedback Op Amp Analysis 8-1
8.1 Introduction 8-1
8.2 CFA Model 8-1
8.3 Development of the Stability Equation 8-2
8.4 The Noninverting CFA 8-3
8.5 The Inverting CFA 8-5
8.6 Stability Analysis 8-7
8.7 Selection of the Feedback Resistor 8-9
8.8 Stability and Input Capacitance 8-11
8.9 Stability and Feedback Capacitance 8-12
8.10 Compensation of CF and CG 8-13
8.11 Summary 8-14

9 Voltage- and Current-Feedback Op Amp Comparison 9-1
9.1 Introduction 9-1
9.2 Precision 9-2
9.3 Bandwidth 9-3
9.4 Stability 9-6
9.5 Impedance 9-7
9.6 Equation Comparison 9-8

10 Op Amp Noise Theory and Applications 10-1
10.1 Introduction 10-1
10.2 Characterization 10-1
10.2.1 rms versus P-P Noise 10-1
10.2.2 Noise Floor 10-3
10.2.3 Signal-to-Noise Ratio 10-3
10.2.4 Multiple Noise Sources 10-3
10.2.5 Noise Units 10-4
10.3 Types of Noise 10-4
10.3.1 Shot Noise 10-5
10.3.2 Thermal Noise 10-7
10.3.3 Flicker Noise 10-8
10.3.4 Burst Noise 10-9
10.3.5 Avalanche Noise 10-9
10.4 Noise Colors 10-10
10.4.1 White Noise 10-11
10.4.2 Pink Noise 10-11
10.4.3 Red/Brown Noise 10-12
10.5 Op Amp Noise 10-12
10.5.1 The Noise Corner Frequency and Total Noise 10-12
10.5.2 The Corner Frequency 10-13
10.5.3 Op Amp Circuit Noise Model 10-14
10.5.4 Inverting Op Amp Circuit Noise 10-16
10.5.5 Noninverting Op Amp Circuit Noise 10-17
10.5.6 Differential Op Amp Circuit Noise 10-18
10.5.7 Summary 10-18
10.6 Putting It All Together 10-19
10.7 References 10-23

11 Understanding Op Amp Parameters 11-1
11.1 Introduction 11-1
11.2 Operational Amplifier Parameter Glossary 11-2
11.3 Additional Parameter Information 11-8
11.3.1 Input Offset Voltage 11-8
11.3.2 Input Current 11-10
11.3.3 Input Common Mode Voltage Range 11-11
11.3.4 Differential Input Voltage Range 11-11
11.3.5 Maximum Output Voltage Swing 11-12
11.3.6 Large Signal Differential Voltage Amplification 11-13
11.3.7 Input Parasitic Elements 11-13
11.3.8 Output Impedance 11-14
11.3.9 Common-Mode Rejection Ratio 11-15
11.3.10 Supply Voltage Rejection Ratio 11-15
11.3.11 Supply Current 11-16

11.3.12 Slew Rate at Unity Gain 11-16
11.3.13 Equivalent Input Noise 11-17
11.3.14 Total Harmonic Distortion Plus Noise 11-18
11.3.15 Unity Gain Bandwidth and Phase Margin 11-19
11.3.16 Settling Time 11-22

12 Instrumentation: Sensors to A/D Converters 12-1
12.1 Introduction 12-1
12.2 Transducer Types 12-6
12.3 Design Procedure 12-11
12.4 Review of the System Specifications 12-12
12.5 Reference Voltage Characterization 12-12
12.6 Transducer Characterization 12-13
12.7 ADC Characterization 12-15
12.8 Op Amp Selection 12-15
12.9 Amplifier Circuit Design 12-16
12.10 Test 12-23
12.11 Summary 12-23
12.12 References 12-23

13 Wireless Communication: Signal Conditioning for IF Sampling 13-1
13.1 Introduction 13-1
13.2 Wireless Systems 13-1
13.3 Selection of ADCs/DACs 13-6
13.4 Factors Influencing the Choice of Op Amps 13-10
13.5 Anti-Aliasing Filters 13-11
13.6 Communication D/A Converter Reconstruction Filter 13-13
13.7 External Vref Circuits for ADCs/DACs 13-15
13.8 High-Speed Analog Input Drive Circuits 13-18
13.9 References 13-22

14 Interfacing D/A Converters to Loads 14-1
14.1 Introduction 14-1
14.2 Load Characteristics 14-1
14.2.1 DC Loads 14-1
14.2.2 AC Loads 14-2
14.3 Understanding the D/A Converter and its Specifications 14-2
14.3.1 Types of D/A Converters — Understanding the Tradeoffs 14-2
14.3.2 The Resistor Ladder D/A Converter 14-2
14.3.3 The Weighted Resistor D/A Converter 14-3
14.3.4 The R/2R D/A Converter 14-4
14.3.5 The Sigma Delta D/A Converter 14-5
14.4 D/A Converter Error Budget 14-6
14.4.1 Accuracy versus Resolution 14-7
14.4.2 DC Application Error Budget 14-7

14.4.3 AC Application Error Budget 14-8
14.4.4 RF Application Error Budget 14-10
14.5 D/A Converter Errors and Parameters 14-10
14.5.1 DC Errors and Parameters 14-10
14.5.2 AC Application Errors and Parameters 14-14
14.6 Compensating For DAC Capacitance 14-18
14.7 Increasing Op Amp Buffer Amplifier Current and Voltage 14-19
14.7.1 Current Boosters 14-20
14.7.2 Voltage Boosters 14-20
14.7.3 Power Boosters 14-22
14.7.4 Single-Supply Operation and DC Offsets 14-22

15 Sine Wave Oscillators 15-1
15.1 What is a Sine Wave Oscillator? 15-1
15.2 Requirements for Oscillation 15-1
15.3 Phase Shift in the Oscillator 15-3
15.4 Gain in the Oscillator 15-4
15.5 Active Element (Op Amp) Impact on the Oscillator 15-5
15.6 Analysis of the Oscillator Operation (Circuit) 15-7
15.7 Sine Wave Oscillator Circuits 15-9
15.7.1 Wien Bridge Oscillator 15-9
15.7.2 Phase Shift Oscillator, Single Amplifier 15-14
15.7.3 Phase Shift Oscillator, Buffered 15-15
15.7.4 Bubba Oscillator 15-17
15.7.5 Quadrature Oscillator 15-18
15.7.6 Conclusion 15-20
15.8 References 15-21

16 Active Filter Design Techniques 16-1
16.1 Introduction 16-1
16.2 Fundamentals of Low-Pass Filters 16-2
16.2.1 Butterworth Low-Pass FIlters 16-6
16.2.2 Tschebyscheff Low-Pass Filters 16-7
16.2.3 Bessel Low-Pass Filters 16-7
16.2.4 Quality Factor Q 16-9
16.2.5 Summary 16-10
16.3 Low-Pass Filter Design 16-11
16.3.1 First-Order Low-Pass Filter 16-12
16.3.2 Second-Order Low-Pass Filter 16-14
16.3.3 Higher-Order Low-Pass Filters 16-19
16.4 High-Pass Filter Design 16-21
16.4.1 First-Order High-Pass Filter 16-23
16.4.2 Second-Order High-Pass Filter 16-24
16.4.3 Higher-Order High-Pass Filter 16-26

16.5 Band-Pass Filter Design 16-27
16.5.1 Second-Order Band-Pass Filter 16-29
16.5.2 Fourth-Order Band-Pass Filter (Staggered Tuning) 16-32
16.6 Band-Rejection Filter Design 16-36
16.6.1 Active Twin-T Filter 16-37
16.6.2 Active Wien-Robinson Filter 16-39
16.7 All-Pass Filter Design 16-41
16.7.1 First-Order All-Pass Filter 16-44
16.7.2 Second-Order All-Pass Filter 16-44
16.7.3 Higher-Order All-Pass Filter 16-45
16.8 Practical Design Hints 16-47
16.8.1 Filter Circuit Biasing 16-47
16.8.2 Capacitor Selection 16-50
16.8.3 Component Values 16-52
16.8.4 Op Amp Selection 16-53
16.9 Filter Coefficient Tables 16-55
16.10 References 16-63

17 Circuit Board Layout Techniques 17-1
17.1 General Considerations 17-1
17.1.1 The PCB is a Component of the Op Amp Design 17-1
17.1.2 Prototype, Prototype, PROTOTYPE! 17-1
17.1.3 Noise Sources 17-2
17.2 PCB Mechanical Construction 17-3
17.2.1 Materials — Choosing the Right One for the Application 17-3
17.2.2 How Many Layers are Best? 17-4
17.2.3 Board Stack-Up — The Order of Layers 17-6
17.3 Grounding 17-7
17.3.1 The Most Important Rule: Keep Grounds Separate 17-7
17.3.2 Other Ground Rules 17-7
17.3.3 A Good Example 17-9
17.3.4 A Notable Exception 17-10
17.4 The Frequency Characteristics of Passive Components 17-11
17.4.1 Resistors 17-11
17.4.2 Capacitors 17-12
17.4.3 Inductors 17-13
17.4.4 Unexpected PCB Passive Components 17-14
17.5 Decoupling 17-20
17.5.1 Digital Circuitry — A Major Problem for Analog Circuitry 17-20
17.5.2 Choosing the Right Capacitor 17-21
17.5.3 Decoupling at the IC Level 17-22
17.5.4 Decoupling at the Board Level 17-23
17.6 Input and Output Isolation 17-23
17.7 Packages 17-24

17.7.1 Through-Hole Considerations 17-26
17.7.2 Surface Mount 17-27
17.7.3 Unused Sections 17-27
17.8 Summary 17-28
17.8.1 General 17-28
17.8.2 Board Structure 17-28
17.8.3 Components 17-28
17.8.4 Routing 17-29
17.8.5 Bypass 17-29
17.9 References 17-29

18 Designing Low-Voltage Op Amp Circuits 18-1
18.1 Introduction 18-1
18.2 Dynamic Range 18-3
18.3 Signal-to-Noise Ratio 18-5
18.4 Input Common-Mode Range 18-6
18.5 Output Voltage Swing 18-11
18.6 Shutdown and Low Current Drain 18-12
18.7 Single-Supply Circuit Design 18-13
18.8 Transducer to ADC Analog Interface 18-13
18.9 DAC to Actuator Analog Interface 18-16
18.10 Comparison of Op Amps 18-20
18.11 Summary 18-22

A Single-Supply Circuit Collection A-1
A.1 Introduction A-1
A.2 Boundary Conditions A-4
A.3 Amplifiers A-5
A.3.1 Inverting Op Amp with Noninverting Positive Reference A-5
A.3.2 Inverting Op Amp with Inverting Negative Reference A-6
A.3.3 Inverting Op Amp with Noninverting Negative Reference A-7
A.3.4 Inverting Op Amp with Inverting Positive Reference A-8
A.3.5 Noninverting Op Amp with Inverting Positive Reference A-9
A.3.6 Noninverting Op Amp with Noninverting Negative Reference A-10
A.3.7 Noninverting Op Amp with Inverting Negative Reference A-11
A.3.8 Noninverting Op Amp with Noninverting Positive Reference A-12
A.3.9 Differential Amplifier A-13
A.3.10 Differential Amplifier With Bias Correction A-14
A.3.11 High Input Impedance Differential Amplifier A-15
A.3.12 High Common-Mode Range Differential Amplifier A-16
A.3.13 High-Precision Differential Amplifier A-17
A.3.14 Simplified High-Precision Differential Amplifier A-18
A.3.15 Variable Gain Differential Amplifier A-19
A.3.16 T Network in the Feedback Loop A-20

A.3.17 Buffer A-21
A.3.18 Inverting AC Amplifier A-22
A.3.19 Noninverting AC Amplifier A-23
A.4 Computing Circuits A-24
A.4.1 Inverting Summer A-24
A.4.2 Noninverting Summer A-25
A.4.3 Noninverting Summer with Buffers A-26
A.4.4 Inverting Integrator A-27
A.4.5 Inverting Integrator with Input Current Compensation A-28
A.4.6 Inverting Integrator with Drift Compensation A-29
A.4.7 Inverting Integrator with Mechanical Reset A-30
A.4.8 Inverting Integrator with Electronic Reset A-31
A.4.9 Inverting Integrator with Resistive Reset A-32
A.4.10 Noninverting Integrator with Inverting Buffer A-33
A.4.11 Noninverting Integrator Approximation A-34
A.4.12 Inverting Differentiator A-35
A.4.13 Inverting Differentiator with Noise Filter A-36
A.5 Oscillators A-37
A.5.1 Basic Wien Bridge Oscillator A-37
A.5.2 Wien Bridge Oscillator with Nonlinear Feedback A-38
A.5.3 Wien Bridge Oscillator with AGC A-39
A.5.4 Quadrature Oscillator A-40
A.5.5 Classical Phase Shift Oscillator A-41
A.5.6 Buffered Phase Shift Oscillator A-42
A.5.7 Bubba Oscillator A-43
A.5.8 Triangle Oscillator A-44

B Single-Supply Op Amp Selection Guide B-1

Figures

2–1 Ohm's Law Applied to the Total Circuit 2-2
2–2 Ohm's Law Applied to a Component 2-2
2–3 Kirchoff's Voltage Law 2-2
2–4 Kirchoff's Current Law 2-3
2–5 Voltage Divider Rule 2-3
2–6 Current Divider Rule 2-4
2–7 Original Circuit 2-5
2–8 Thevenin's Equivalent Circuit for Figure 2–7 2-5
2–9 Example of Thevenin's Equivalent Circuit 2-6
2–10 Analysis Done the Hard Way 2-7
2–11 Superposition Example 2-8
2–12 When V1 is Grounded 2-8
2–13 When V2 is Grounded 2-8
2–14 Saturated Transistor Circuit 2-9
2–15 Transistor Amplifier 2-10
2–16 Thevenin Equivalent of the Base Circuit 2-11

3–1 The Ideal Op Amp 3-2
3–2 The Noninverting Op Amp 3-3
3–3 The Inverting Op Amp 3-4
3–4 The Adder Circuit 3-5
3–5 The Differential Amplifier 3-6
3–6 Differential Amplifier With Common-Mode Input Signal 3-7
3–7 T Network in Feedback Loop 3-7
3–8 Thevenin's Theorem Applied to T Network 3-8
3–9 Video Amplifier 3-9
3–10 Low-Pass Filter 3-10
3–11 High-Pass Filter 3-10

4–1 Split-Supply Op Amp Circuit 4-1
4–2 Split-Supply Op Amp Circuit With Reference Voltage Input 4-2
4–3 Split-Supply Op Amp Circuit With Common-Mode Voltage 4-2
4–4 Single-Supply Op Amp Circuit 4-3
4–5 Inverting Op Amp 4-4
4–6 Inverting Op Amp With VCC Bias 4-5
4–7 Transfer Curve for Inverting Op Amp With VCC Bias 4-5
4–8 Noninverting Op Amp 4-7

4–9 Transfer Curve for Noninverting Op Amp 4-7
4–10 Schematic for Case1: VOUT = +mVIN + b 4-9
4–11 Case 1 Example Circuit 4-12
4–12 Case 1 Example Circuit Measured Transfer Curve 4-12
4–13 Schematic for Case 2: VOUT = +mVIN – b 4-13
4–14 Case 2 Example Circuit 4-15
4–15 Case 2 Example Circuit Measured Transfer Curve 4-15
4–16 Schematic for Case 3: VOUT = –mVIN + b 4-16
4–17 Case 3 Example Circuit 4-17
4–18 Case 3 Example Circuit Measured Transfer Curve 4-18
4–19 Schematic for Case 4: VOUT = –mVIN – b 4-19
4–20 Case 4 Example Circuit 4-20
4–21 Case 4 Example Circuit Measured Transfer Curve 4-21

5–1 Definition of Blocks 5-2
5–2 Summary Points 5-3
5–3 Definition of Control System Terms 5-3
5–4 Definition of an Electronic Feedback Circuit 5-3
5–5 Multiloop Feedback System 5-4
5–6 Block Diagram Transforms 5-5
5–7 Comparison of Control and Electronic Canonical Feedback Systems 5-6
5–8 Low-Pass Filter 5-8
5–9 Bode Plot of Low-Pass Filter Transfer Function 5-9
5–10 Band Reject Filter 5-9
5–11 Individual Pole Zero Plot of Band Reject Filter 5-10
5–12 Combined Pole Zero Plot of Band Reject Filter 5-10
5–13 When No Pole Exists in Equation (5–12) 5-11
5–14 When Equation 5–12 has a Single Pole 5-11
5–15 Magnitude and Phase Plot of Equation 5–14 5-13
5–16 Magnitude and Phase Plot of the Loop Gain Increased to (K+C) 5-14
5–17 Magnitude and Phase Plot of the Loop Gain With Pole Spacing Reduced 5-14
5–18 Phase Margin and Overshoot vs Damping Ratio 5-16

6–1 Feedback System Block Diagram 6-2
6–2 Feedback Loop Broken to Calculate Loop Gain 6-4
6–3 Noninverting Op Amp 6-5
6–4 Open Loop Noninverting Op Amp 6-6
6–5 Inverting Op Amp 6-6
6–6 Inverting Op Amp: Feedback Loop Broken for Loop Gain Calculation 6-7
6–7 Differential Amplifier Circuit 6-8

7–1 Miller Effect Compensation 7-2
7–2 TL03X Frequency and Time Response Plots 7-3
7–3 Phase Margin and Percent Overshoot Versus Damping Ratio 7-4
7–4 TL07X Frequency and Time Response Plots 7-5

7–5 TL08X Frequency and Time Response Plots 7-6
7–6 TLV277X Frequency Response Plots 7-7
7–7 TLV227X Time Response Plots 7-7
7–8 Capacitively-Loaded Op Amp 7-9
7–9 Capacitively-Loaded Op Amp With Loop Broken for Loop Gain (Aβ) Calculation 7-9
7–10 Possible Bode Plot of the Op Amp Described in Equation 7–7 7-11
7–11 Dominant-Pole Compensation Plot 7-11
7–12 Gain Compensation 7-12
7–13 Lead-Compensation Circuit 7-13
7–14 Lead-Compensation Bode Plot 7-14
7–15 Inverting Op Amp With Lead Compensation 7-15
7–16 Noninverting Op Amp With Lead Compensation 7-16
7–17 Op Amp With Stray Capacitance on the Inverting Input 7-16
7–18 Compensated Attenuator Circuit 7-17
7–19 Compensated Attenuator Bode Plot 7-18
7–20 Lead-Lag Compensated Op Amp 7-19
7–21 Bode Plot of Lead-Lag Compensated Op Amp 7-19
7–22 Closed-Loop Plot of Lead-Lag Compensated Op Amp 7-20

8–1 Current-Feedback Amplifier Model 8-2
8–2 Stability Analysis Circuit 8-2
8–3 Stability Analysis Circuit 8-3
8–4 Noninverting CFA 8-4
8–5 Inverting CFA 8-5
8–6 Bode Plot of Stability Equation 8-7
8–7 Plot of CFA RF, G, and BW 8-10
8–8 Effects of Stray Capacitance on CFAs 8-12
8–9 Bode Plot with CF 8-13

9–1 Long-Tailed Pair 9-2
9–2 Ideal CFA 9-3
9–3 VFA Gain versus Frequency 9-4
9–4 CFA Gain vs Frequency 9-5

10–1 Gaussian Distribution of Noise Energy 10-2
10–2 Shot Noise Generation 10-5
10–3 Thermal Noise 10-7
10–4 Avalanche Noise 10-10
10–5 Noise Colors 10-11
10–6 TLV2772 Op Amp Noise Characteristics 10-13
10–7 Op Amp Circuit Noise Model 10-15
10–8 Equivalent Op Amp Circuit Noise Model 10-15
10–9 Inverting Op Amp Circuit Noise Model 10-16
10–10 Inverting Equivalent Op Amp Circuit Noise Model 10-16
10–11 Noninverting Equivalent Op Amp Circuit Noise Model 10-17

10–12 Differential Equivalent Op Amp Circuit Noise Model 10-18
10–13 Split Supply Op Amp Circuit 10-19
10–14 TLC2201 Op Amp Noise Performance 10-20
10–15 TLC2201 Op Amp Circuit 10-21
10–16 Improved TLC2201 Op Amp Circuit 10-23

11–1 Test Circuits for Input Offset Voltage 11-9
11–2 Offset Voltage Adjust 11-9
11–3 Test Circuit – IIB 11-10
11–4 $V_{OM\pm}$ 11-12
11–5 Input Parasitic Elements 11-13
11–6 Effect of Output Impedance 11-14
11–7 Figure 6. Slew Rate 11-16
11–8 Figure 7. Simplified Op Amp Schematic 11-17
11–9 Typical Op amp Input Noise Spectrum 11-18
11–10 Output Spectrum with THD + N = 1% 11-19
11–11 Voltage Amplification and Phase Shift vs. Frequency 11-21
11–12 Settling Time 11-22

12–1 Block Diagram of a Transducer Measurement System 12-1
12–2 Significant Bits versus Binary Bits 12-2
12–3 Example of Spans That Require Correction 12-4
12–4 Voltage Divider Circuit for a Resistive Transducer 12-6
12–5 Current Source Excitation for a Resistive Transducer 12-7
12–6 Precision Current Source 12-7
12–7 Wheatstone Bridge Circuit 12-8
12–8 Photodiode Amplifier 12-8
12–9 Phototransistor Amplifier 12-9
12–10 Photovoltaic Cell Amplifier 12-9
12–11 Active Full-Wave Rectifier and Filter 12-10
12–12 Reference and Transducer Bias Circuit 12-13
12–13 AIA Circuit 12-17
12–14 Final Analog Interface Circuit 12-21

13–1 A Typical GSM Cellular Base Station Receiver Block Diagram 13-2
13–2 An Implementation of a Software-Configurable Dual-IF Receiver 13-4
13–3 Basic W-CDMA Cellular Base Station Transmitter Block 13-5
13–4 Communication DAC with Interpolation and Reconstruction Filters 13-5
13–5 QPSK Power Spectral Density Without Raised Cosine Filter — W-CDMA 13-14
13–6 Reconstruction Filter Characteristics 13-14
13–7 A Single-Pole Reconstruction Filter 13-15
13–8 Voltage Reference Filter Circuit 13-16
13–9 Voltage Follower Frequency Response Plot 13-17
13–10 External Voltage Reference Circuit for ADC/DAC. 13-17
13–11 Single-Ended to Differential Output Drive Circuit 13-18

13–12 Differential Amplifier Closed-Loop Response 13-19
13–13 ADC Single-Ended Input Drive Circuit 13-20
13–14 Gain vs Frequency Plot for THS3201 13-21
13–15 Phase vs Frequency for THS3201 13-21

14–1 Resistor Ladder D/A Converter 14-2
14–2 Binary Weighted D/A Converter 14-3
14–3 R/2R Resistor Array 14-4
14–4 R/2R D/A Converter 14-5
14–5 Sigma Delta D/A Converter 14-6
14–6 Total Harmonic Distortion 14-9
14–7 D/A Offset Error 14-11
14–8 D/A Gain Error 14-12
14–9 Differential Nonlinearity Error 14-13
14–10 Integral Nonlinearity Error 14-13
14–11 Spurious Free Dynamic Range 14-15
14–12 Intermodulation Distortion 14-16
14–13 D/A Settling Time 14-17
14–14 D/A Deglitch Circuit 14-17
14–15 Compensating for CMOS DAC Output Capacitance 14-18
14–16 D/A Output Current Booster 14-20
14–17 Incorrect Method of Increasing Voltage Swing of D/A Converters 14-21
14–18 Correct Method of Increasing Voltage Range 14-22
14–19 Single-Supply DAC Operation 14-23

15–1 Canonical Form of a Feedback System with Positive or Negative Feedback 15-2
15–2 Phase Plot of RC Sections 15-3
15–3 Op Amp Frequency Response 15-6
15–4 Op Amp Bandwidth and Oscillator Output 15-7
15–5 Block Diagram of an Oscillator with Positive Feedback 15-8
15–6 Amplifier with Positive and Negative Feedback 15-8
15–7 Wien Bridge Circuit Schematic 15-9
15–8 Final Wien Bridge Oscillator Circuit 15-11
15–9 Wien Bridge Output Waveforms 15-11
15–10 Wien Bridge Oscillator with Nonlinear Feedback 15-12
15–11 Output of the Circuit in Figure 15–10 15-12
15–12 Wien Bridge Oscillator with AGC 15-13
15–13 Output of the Circuit in Figure 15–12 15-14
15–14 Phase Shift Oscillator (Single Op Amp) 15-14
15–15 Output of the Circuit in Figure 15–14 15-15
15–16 Phase Shift Oscillator, Buffered 15-16
15–17 Output of the Circuit Figure 15–16 15-16
15–18 Bubba Oscillator 15-17
15–19 Output of the Circuit in Figure 15–18. 15-18

15–20 Quadrature Oscillator 15-19
15–21 Output of the Circuit in Figure 15–20 15-20

16–1 Second-Order Passive Low-Pass and Second-Order Active Low-Pass 16-1
16–2 First-Order Passive RC Low-Pass 16-2
16–3 Fourth-Order Passive RC Low-Pass with Decoupling Amplifiers 16-3
16–4 Frequency and Phase Responses of a Fourth-Order Passive RC Low-Pass Filter 16-4
16–5 Amplitude Responses of Butterworth Low-Pass Filters 16-6
16–6 Gain Responses of Tschebyscheff Low-Pass Filters 16-7
16–7 Comparison of Phase Responses of Fourth-Order Low-Pass Filters 16-8
16–8 Comparison of Normalized Group Delay (Tgr) of Fourth-Order Low-Pass Filters 16-8
16–9 Comparison of Gain Responses of Fourth-Order Low-Pass Filters 16-9
16–10 Graphical Presentation of Quality Factor Q on a Tenth-Order Tschebyscheff Low-Pass Filter with 3-dB Passband Ripple 16-10
16–11 Cascading Filter Stages for Higher-Order Filters 16-12
16–12 First-Order Noninverting Low-Pass Filter 16-12
16–13 First-Order Inverting Low-Pass Filter 16-13
16–14 First-Order Noninverting Low-Pass Filter with Unity Gain 16-14
16–15 General Sallen-Key Low-Pass Filter 16-15
16–16 Unity-Gain Sallen-Key Low-Pass Filter 16-15
16–17 Second-Order Unity-Gain Tschebyscheff Low-Pass with 3-dB Ripple 16-16
16–18 Adjustable Second-Order Low-Pass Filter 16-17
16–19 Second-Order MFB Low-Pass Filter 16-18
16–20 First-Order Unity-Gain Low-Pass 16-19
16–21 Second-Order Unity-Gain Sallen-Key Low-Pass Filter 16-20
16–22 Fifth-Order Unity-Gain Butterworth Low-Pass Filter 16-21
16–23 Low-Pass to High-Pass Transition Through Components Exchange 16-21
16–24 Developing The Gain Response of a High-Pass Filter 16-22
16–25 First-Order Noninverting High-Pass Filter 16-23
16–26 First-Order Inverting High-Pass Filter 16-23
16–27 General Sallen-Key High-Pass Filter 16-24
16–28 Unity-Gain Sallen-Key High-Pass Filter 16-24
16–29 Second-Order MFB High-Pass Filter 16-25
16–30 Third-Order Unity-Gain Bessel High-Pass 16-27
16–31 Low-Pass to Band-Pass Transition 16-28
16–32 Gain Response of a Second-Order Band-Pass Filter 16-29
16–33 Sallen-Key Band-Pass 16-30
16–34 MFB Band-Pass 16-31
16–35 Gain Responses of a Fourth-Order Butterworth Band-Pass and its Partial Filters 16-36
16–36 Low-Pass to Band-Rejection Transition 16-37
16–37 Passive Twin-T Filter 16-37
16–38 Active Twin-T Filter 16-38
16–39 Passive Wien-Robinson Bridge 16-39
16–40 Active Wien-Robinson Filter 16-39

16–41 Comparison of Q Between Passive and Active Band-Rejection Filters 16-41
16–42 Frequency Response of the Group Delay for the First 10 Filter Orders 16-43
16–43 First-Order All-Pass 16-44
16–44 Second-Order All-Pass Filter 16-44
16–45 Seventh-Order All-Pass Filter 16-46
16–46 Dual-Supply Filter Circuit 16-47
16–47 Single-Supply Filter Circuit 16-47
16–48 Biasing a Sallen-Key Low-Pass 16-48
16–49 Biasing a Second-Order MFB Low-Pass Filter 16-49
16–50 Biasing a Sallen-Key and an MFB High-Pass Filter 16-50
16–51 Differences in Q and fC in the Partial Filters of an Eighth-Order Butterworth Low-Pass Filter 16-51
16–52 Modification of the Intended Butterworth Response to aTschebyscheff-Type Characteristic 16-51
16–53 Open-Loop Gain (AOL) and Filter Response (A) 16-53

17–1 Op Amp Terminal Model 17-5
17–2 Digital and Analog Plane Placement 17-8
17–3 Separate Grounds 17-8
17–4 Broadcasting From PCB Traces 17-9
17–5 A Careful Board Layout 17-10
17–6 Resistor High-Frequency Performance 17-11
17–7 Capacitor High-Frequency Performance 17-12
17–8 Inductor High-Frequency Performance 17-13
17–9 Loop and Slot Antenna Board Trace Layouts 17-15
17–10 PCB Trace Corners 17-16
17–11 PCB Trace-to-Plane Capacitance Formula 17-17
17–12 Effect of 1-pF Capacitance on Op Amp Inverting Input 17-18
17–13 Coupling Between Parallel Signal Traces 17-19
17–14 Via Inductance Measurements 17-19
17–15 Logic Gate Output Structure 17-21
17–16 Capacitor Self Resonance 17-22
17–17 Common Op Amp Pinouts 17-24
17–18 Trace Length for an Inverting Op Amp Stage 17-25
17–19 Mirror-Image Layout for Quad Op Amp Package 17-25
17–20 Quad Op Amp Package Layout with Half-Supply Generator 17-26
17–21 Proper Termination of Unused Op Amp Sections 17-28

18–1 Op Amp Error Sources 18-4
18–2 Noninverting Op Amp 18-7
18–3 Input Circuit of a NonRRI Op Amp 18-8
18–4 Input Circuit of an RRI Op Amp 18-8
18–5 Input Bias Current Changes with Input Common-Mode Voltage 18-9
18–6 Input Offset Voltage Changes with Input Common-Mode Voltage 18-10

18–7 RRO Output Stage 18-11
18–8 Data Acquisition System 18-14
18–9 Schematic for the Transducer to ADC Interface Circuit 18-15
18–10 Final Schematic for the Transducer to ADC Interface Circuit 18-16
18–11 Digital Control System 18-16
18–12 DAC Current Source to Actuator Interface Circuit 18-17
18–13 DAC Current Sink to Actuator Interface Circuit 18-19

A–1 Split-Supply Op Amp Circuit A-1
A–2 Split-Supply Op Amp Circuit With Reference Voltage Input A-2
A–3 Split-Supply Op Amp Circuit With Common-Mode Voltage A-3
A–4 Single-Supply Op Amp Circuit A-3
A–5 Inverting Op Amp with Noninverting Positive Reference A-5
A–6 Inverting Op Amp with Inverting Negative Reference A-6
A–7 Inverting Op Amp with Noninverting Negative Reference A-7
A–8 Inverting Op Amp with Inverting Positive Reference A-8
A–9 Noninverting Op Amp with Inverting Positive Reference A-9
A–10 Noninverting Op Amp with Noninverting Negative Reference A-10
A–11 Noninverting Op Amp with Inverting Positive Reference A-11
A–12 Noninverting Op Amp with Noninverting Positive Reference A-12
A–13 Differential Amplifier A-13
A–14 Differential Amplifier with Bias Correction A-14
A–15 High Input Impedance Differential Amplifier A-15
A–16 High Common-Mode Range Differential Amplifier A-16
A–17 High-Precision Differential Amplifier A-17
A–18 Simplified High-Precision Differential Amplifier A-18
A–19 Variable Gain Differential Amplifier A-19
A–20 T Network in the Feedback Loop A-20
A–21 Buffer A-21
A–22 Inverting AC Amplifier A-22
A–23 Noninverting AC Amplifier A-23
A–24 Inverting Summer A-24
A–25 Noninverting Summer A-25
A–26 Noninverting Summer with Buffers A-26
A–27 Inverting Integrator A-27
A–28 Inverting Integrator with Input Current Compensation A-28
A–29 Inverting Integrator with Drift Compensation A-29
A–30 Inverting Integrator with Mechanical Reset A-30
A–31 Inverting Integrator with Electronic Reset A-31
A–32 Inverting Integrator with Resistive Reset A-32
A–33 Noninverting Integrator with Inverting Buffer A-33
A–34 Noninverting Integrator Approximation A-34
A–35 Inverting Differentiator A-35
A–36 Inverting Differentiator with Noise Filter A-36

A–37 Basic Wien Bridge Oscillator A-37
A–38 Wien Bridge Oscillator with Nonlinear Feedback A-38
A–39 Wien Bridge Oscillator with AGC A-39
A–40 Quadrature Oscillator A-40
A–41 Classical Phase Shift Oscillator A-41
A–42 Buffered Phase Shift Oscillator A-42
A–43 Bubba Oscillator A-43
A–44 Triangle Oscillator A-44

Tables

3–1 Basic Ideal Op Amp Assumptions 3-2

8–1 Data Set for Curves in Figure 8–7 8-11

9–1 Tabulation of Pertinent VFA and CFA Equations 9-9

10–1 Noise Colors 10-10

11–1 Op Amp Parameter GLossary 11-2
11–2 Cross-Reference of Op Amp Parameters 11-6

12–1 Transducer Output Voltage 12-14
12–2 ADC Input Voltage 12-15
12–3 Op Amp Selection 12-16
12–4 AIA Input and Output Voltages 12-16
12–5 Offset and Gain Error Budget 12-19

13–1 GSM Receiver Block System Budget 13-3
13–2 High-Speed Op Amp Requirements 13-11

14–1 DC Step Size for D/A converters 14-7
14–2 Converter Bits, THD, and Dynamic Range 14-10

16–1 Second-Order FIlter Coefficients 16-17
16–2 Values of α For Different Filter Types and Different Qs 16-33
16–3 Single-Supply Op Amp Selection Guide (T_A = 25°C, V_{CC} = 5 V) 16-54
16–4 Bessel Coefficients 16-56
16–5 Butterworth Coefficients 16-57
16–6 Tschebyscheff Coefficients for 0.5-dB Passband Ripple 16-58
16–7 Tschebyscheff Coefficients for 1-dB Passband Ripple 16-59
16–8 Tschebyscheff Coefficients for 2-dB Passband Ripple 16-60
16–9 Tschebyscheff Coefficients for 3-dB Passband Ripple 16-61
16–10 All-Pass Coefficients 16-62

17–1 PCB Materials 17-3
17–2 Recommended Maximum Frequencies for Capacitors 17-21

18–1 Comparison of Op Amp Error Terms 18-5
18–2 Op Amp Parameters 18-21

A–1 Ideal Op Amp Assumptions A-4

Examples

16–1 First-Order Unity-Gain Low-Pass Filter 16-14
16–2 Second-Order Unity-Gain Tschebyscheff Low-Pass Filter 16-16
16–3 Fifth-Order Filter 16-19
16–4 Third-Order High-Pass Filter with f_C = 1 kHz 16-26
16–5 Second-Order MFB Band-Pass Filter with f_m = 1 kHz 16-32
16–6 Fourth-Order Butterworth Band-Pass Filter 16-34
16–7 2-ms Delay All-Pass Filter 16-45

Chapter 1

The Op Amp's Place In The World

Ron Mancini

In 1934 Harry Black[1] commuted from his home in New York City to work at Bell Labs in New Jersey by way of a railroad/ferry. The ferry ride relaxed Harry enabling him to do some conceptual thinking. Harry had a tough problem to solve; when phone lines were extended long distances, they needed amplifiers, and undependable amplifiers limited phone service. First, initial tolerances on the gain were poor, but that problem was quickly solved with an adjustment. Second, even when an amplifier was adjusted correctly at the factory, the gain drifted so much during field operation that the volume was too low or the incoming speech was distorted.

Many attempts had been made to make a stable amplifier, but temperature changes and power supply voltage extremes experienced on phone lines caused uncontrollable gain drift. Passive components had much better drift characteristics than active components had, thus if an amplifier's gain could be made dependent on passive components, the problem would be solved. During one of his ferry trips, Harry's fertile brain conceived a novel solution for the amplifier problem, and he documented the solution while riding on the ferry.

The solution was to first build an amplifier that had more gain than the application required. Then some of the amplifier output signal was fed back to the input in a manner that makes the circuit gain (circuit is the amplifier and feedback components) dependent on the feedback circuit rather than the amplifier gain. Now the circuit gain is dependent on the passive feedback components rather than the active amplifier. This is called negative feedback, and it is the underlying operating principle for all modern day op amps. Harry had documented the first intentional feedback circuit during a ferry ride. I am sure unintentional feedback circuits had been built prior to that time, but the designers ignored the effect!

I can hear the squeals of anguish coming from the managers and amplifier designers. I imagine that they said something like this, "it is hard enough to achieve 30-kHz gain–bandwidth (GBW), and now this fool wants me to design an amplifier with 3-MHz GBW. But, he is still going to get a circuit gain GBW of 30 kHz". Well, time has proven Harry right, but there is a minor problem that Harry didn't discuss in detail, and that is the oscillation

problem. It seems that circuits designed with large open loop gains sometimes oscillate when the loop is closed. A lot of people investigated the instability effect, and it was pretty well understood in the 1940s, but solving stability problems involved long, tedious, and intricate calculations. Years passed without anybody making the problem solution simpler or more understandable.

In 1945 H. W. Bode presented a system for analyzing the stability of feedback systems by using graphical methods. Until this time, feedback analysis was done by multiplication and division, so calculation of transfer functions was a time consuming and laborious task. Remember, engineers did not have calculators or computers until the ’70s. Bode presented a log technique that transformed the intensely mathematical process of calculating a feedback system’s stability into graphical analysis that was simple and perceptive. Feedback system design was still complicated, but it no longer was an art dominated by a few electrical engineers kept in a small dark room. Any electrical engineer could use Bode’s methods to find the stability of a feedback circuit, so the application of feedback to machines began to grow. There really wasn’t much call for electronic feedback design until computers and transducers become of age.

The first real-time computer was the analog computer! This computer used preprogrammed equations and input data to calculate control actions. The programming was hard wired with a series of circuits that performed math operations on the data, and the hard wiring limitation eventually caused the declining popularity of the analog computer. The heart of the analog computer was a device called an operational amplifier because it could be configured to perform many mathematical operations such as multiplication, addition, subtraction, division, integration, and differentiation on the input signals. The name was shortened to the familiar *op amp*, as we have come to know and love them. The op amp used an amplifier with a large open loop gain, and when the loop was closed, the amplifier performed the mathematical operations dictated by the external passive components. This amplifier was very large because it was built with vacuum tubes and it required a high-voltage power supply, but it was the heart of the analog computer, thus its large size and huge power requirements were accepted as the price of doing business. Many early op amps were designed for analog computers, and it was soon found out that op amps had other uses and were very handy to have around the physics lab.

At this time general-purpose analog computers were found in universities and large company laboratories because they were critical to the research work done there. There was a parallel requirement for transducer signal conditioning in lab experiments, and op amps found their way into signal conditioning applications. As the signal conditioning applications expanded, the demand for op amps grew beyond the analog computer requirements, and even when the analog computers lost favor to digital computers, the op amp survived because of its importance in universal analog applications. Eventually digital computers replaced the analog computers (a sad day for real-time measurements), but the demand for op amps increased as measurement applications increased.

The first signal conditioning op amps were constructed with vacuum tubes prior to the introduction of transistors, so they were large and bulky. During the ’50s, miniature vacuum tubes that worked from lower voltage power supplies enabled the manufacture of op amps that shrunk to the size of a brick used in house construction, so the op amp modules were nicknamed *bricks*. Vacuum tube size and component size decreased until an op amp was shrunk to the size of a single octal vacuum tube. Transistors were commercially developed in the ’60s, and they further reduced op amp size to several cubic inches, but the nickname brick still held on. Now the nickname brick is attached to any electronic module that uses potting compound or non-integrated circuit (IC) packaging methods. Most of these early op amps were made for specific applications, so they were not necessarily general purpose. The early op amps served a specific purpose, but each manufacturer had different specifications and packages; hence, there was little second sourcing among the early op amps.

ICs were developed during the late 1950s and early 1960s, but it wasn’t till the middle 1960s that Fairchild released the μA709. This was the first commercially successful IC op amp, and Robert J. Widler designed it. The μA709 had its share of problems, but any competent analog engineer could use it, and it served in many different analog applications. The major drawback of the μA709 was stability; it required external compensation and a competent analog engineer to apply it. Also, the μA709 was quite sensitive because it had a habit of self destructing under any adverse condition. The self-destruction habit was so prevalent that one major military equipment manufacturer published a paper titled something like, *The 12 Pearl Harbor Conditions of the μA709*. The μA741 followed the μA709, and it is an internally compensated op amp that does not require external compensation if operated under data sheet conditions. Also, it is much more forgiving than the μA709. There has been a never-ending series of new op amps released each year since then, and their performance and reliability has improved to the point where present day op amps can be used for analog applications by anybody.

The IC op amp is here to stay; the latest generation op amps cover the frequency spectrum from 5-kHz GBW to beyond 1-GHz GBW. The supply voltage ranges from guaranteed operation at 0.9 V to absolute maximum voltage ratings of 1000 V. The input current and input offset voltage has fallen so low that customers have problems verifying the specifications during incoming inspection. The op amp has truly become the universal analog IC because it performs all analog tasks. It can function as a line driver, comparator (one bit A/D), amplifier, level shifter, oscillator, filter, signal conditioner, actuator driver, current source, voltage source, and many other applications. The designer’s problem is how to rapidly select the correct circuit/op amp combination and then, how to calculate the passive component values that yield the desired transfer function in the circuit.

This book deals with op amp circuits — not with the innards of op amps. It treats the calculations from the circuit level, and it doesn’t get bogged down in a myriad of detailed calculations. Rather, the reader can start at the level appropriate for them, and quickly move on to the advanced topics. If you are looking for material about the innards of op amps

you are looking in the wrong place. The op amp is treated as a completed component in this book.

The op amp will continue to be a vital component of analog design because it is such a fundamental component. Each generation of electronics equipment integrates more functions on silicon and takes more of the analog circuitry inside the IC. Don't fear, as digital applications increase, analog applications also increase because the predominant supply of data and interface applications are in the real world, and the real world is an analog world. Thus, each new generation of electronics equipment creates requirements for new analog circuits; hence, new generations of op amps are required to fulfill these requirements. Analog design, and op amp design, is a fundamental skill that will be required far into the future.

References

1 Black, H. S., *Stabilized Feedback Amplifiers*, BSTJ, Vol. 13, January 1934

Chapter 2

Review of Circuit Theory

Ron Mancini

2.1 Introduction

Although this book minimizes math, some algebra is germane to the understanding of analog electronics. Math and physics are presented here in the manner in which they are used later, so no practice exercises are given. For example, after the voltage divider rule is explained, it is used several times in the development of other concepts, and this usage constitutes practice.

Circuits are a mix of passive and active components. The components are arranged in a manner that enables them to perform some desired function. The resulting arrangement of components is called a circuit or sometimes a circuit configuration. The art portion of analog design is developing the circuit configuration. There are many published circuit configurations for almost any circuit task, thus all circuit designers need not be artists.

When the design has progressed to the point that a circuit exists, equations must be written to predict and analyze circuit performance. Textbooks are filled with rigorous methods for equation writing, and this review of circuit theory does not supplant those textbooks. But, a few equations are used so often that they should be memorized, and these equations are considered here.

There are almost as many ways to analyze a circuit as there are electronic engineers, and if the equations are written correctly, all methods yield the same answer. There are some simple ways to analyze the circuit without completing unnecessary calculations, and these methods are illustrated here.

2.2 Laws of Physics

Ohm's law is stated as V=IR, and it is fundamental to all electronics. Ohm's law can be applied to a single component, to any group of components, or to a complete circuit. When the current flowing through any portion of a circuit is known, the voltage dropped across that portion of the circuit is obtained by multiplying the current times the resistance (Equation 2–1).

$$V = IR \tag{2–1}$$

In Figure 2–1, Ohm's law is applied to the total circuit. The current, (I) flows through the total resistance (R), and the voltage (V) is dropped across R.

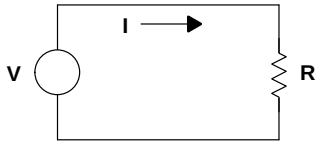


Figure 2–1. Ohm's Law Applied to the Total Circuit

In Figure 2–2, Ohm's law is applied to a single component. The current (I_R) flows through the resistor (R) and the voltage (V_R) is dropped across R. Notice, the same formula is used to calculate the voltage drop across R even though it is only a part of the circuit.

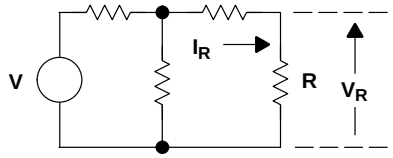


Figure 2–2. Ohm's Law Applied to a Component

Kirchoff's voltage law states that the sum of the voltage drops in a series circuit equals the sum of the voltage sources. Otherwise, the source (or sources) voltage must be dropped across the passive components. When taking sums keep in mind that the sum is an algebraic quantity. Kirchoff's voltage law is illustrated in Figure 2–3 and Equations 2–2 and 2–3.

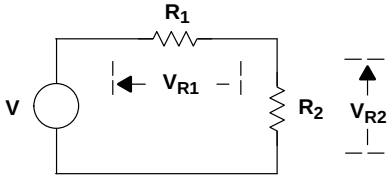


Figure 2–3. Kirchoff's Voltage Law

$$\sum V_{SOURCES} = \sum V_{DROPS} \tag{2–2}$$

$$V = V_{R1} + V_{R2} \tag{2–3}$$

Kirchoff's current law states: the sum of the currents entering a junction equals the sum of the currents leaving a junction. It makes no difference if a current flows from a current

source, through a component, or through a wire, because all currents are treated identically. Kirchoff's current law is illustrated in Figure 2–4 and Equations 2–4 and 2–5.

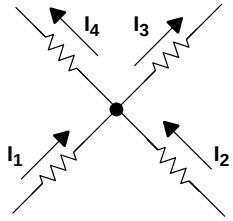


Figure 2–4. Kirchoff's Current Law

$$\sum I_{IN} = \sum I_{OUT} \tag{2–4}$$

$$I_1 + I_2 = I_3 + I_4 \tag{2–5}$$

2.3 Voltage Divider Rule

When the output of a circuit is not loaded, the voltage divider rule can be used to calculate the circuit's output voltage. Assume that the same current flows through all circuit elements (Figure 2–5). Equation 2–6 is written using Ohm's law as $V = I (R_1 + R_2)$. Equation 2–7 is written as Ohm's law across the output resistor.

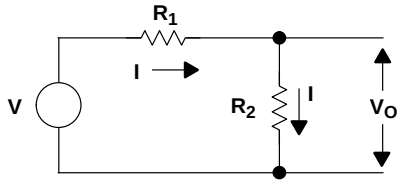


Figure 2–5. Voltage Divider Rule

$$I = \frac{V}{R_1 + R_2} \tag{2–6}$$

$$V_{OUT} = IR_2 \tag{2–7}$$

Substituting Equation 2–6 into Equation 2–7, and using algebraic manipulation yields Equation 2–8.

$$V_{OUT} = V \frac{R_2}{R_1 + R_2} \tag{2–8}$$

A simple way to remember the voltage divider rule is that the output resistor is divided by the total circuit resistance. This fraction is multiplied by the input voltage to obtain the out-

put voltage. Remember that the voltage divider rule always assumes that the output resistor is not loaded; the equation is not valid when the output resistor is loaded by a parallel component. Fortunately, most circuits following a voltage divider are input circuits, and input circuits are usually high resistance circuits. When a fixed load is in parallel with the output resistor, the equivalent parallel value comprised of the output resistor and loading resistor can be used in the voltage divider calculations with no error. Many people ignore the load resistor if it is ten times greater than the output resistor value, but this calculation can lead to a 10% error.

2.4 Current Divider Rule

When the output of a circuit is not loaded, the current divider rule can be used to calculate the current flow in the output branch circuit (R_2). The currents I_1 and I_2 in Figure 2–6 are assumed to be flowing in the branch circuits. Equation 2–9 is written with the aid of Kirchoff's current law. The circuit voltage is written in Equation 2–10 with the aid of Ohm's law. Combining Equations 2–9 and 2–10 yields Equation 2–11.

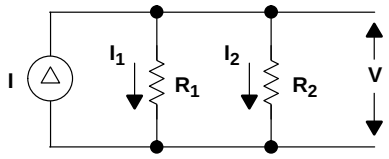


Figure 2–6. Current Divider Rule

$$I = I_1 + I_2 \tag{2–9}$$

$$V = I_1R_1 = I_2R_2 \tag{2–10}$$

$$I = I_1 + I_2 = I_2\frac{R_2}{R_1} + I_2 = I_2\left(\frac{R_1 + R_2}{R_1}\right) \tag{2–11}$$

Rearranging the terms in Equation 2–11 yields Equation 2–12.

$$I_2 = I\left(\frac{R_1}{R_1 + R_2}\right) \tag{2–12}$$

The total circuit current divides into two parts, and the resistance (R_1) divided by the total resistance determines how much current flows through R_2. An easy method of remembering the current divider rule is to remember the voltage divider rule. Then modify the voltage divider rule such that the opposite resistor is divided by the total resistance, and the fraction is multiplied by the input current to get the branch current.

2.5 Thevenin's Theorem

There are times when it is advantageous to isolate a part of the circuit to simplify the analysis of the isolated part of the circuit. Rather than write loop or node equations for the complete circuit, and solving them simultaneously, Thevenin's theorem enables us to isolate the part of the circuit we are interested in. We then replace the remaining circuit with a simple series equivalent circuit, thus Thevenin's theorem simplifies the analysis.

There are two theorems that do similar functions. The Thevenin theorem just described is the first, and the second is called Norton's theorem. Thevenin's theorem is used when the input source is a voltage source, and Norton's theorem is used when the input source is a current source. Norton's theorem is rarely used, so its explanation is left for the reader to dig out of a textbook if it is ever required.

The rules for Thevenin's theorem start with the component or part of the circuit being replaced. Referring to Figure 2–7, look back into the terminals (left from C and R_3 toward point XX in the figure) of the circuit being replaced. Calculate the no load voltage (V_{TH}) as seen from these terminals (use the voltage divider rule).

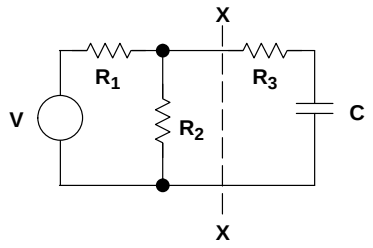


Figure 2–7. Original Circuit

Look into the terminals of the circuit being replaced, short independent voltage sources, and calculate the impedance between these terminals. The final step is to substitute the Thevenin equivalent circuit for the part you wanted to replace as shown in Figure 2–8.

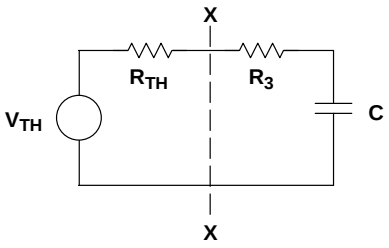
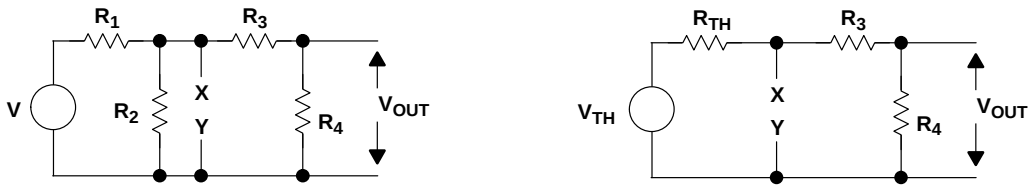


Figure 2–8. Thevenin's Equivalent Circuit for Figure 2–7

The Thevenin equivalent circuit is a simple series circuit, thus further calculations are simplified. The simplification of circuit calculations is often sufficient reason to use Thevenin's

theorem because it eliminates the need for solving several simultaneous equations. The detailed information about what happens in the circuit that was replaced is not available when using Thevenin's theorem, but that is no consequence because you had no interest in it.

As an example of Thevenin's theorem, let's calculate the output voltage (V_{OUT}) shown in Figure 2–9A. The first step is to stand on the terminals X–Y with your back to the output circuit, and calculate the open circuit voltage seen (V_{TH}). This is a perfect opportunity to use the voltage divider rule to obtain Equation 2–13.



(a) The Original Circuit **(b) The Thevenin Equivalent Circuit**

Figure 2–9. Example of Thevenin's Equivalent Circuit

$$V_{TH} = V \frac{R_2}{R_1 + R_2} \tag{2–13}$$

Still standing on the terminals X-Y, step two is to calculate the impedance seen looking into these terminals (short the voltage sources). The Thevenin impedance is the parallel impedance of R_1 and R_2 as calculated in Equation 2–14. Now get off the terminals X-Y before you damage them with your big feet. Step three replaces the circuit to the left of X-Y with the Thevenin equivalent circuit V_{TH} and R_{TH}.

$$R_{TH} = \frac{R_1 R_2}{R_1 + R_2} = R_1 \parallel R_2 \tag{2–14}$$

Note:

Two parallel vertical bars (||) are used to indicate parallel components as shown in Equation 2–14.

The final step is to calculate the output voltage. Notice the voltage divider rule is used again. Equation 2–15 describes the output voltage, and it comes out naturally in the form of a series of voltage dividers, which makes sense. That's another advantage of the voltage divider rule; the answers normally come out in a recognizable form rather than a jumble of coefficients and parameters.

$$V_{OUT} = V_{TH}\frac{R_4}{R_{TH} + R_3 + R_4} = V\left(\frac{R_2}{R_1 + R_2}\right)\frac{R_4}{\frac{R_1R_2}{R_1+R_2} + R_3 + R_4} \tag{2–15}$$

The circuit analysis is done the hard way in Figure 2–10, so you can see the advantage of using Thevenin's Theorem. Two loop currents, I_1 and I_2, are assigned to the circuit. Then the loop Equations 2–16 and 2–17 are written.

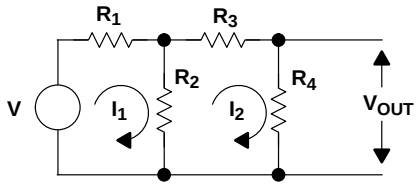


Figure 2–10. Analysis Done the Hard Way

$$V = I_1(R_1 + R_2) - I_2R_2 \tag{2–16}$$

$$I_2(R_2 + R_3 + R_4) = I_1R_2 \tag{2–17}$$

Equation 2–17 is rewritten as Equation 2–18 and substituted into Equation 2–16 to obtain Equation 2–19.

$$I_1 = I_2\frac{R_2 + R_3 + R_4}{R_2} \tag{2–18}$$

$$V = I_2\left(\frac{R_2 + R_3 + R_4}{R_2}\right)(R_1 + R_2) - I_2R_2 \tag{2–19}$$

The terms are rearranged in Equation 2–20. Ohm's law is used to write Equation 2–21, and the final substitutions are made in Equation 2–22.

$$I_2 = \frac{V}{\frac{R_2+R_3+R_4}{R_2}(R_1 + R_2) - R_2} \tag{2–20}$$

$$V_{OUT} = I_2R_4 \tag{2–21}$$

$$V_{OUT} = V\frac{R_4}{\frac{(R_2+R_3+R_4)(R_1+R_2)}{R_2} - R_2} \tag{2–22}$$

This is a lot of extra work for no gain. Also, the answer is not in a usable form because the voltage dividers are not recognizable, thus more algebra is required to get the answer into usable form.

2.6 Superposition

Superposition is a theorem that can be applied to any linear circuit. Essentially, when there are independent sources, the voltages and currents resulting from each source can be calculated separately, and the results are added algebraically. This simplifies the calculations because it eliminates the need to write a series of loop or node equations. An example is shown in Figure 2–11.

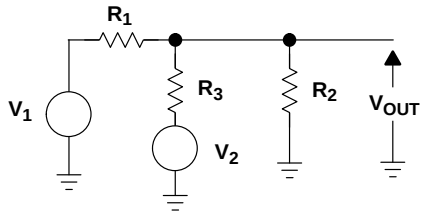


Figure 2–11. Superposition Example

When V_1 is grounded, V_2 forms a voltage divider with R_3 and the parallel combination of R_2 and R_1. The output voltage for this circuit (V_{OUT2}) is calculated with the aid of the voltage divider equation (2–23). The circuit is shown in Figure 2–12. The voltage divider rule yields the answer quickly.

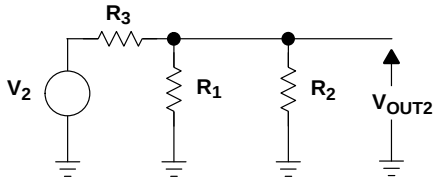


Figure 2–12. When V_1 is Grounded

$$V_{OUT2} = V_2 \frac{R_1 \parallel R_2}{R_3 + R_1 \parallel R_2} \tag{2–23}$$

Likewise, when V_2 is grounded (Figure 2–13), V_1 forms a voltage divider with R_1 and the parallel combination of R_3 and R_2, and the voltage divider theorem is applied again to calculate V_{OUT} (Equation 2–24).

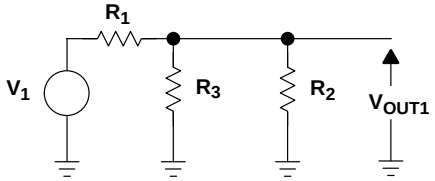


Figure 2–13. When V_2 is Grounded

$$V_{OUT1} = V_1 \frac{R_2 \parallel R_3}{R_1 + R_2 \parallel R_3} \tag{2–24}$$

After the calculations for each source are made the components are added to obtain the final solution (Equation 2–25).

$$V_{OUT} = V_1 \frac{R_2 \parallel R_3}{R_1 + R_2 \parallel R_3} + V_2 \frac{R_1 \parallel R_2}{R_3 + R_1 \parallel R_2} \tag{2–25}$$

The reader should analyze this circuit with loop or node equations to gain an appreciation for superposition. Again, the superposition results come out as a simple arrangement that is easy to understand. One looks at the final equation and it is obvious that if the sources are equal and opposite polarity, and when $R_1 = R_3$, then the output voltage is zero. Conclusions such as this are hard to make after the results of a loop or node analysis unless considerable effort is made to manipulate the final equation into symmetrical form.

2.7 Calculation of a Saturated Transistor Circuit

The circuit specifications are: when $V_{IN} = 12$ V, $V_{OUT} < 0.4$ V at $I_{SINK} < 10$ mA, and $V_{IN} < 0.05$ V, $V_{OUT} > 10$ V at $I_{OUT} = 1$ mA. The circuit diagram is shown in Figure 2–14.

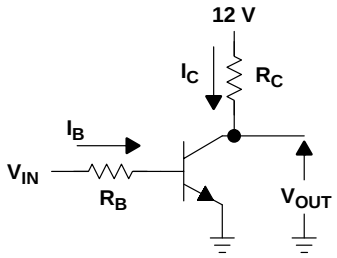


Figure 2–14. Saturated Transistor Circuit

The collector resistor must be sized (Equation 2–26) when the transistor is off, because it has to be small enough to allow the output current to flow through it without dropping more than two volts to meet the specification for a 10-V output.

$$R_C \leq \frac{V_{+12} - V_{OUT}}{I_{OUT}} = \frac{12 - 10}{1} = 2\ \text{k} \tag{2–26}$$

When the transistor is off, 1 mA can be drawn out of the collector resistor without pulling the collector or output voltage to less than ten volts (Equation 2–27). When the transistor is on, the base resistor must be sized (Equation 2–28) to enable the input signal to drive enough base current into the transistor to saturate it. The transistor beta is 50.

$$I_C = \beta I_B = \frac{V_{+12} - V_{CE}}{R_C} + I_L \approx \frac{V_{+12}}{R_C} + I_L \quad (2\text{–}27)$$

$$R_B \leq \frac{V_{IN} - V_{BE}}{I_B} \quad (2\text{–}28)$$

Substituting Equation 2–27 into Equation 2–28 yields Equation 2–29.

$$R_B \leq \frac{(V_{IN} - V_{BE})\beta}{I_C} = \frac{(12 - 0.6)50\ V}{\left[\frac{12}{2} + (10)\right] mA} = 35.6\ k \quad (2\text{–}29)$$

When the transistor goes on it sinks the load current, and it still goes into saturation. These calculations neglect some minor details, but they are in the 98% accuracy range.

2.8 Transistor Amplifier

The amplifier is an analog circuit (Figure 2–15), and the calculations, plus the points that must be considered during the design, are more complicated than for a saturated circuit. This extra complication leads people to say that analog design is harder than digital design (the saturated transistor is digital i.e.; on or off). Analog design is harder than digital design because the designer must account for all states in analog, whereas in digital only two states must be accounted for. The specifications for the amplifier are an ac voltage gain of four and a peak-to-peak signal swing of 4 volts.

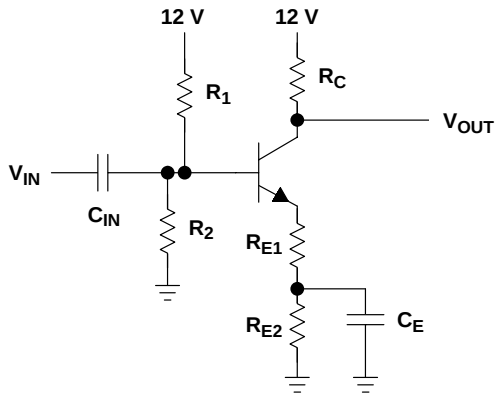


Figure 2–15. Transistor Amplifier

I_C is selected as 10 mA because the transistor has a current gain (β) of 100 at that point. The collector voltage is arbitrarily set at 8 V; when the collector voltage swings positive

2 V (from 8 V to 10 V) there is still enough voltage dropped across R_C to keep the transistor on. Set the collector-emitter voltage at 4 V; when the collector voltage swings negative 2 V (from 8 V to 6 V) the transistor still has 2 V across it, so it stays linear. This sets the emitter voltage (V_E) at 4 V.

$$R_C \leq \frac{V_{+12} - V_C}{I_C} = \frac{12\text{ V} - 8\text{ V}}{10\text{ mA}} = 400\ \Omega \tag{2–30}$$

$$R_E = R_{E1} + R_{E2} = \frac{V_E}{I_E} = \frac{V_E}{I_B + I_C} \cong \frac{V_E}{I_C} = \frac{4\text{ V}}{10\text{ mA}} = 400\ \Omega \tag{2–31}$$

Use Thevenin's equivalent circuit to calculate R_1 and R_2 as shown in Figure 2–16.

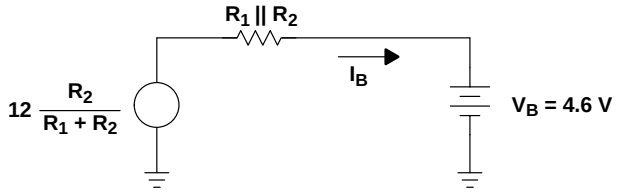


Figure 2–16. Thevenin Equivalent of the Base Circuit

$$I_B = \frac{I_C}{\beta} = \frac{10\text{ mA}}{100} = 0.1\text{ mA} \tag{2–32}$$

$$V_{TH} = \frac{12R_2}{R_1 + R_2} \tag{2–33}$$

$$R_{TH} = \frac{R_1 R_2}{R_1 + R_2} \tag{2–34}$$

We want the base voltage to be 4.6 V because the emitter voltage is then 4 V. Assume a voltage drop of 0.4 V across R_{TH}, so Equation 2–35 can be written. The drop across R_{TH} may not be exactly 0.4 V because of beta variations, but a few hundred mV does not matter is this design. Now, calculate the ratio of R_1 and R_2 using the voltage divider rule (the load current has been accounted for).

$$R_{TH} = \frac{0.4}{0.1}\text{k} = 4\text{ k} \tag{2–35}$$

$$V_{TH} = I_B R_{Th} + V_B = 0.4 + 4.6 = 5 = 12\frac{R_2}{R_1 + R_2} \tag{2–36}$$

$$R_2 = \frac{7}{5} R_1 \tag{2–37}$$

R_1 is almost equal to R_2, thus selecting R_1 as twice the Thevenin resistance yields approximately 4 K as shown in Equation 2–35. Hence, R_1 = 11.2 k and R_2 = 8 k. The ac gain is

approximately R_C/R_{E1} because C_E shorts out R_{E2} at high frequencies, so we can write Equation 2–38.

$$R_{E1} = \frac{R_C}{G} = \frac{400}{4} = 100\ \Omega \tag{2–38}$$

$$R_{E2} = R_E - R_{E1} = 400 - 100 = 300\ \Omega \tag{2–39}$$

The capacitor selection depends on the frequency response required for the amplifier, but 10 μF for C_{IN} and 1000 μF for C_E suffice for a starting point.

Chapter 3

Development of the Ideal Op Amp Equations

Ron Mancini

3.1 Ideal Op Amp Assumptions

The name *Ideal Op Amp* is applied to this and similar analysis because the salient parameters of the op amp are assumed to be perfect. There is no such thing as an ideal op amp, but present day op amps come so close to ideal that *Ideal Op Amp* analysis approaches actual analysis. Op amps depart from the ideal in two ways. First, dc parameters such as input offset voltage are large enough to cause departure from the ideal. The ideal assumes that input offset voltage is zero. Second, ac parameters such as gain are a function of frequency, so they go from large values at dc to small values at high frequencies.

This assumption simplifies the analysis, thus it clears the path for insight. It is so much easier to see the forest when the brush and huge trees are cleared away. Although the ideal op amp analysis makes use of perfect parameters, the analysis is often valid because some op amps approach perfection. In addition, when working at low frequencies, several kHz, the ideal op amp analysis produces accurate answers. Voltage feedback op amps are covered in this chapter, and current feedback op amps are covered in Chapter 8.

Several assumptions have to be made before the ideal op amp analysis can proceed. First, assume that the current flow into the input leads of the op amp is zero. This assumption is almost true in FET op amps where input currents can be less than a pA, but this is not always true in bipolar high-speed op amps where tens of μA input currents are found.

Second, the op amp gain is assumed to be infinite, hence it drives the output voltage to any value to satisfy the input conditions. This assumes that the op amp output voltage can achieve any value. In reality, saturation occurs when the output voltage comes close to a power supply rail, but reality does not negate the assumption, it only bounds it.

Also, implicit in the infinite gain assumption is the need for zero input signal. The gain drives the output voltage until the voltage between the input leads (the error voltage) is zero. This leads to the third assumption that the voltage between the input leads is zero. The implication of zero voltage between the input leads means that if one input is tied to

a hard voltage source such as ground, then the other input is at the same potential. The current flow into the input leads is zero, so the input impedance of the op amp is infinite.

Fourth, the output impedance of the ideal op amp is zero. The ideal op amp can drive any load without an output impedance dropping voltage across it. The output impedance of most op amps is a fraction of an ohm for low current flows, so this assumption is valid in most cases. Fifth, the frequency response of the ideal op amp is flat; this means that the gain does not vary as frequency increases. By constraining the use of the op amp to the low frequencies, we make the frequency response assumption true.

Table 3–1 lists the basic ideal op amp assumptions and FIgure 3–1shows the ideal op amp.

Table 3–1. Basic Ideal Op Amp Assumptions

PARAMETER NAME	PARAMETERS SYMBOL	VALUE
Input current	I_{IN}	0
Input offset voltage	V_{OS}	0
Input impedance	Z_{IN}	∞
Output impedance	Z_{OUT}	0
Gain	a	∞

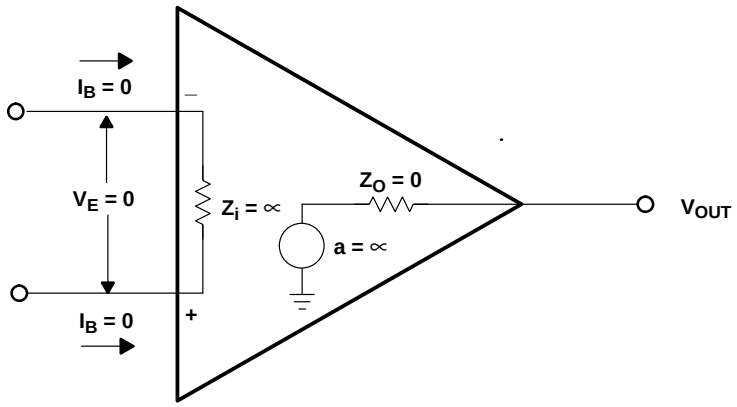


Figure 3–1. The Ideal Op Amp

3.2 The Noninverting Op Amp

The noninverting op amp has the input signal connected to its noninverting input (Figure 3–2), thus its input source sees an infinite impedance. There is no input offset voltage because $V_{OS} = V_E = 0$, hence the negative input must be at the same voltage as the positive input. The op amp output drives current into R_F until the negative input is at the voltage, V_{IN}. This action causes V_{IN} to appear across R_G.

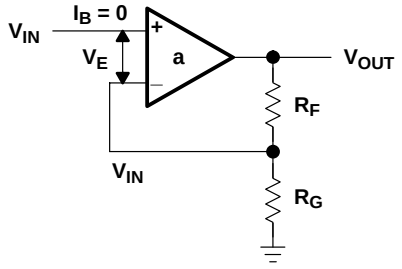


Figure 3–2. The Noninverting Op Amp

The voltage divider rule is used to calculate V_{IN}; V_{OUT} is the input to the voltage divider, and V_{IN} is the output of the voltage divider. Since no current can flow into either op amp lead, use of the voltage divider rule is allowed. Equation 3–1 is written with the aid of the voltage divider rule, and algebraic manipulation yields Equation 3–2 in the form of a gain parameter.

$$V_{IN} = V_{OUT} \frac{R_G}{R_G + R_F} \tag{3–1}$$

$$\frac{V_{OUT}}{V_{IN}} = \frac{R_G + R_F}{R_G} = 1 + \frac{R_F}{R_G} \tag{3–2}$$

When R_G becomes very large with respect to R_F, $(R_F/R_G) \Rightarrow 0$ and Equation 3–2 reduces to Equation 3–3.

$$V_{OUT} = 1 \tag{3–3}$$

Under these conditions $V_{OUT} = 1$ and the circuit becomes a unity gain buffer. R_G is usually deleted to achieve the same results, and when R_G is deleted, R_F can also be deleted (RF must be shorted when it is deleted). When R_F and R_G are deleted, the op amp output is connected to its inverting input with a wire. Some op amps are self-destructive when R_F is left out of the circuit, so R_F is used in many buffer designs. When R_F is included in a buffer circuit, its function is to protect the inverting input from an over voltage to limit the current through the input ESD (electro-static discharge) structure (typically < 1 mA), and it can have almost any value (20 k is often used). R_F can never be left out of the circuit

in a current feedback amplifier design because R_F determines stability in current feedback amplifiers.

Notice that the gain is only a function of the feedback and gain resistors; therefore the feedback has accomplished its function of making the gain independent of the op amp parameters. The gain is adjusted by varying the ratio of the resistors. The actual resistor values are determined by the impedance levels that the designer wants to establish. If R_F = 10 k and R_G = 10 k the gain is two as shown in Equation 2, and if R_F = 100 k and R_G = 100 k the gain is still two. The impedance levels of 10 k or 100 k determine the current drain, the effect of stray capacitance, and a few other points. The impedance level does not set the gain; the ratio of R_F/R_G does.

3.3 The Inverting Op Amp

The noninverting input of the inverting op amp circuit is grounded. One assumption made is that the input error voltage is zero, so the feedback keeps inverting the input of the op amp at a virtual ground (not actual ground but acting like ground). The current flow in the input leads is assumed to be zero, hence the current flowing through R_G equals the current flowing through R_F. Using Kirchoff's law, we write Equation 3–4; and the minus sign is inserted because this is the inverting input. Algebraic manipulation gives Equation 3–5.

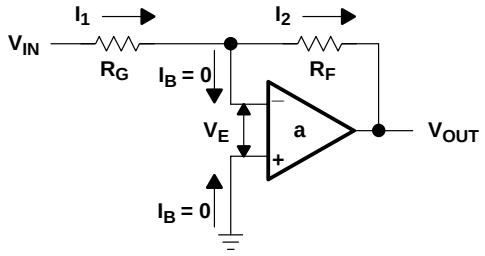


Figure 3–3. The Inverting Op Amp

$$I_1 = \frac{V_{IN}}{R_G} = -I_2 = -\frac{V_{OUT}}{R_F} \tag{3–4}$$

$$\frac{V_{OUT}}{V_{IN}} = -\frac{R_F}{R_G} \tag{3–5}$$

Notice that the gain is only a function of the feedback and gain resistors, so the feedback has accomplished its function of making the gain independent of the op amp parameters. The actual resistor values are determined by the impedance levels that the designer wants to establish. If R_F = 10 k and R_G = 10 k the gain is minus one as shown in Equation

3–5, and if R_F = 100 k and R_G = 100 k the gain is still minus one. The impedance levels of 10 k or 100 k determine the current drain, the effect of stray capacitance, and a few other points. The impedance level does not set the gain; the ratio of R_F/R_G does.

One final note; the output signal is the input signal amplified and inverted. The circuit input impedance is set by R_G because the inverting input is held at a virtual ground.

3.4 The Adder

An adder circuit can be made by connecting more inputs to the inverting op amp (Figure 3–4). The opposite end of the resistor connected to the inverting input is held at virtual ground by the feedback; therefore, adding new inputs does not affect the response of the existing inputs.

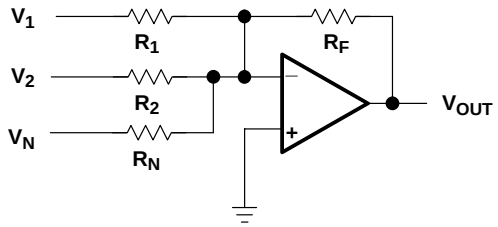


Figure 3–4. The Adder Circuit

Superposition is used to calculate the output voltages resulting from each input, and the output voltages are added algebraically to obtain the total output voltage. Equation 3–6 is the output equation when V_1 and V_2 are grounded. Equations 3–7 and 3–8 are the other superposition equations, and the final result is given in Equation 3–9.

$$V_{OUTN} = -\frac{R_F}{R_N} V_N \tag{3–6}$$

$$V_{OUT1} = -\frac{R_F}{R_1} V_1 \tag{3–7}$$

$$V_{OUT2} = -\frac{R_F}{R_2} V_2 \tag{3–8}$$

$$V_{OUT} = -\left(\frac{R_F}{R_1} V_1 + \frac{R_F}{R_2} V_2 + \frac{R_F}{R_N} V_N\right) \tag{3–9}$$

3.5 The Differential Amplifier

The differential amplifier circuit amplifies the difference between signals applied to the inputs (Figure 3–5). Superposition is used to calculate the output voltage resulting from each input voltage, and then the two output voltages are added to arrive at the final output voltage.

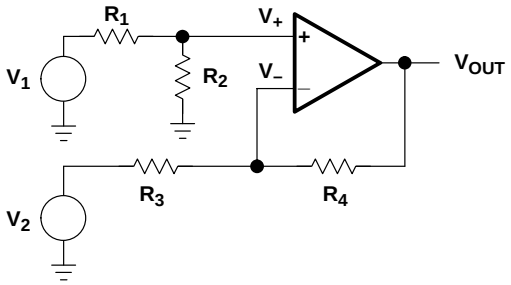


Figure 3–5. The Differential Amplifier

The op amp input voltage resulting from the input source, V_1, is calculated in Equations 3–10 and 3–11. The voltage divider rule is used to calculate the voltage, V_+, and the noninverting gain equation (Equation 3–2) is used to calculate the noninverting output voltage, V_{OUT1}.

$$V_+ = V_1 \frac{R_2}{R_1 + R_2} \tag{3–10}$$

$$V_{OUT1} = V_+(G_+) = V_1 \frac{R_2}{R_1 + R_2}\left(\frac{R_3 + R_4}{R_3}\right) \tag{3–11}$$

The inverting gain equation (Equation 3–5) is used to calculate the stage gain for V_{OUT2} in Equation 3–12. These inverting and noninverting gains are added in Equation 3–13.

$$V_{OUT2} = V_2\left(-\frac{R_4}{R_3}\right) \tag{3–12}$$

$$V_{OUT} = V_1 \frac{R_2}{R_1 + R_2}\left(\frac{R_3 + R_4}{R_3}\right) - V_2 \frac{R_4}{R_3} \tag{3–13}$$

When $R_2 = R_4$ and $R_1 = R_3$, Equation 3–13 reduces to Equation 3–14.

$$V_{OUT} = (V_1 - V_2)\frac{R_4}{R_3} \tag{3–14}$$

It is now obvious that the differential signal, $(V_1–V_2)$, is multiplied by the stage gain, so the name differential amplifier suits the circuit. Because it only amplifies the differential

portion of the input signal, it rejects the common-mode portion of the input signal. A common-mode signal is illustrated in Figure 3–6. Because the differential amplifier strips off or rejects the common-mode signal, this circuit configuration is often employed to strip dc or injected common-mode noise off a signal.

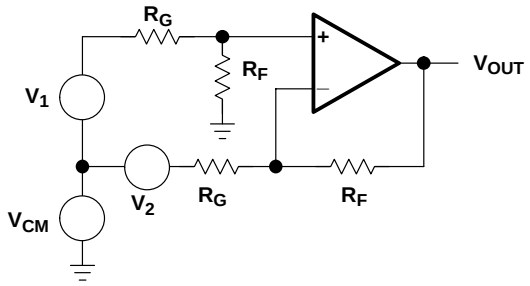


Figure 3–6. Differential Amplifier With Common-Mode Input Signal

The disadvantage of this circuit is that the two input impedances cannot be matched when it functions as a differential amplifier, thus there are two and three op amp versions of this circuit specially designed for high performance applications requiring matched input impedances.

3.6 Complex Feedback Networks

When complex networks are put into the feedback loop, the circuits get harder to analyze because the simple gain equations cannot be used. The usual technique is to write and solve node or loop equations. There is only one input voltage, so superposition is not of any use, but Thevenin's theorem can be used as is shown in the example problem given below.

Sometimes it is desirable to have a low resistance path to ground in the feedback loop. Standard inverting op amps can not do this when the driving circuit sets the input resistor value, and the gain specification sets the feedback resistor value. Inserting a *T* network in the feedback loop (FIgure 3–7) yields a degree of freedom that enables both specifications to be met with a low dc resistance path in the feedback loop.

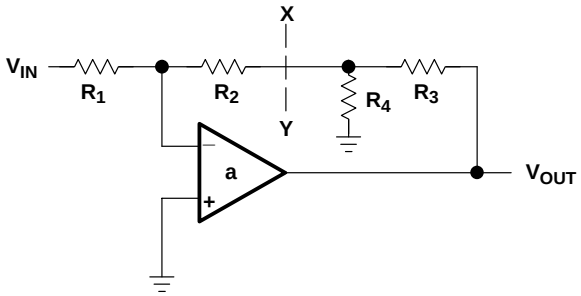


Figure 3–7. T Network in Feedback Loop

Break the circuit at point X–Y, stand on the terminals looking into R_4, and calculate the Thevenin equivalent voltage as shown in Equation 3–15. The Thevenin equivalent impedance is calculated in Equation 3–16.

$$V_{TH} = V_{OUT} \frac{R_4}{R_3 + R_4} \tag{3–15}$$

$$R_{TH} = R_3 \parallel R_4 \tag{3–16}$$

Replace the output circuit with the Thevenin equivalent circuit as shown in Figure 5–8, and calculate the gain with the aid of the inverting gain equation as shown in Equation 3–17.

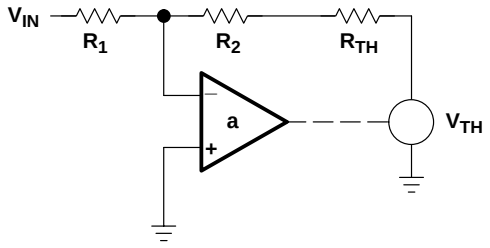


Figure 3–8. Thevenin's Theorem Applied to T Network

Substituting the Thevenin equivalents into Equation 3–17 yields Equation 3–18.

$$-\frac{V_{TH}}{V_{IN}} = \frac{R_2 + R_{TH}}{R_1} \tag{3–17}$$

$$-\frac{V_{OUT}}{V_{IN}} = \frac{R_2 + R_{TH}}{R_1}\left(\frac{R_3 + R_4}{R_4}\right) = \frac{R_2 + \left(R_3 \parallel R_4\right)}{R_1}\left(\frac{R_3 + R_4}{R_4}\right) \tag{3–18}$$

Algebraic manipulation yields Equation 3–19.

$$-\frac{V_{OUT}}{V_{IN}} = \frac{R_2 + R_3 + \frac{R_2 R_3}{R_4}}{R_1} \tag{3–19}$$

Specifications for the circuit you are required to build are an inverting amplifier with an input resistance of 10 k (R_G = 10 k), a gain of 100, and a feedback resistance of 20 K or less. The inverting op amp circuit can not meet these specifications because R_F must equal 1000 k. Inserting a *T* network with $R_2 = R_4$ = 10 k and R_3 = 485 k approximately meets the specifications.

3.7 Video Amplifiers

Video signals contain high frequencies, and they use coaxial cable to transmit and receive signals. The cable connecting these circuits has a characteristic impedance of 75 Ω. To prevent reflections, which cause distortion and ghosting, the input and output circuit impedances must match the 75 Ω cable.

Matching the input impedance is simple for a noninverting amplifier because its input impedance is very high; just make R_{IN} = 75 Ω. R_F and R_G can be selected as high values, in the hundreds of Ohms range, so that they have minimal affect on the impedance of the input or output circuit. A matching resistor, R_M, is placed in series with the op amp output to raise its output impedance to 75 Ω; a terminating resistor, R_T, is placed at the input of the next stage to match the cable (Figure 3–9).

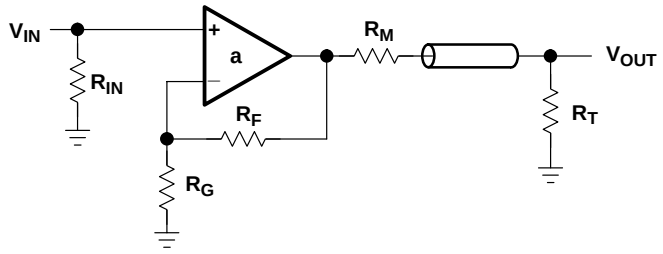


Figure 3–9. Video Amplifier

The matching and terminating resistors are equal in value, and they form a voltage divider of 1/2 because R_T is not loaded. Very often R_F is selected equal to R_G so that the op amp gain equals two. Then the system gain, which is the op amp gain multiplied by the divider gain, is equal to one ($2 \times 1/2 = 1$).

3.8 Capacitors

Capacitors are a key component in a circuit designer's tool kit, thus a short discussion on evaluating their affect on circuit performance is in order. Capacitors have an impedance of $X_C = 1/2\pi fC$. Note that when the frequency is zero the capacitive impedance (also known as reactance) is infinite, and that when the frequency is infinite the capacitive impedance is zero. These end-points are derived from the final value theorem, and they are used to get a rough idea of the effect of a capacitor. When a capacitor is used with a resistor, they form what is called a break-point. Without going into complicated math, just accept that the break frequency occurs at $f = 1/(2\pi RC)$ and the gain is –3 dB at the break frequency.

The low pass filter circuit shown in Figure 3–10 has a capacitor in parallel with the feedback resistor. The gain for the low pass filter is given in Equation 3–20.

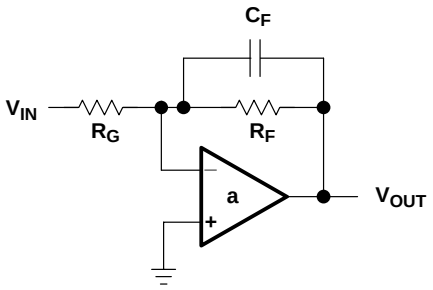


Figure 3–10. Low-Pass Filter

$$\frac{V_{OUT}}{V_{IN}} = -\frac{X_C \parallel R_F}{R_G} \tag{3–20}$$

At very low frequencies $X_C \Rightarrow \infty$, so R_F dominates the parallel combination in Equation 20, and the capacitor has no effect. The gain at low frequencies is $-R_F/R_G$. At very high frequencies $X_C \Rightarrow 0$, so the feedback resistor is shorted out, thus reducing the circuit gain to zero. At the frequency where $X_C = R_F$ the gain is reduced by $\sqrt{2}$ because complex impedances in parallel equal half the vector sum of both impedances.

Connecting the capacitor in parallel with R_G where it has the opposite effect makes a high pass filter (Figure 3–11). Equation 3–21 gives the equation for the high pass filter.

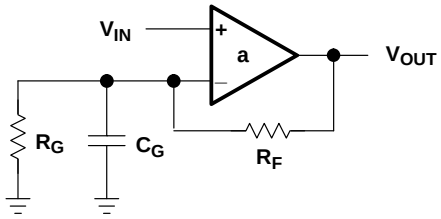


Figure 3–11. High-Pass Filter

$$\frac{V_{OUT}}{V_{IN}} = 1 + \frac{R_F}{X_C \parallel R_G} \tag{3–21}$$

At very low frequencies $X_C \Rightarrow \infty$, so R_G dominates the parallel combination in Equation 3–21, and the capacitor has no effect. The gain at low frequencies is $1+R_F/R_G$. At very high frequencies $X_C \Rightarrow 0$, so the gain setting resistor is shorted out thus increasing the circuit gain to maximum.

This simple technique is used to predict the form of a circuit transfer function rapidly. Better analysis techniques are presented in later chapters for those applications requiring more precision.

3.9 Summary

When the proper assumptions are made, the analysis of op amp circuits is straightforward. These assumptions, which include zero input current, zero input offset voltage, and infinite gain, are realistic assumptions because the new op amps make them true in most applications.

When the signal is comprised of low frequencies, the gain assumption is valid because op amps have very high gain at low frequencies. When CMOS op amps are used, the input current is in the femto amp range; close enough to zero for most applications. Laser trimmed input circuits reduce the input offset voltage to a few micro volts; close enough to zero for most applications. The ideal op amp is becoming real; especially for undemanding applications.

Chapter 4

Single-Supply Op Amp Design Techniques

Ron Mancini

4.1 Single Supply versus Dual Supply

The previous chapter assumed that all op amps were powered from dual or split supplies, and this is not the case in today's world of portable, battery-powered equipment. When op amps are powered from dual supplies (see Figure 4–1), the supplies are normally equal in magnitude, opposing in polarity, and the center tap of the supplies is connected to ground. Any input sources connected to ground are automatically referenced to the center of the supply voltage, so the output voltage is automatically referenced to ground.

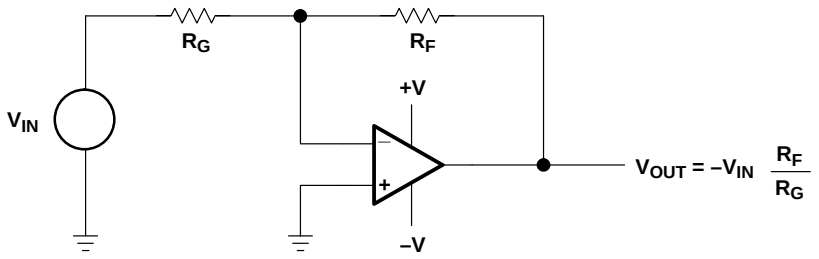


Figure 4–1. Split-Supply Op Amp Circuit

Single-supply systems do not have the convenient ground reference that dual-supply systems have, thus biasing must be employed to ensure that the output voltage swings between the correct voltages. Input sources connected to ground are actually connected to a supply rail in single-supply systems. This is analogous to connecting a dual-supply input to the minus power rail. This requirement for biasing the op amp inputs to achieve the desired output voltage swing complicates single-supply designs.

When the signal source is not referenced to ground (see Figure 4–2), the voltage difference between ground and the reference voltage is amplified along with the signal. Unless the reference voltage was inserted as a bias voltage, and such is not the case when the input signal is connected to ground, the reference voltage must be stripped from the signal so that the op amp can provide maximum dynamic range.

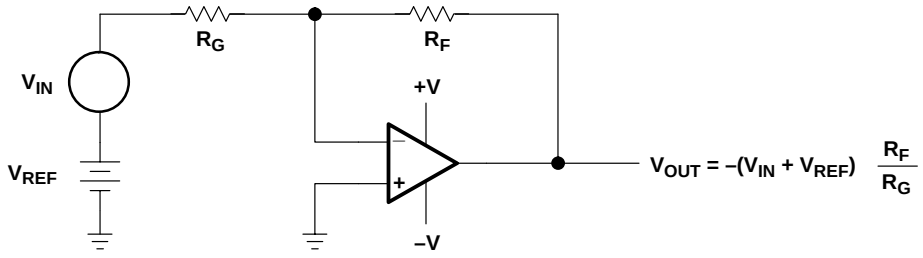


Figure 4–2. Split-Supply Op Amp Circuit With Reference Voltage Input

An input bias voltage is used to eliminate the reference voltage when it must not appear in the output voltage (see Figure 4–3). The voltage, V_{REF}, is in both input circuits, hence it is named a common-mode voltage. Voltage feedback op amps reject common-mode voltages because their input circuit is constructed with a differential amplifier (chosen because it has natural common-mode voltage rejection capabilities).

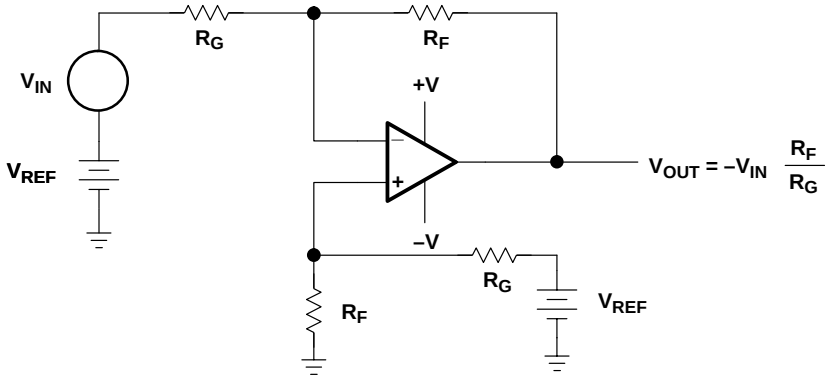


Figure 4–3. Split-Supply Op Amp Circuit With Common-Mode Voltage

When signal sources are referenced to ground, single-supply op amp circuits exhibit a large input common-mode voltage. Figure 4–4 shows a single-supply op amp circuit that has its input voltage referenced to ground. The input voltage is not referenced to the midpoint of the supplies like it would be in a split-supply application, rather it is referenced to the lower power supply rail. This circuit does not operate when the input voltage is positive because the output voltage would have to go to a negative voltage, hard to do with a positive supply. It operates marginally with small negative input voltages because most op amps do not function well when the inputs are connected to the supply rails.

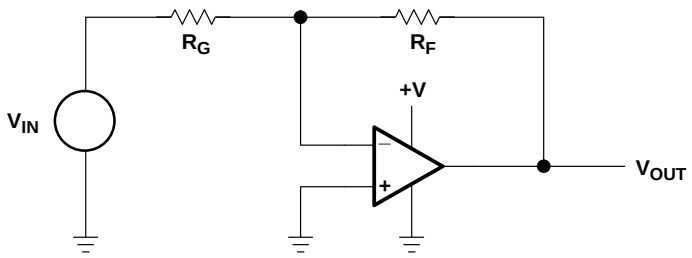


Figure 4–4. Single-Supply Op Amp Circuit

The constant requirement to account for inputs connected to ground or different reference voltages makes it difficult to design single-supply op amp circuits. Unless otherwise specified, all op amp circuits discussed in this chapter are single-supply circuits. The single-supply may be wired with the negative or positive lead connected to ground, but as long as the supply polarity is correct, the wiring does not affect circuit operation.

Use of a single-supply limits the polarity of the output voltage. When the supply voltage V_{CC} = 10 V, the output voltage is limited to the range $0 \leq V_{out} \leq 10$. This limitation precludes negative output voltages when the circuit has a positive supply voltage, but it does not preclude negative input voltages when the circuit has a positive supply voltage. As long as the voltage on the op amp input leads does not become negative, the circuit can handle negative input voltages.

Beware of working with negative (positive) input voltages when the op amp is powered from a positive (negative) supply because op amp inputs are highly susceptible to reverse voltage breakdown. Also, insure that all possible start-up conditions do not reverse bias the op amp inputs when the input and supply voltage are opposite polarity.

4.2 Circuit Analysis

The complexities of single-supply op amp design are illustrated with the following example. Notice that the biasing requirement complicates the analysis by presenting several conditions that are not realizable. It is best to wade through this material to gain an understanding of the problem, especially since a cookbook solution is given later in this chapter. The previous chapter assumed that the op amps were ideal, and this chapter starts to deal with op amp deficiencies. The input and output voltage swing of many op amps are limited as shown in Figure 4–7, but if one designs with the selected rail-to-rail op amps, the input/output swing problems are minimized. The inverting circuit shown in Figure 4–5 is analyzed first.

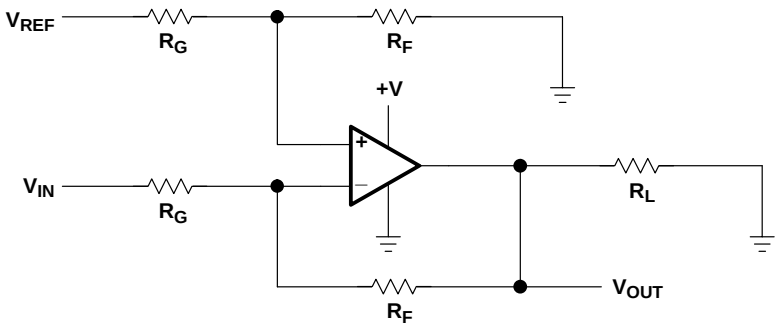


Figure 4–5. Inverting Op Amp

Equation 4–1 is written with the aid of superposition, and simplified algebraically, to acquire Equation 4–2.

$$V_{OUT} = V_{REF}\left(\frac{R_F}{R_G + R_F}\right)\left(\frac{R_F + R_G}{R_G}\right) - V_{IN}\frac{R_F}{R_G} \tag{4–1}$$

$$V_{OUT} = \left(V_{REF} - V_{IN}\right)\frac{R_F}{R_G} \tag{4–2}$$

As long as the load resistor, R_L, is a large value, it does not enter into the circuit calculations, but it can introduce some second order effects such as limiting the output voltage swings. Equation 4–3 is obtained by setting V_{REF} equal to V_{IN}, and there is no output voltage from the circuit regardless of the input voltage. The author unintentionally designed a few of these circuits before he created an orderly method of op amp circuit design. Actually, a real circuit has a small output voltage equal to the lower transistor saturation voltage, which is about 150 mV for a TLC07X.

$$V_{OUT} = \left(V_{REF} - V_{IN}\right)\frac{R_F}{R_G} = \left(V_{IN} - V_{IN}\right)\frac{R_F}{R_G} = 0 \tag{4–3}$$

When $V_{REF} = 0$, $V_{OUT} = -V_{IN}(R_F/R_G)$, there are two possible solutions to Equation 4–2. First, when V_{IN} is any positive voltage, V_{OUT} should be negative voltage. The circuit can not achieve a negative voltage with a positive supply, so the output saturates at the lower power supply rail. Second, when V_{IN} is any negative voltage, the output spans the normal range according to Equation 4–5.

$$V_{IN} \geq 0, \qquad V_{OUT} = 0 \tag{4–4}$$

$$V_{IN} \leq 0, \qquad V_{OUT} = \left|V_{IN}\right| \frac{R_F}{R_G} \tag{4–5}$$

When V_{REF} equals the supply voltage, V_{CC}, we obtain Equation 4–6. In Equation 4–6, when V_{IN} is negative, V_{OUT} should exceed V_{CC}; that is impossible, so the output saturates. When V_{IN} is positive, the circuit acts as an inverting amplifier.

$$V_{OUT} = \left(V_{CC} - V_{IN}\right)\frac{R_F}{R_G} \quad (4–6)$$

The transfer curve for the circuit shown in Figure 4–6 (V_{CC} = 5 V, $R_G = R_F$ = 100 kΩ, R_L = 10 kΩ) is shown in Figure 4–7.

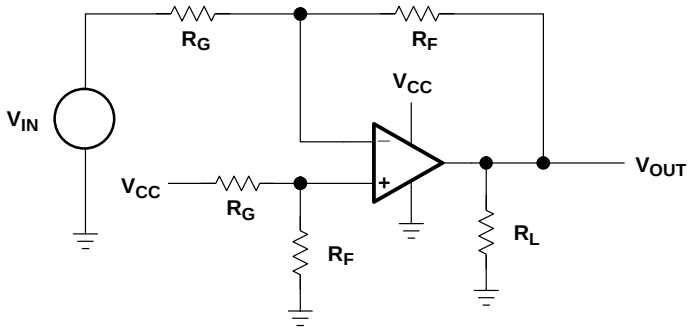


Figure 4–6. Inverting Op Amp With V_{CC} Bias

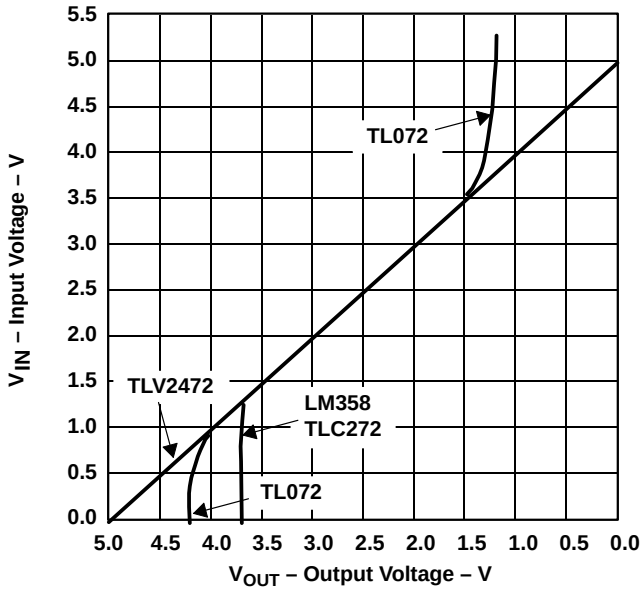


Figure 4–7. Transfer Curve for Inverting Op Amp With V_{CC} Bias

Four op amps were tested in the circuit configuration shown in Figure 4–6. Three of the old generation op amps, LM358, TL07X, and TLC272 had output voltage spans of 2.3 V to 3.75 V. This performance does not justify the ideal op amp assumption that was made in the previous chapter unless the output voltage swing is severely limited. Limited output or input voltage swing is one of the worst deficiencies a single-supply op amp can have because the limited voltage swing limits the circuit's dynamic range. Also, limited voltage swing frequently results in distortion of large signals. The fourth op amp tested was the newer TLV247X, which was designed for rail-to-rail operation in single-supply circuits. The TLV247X plotted a perfect curve (results limited by the instrumentation), and it amazed the author with a textbook performance that justifies the use of ideal assumptions. Some of the older op amps must limit their transfer equation as shown in Equation 4–7.

$$V_{OUT} = \left(V_{CC} - V_{IN}\right)\frac{R_F}{R_G} \quad \text{for} \quad V_{OH} \geq V_{OUT} \geq V_{OL} \tag{4–7}$$

The noninverting op amp circuit is shown in Figure 4–8. Equation 4–8 is written with the aid of superposition, and simplified algebraically, to acquire Equation 4–9.

$$V_{OUT} = V_{IN}\left(\frac{R_F}{R_G + R_F}\right)\left(\frac{R_F + R_G}{R_G}\right) - V_{REF}\,\frac{R_F}{R_G} \tag{4–8}$$

$$V_{OUT} = \left(V_{IN} - V_{REF}\right)\frac{R_F}{R_G} \tag{4–9}$$

When $V_{REF} = 0$, $V_{OUT} = V_{IN}\,\frac{R_F}{R_G}$, there are two possible circuit solutions. First, when V_{IN} is a negative voltage, V_{OUT} must be a negative voltage. The circuit can not achieve a negative output voltage with a positive supply, so the output saturates at the lower power supply rail. Second, when V_{IN} is a positive voltage, the output spans the normal range as shown by Equation 4–11.

$$V_{IN} \leq 0, \qquad V_{OUT} = 0 \tag{4–10}$$

$$V_{IN} \geq 0, \qquad V_{OUT} = V_{IN} \tag{4–11}$$

The noninverting op amp circuit is shown in Figure 4–8 with $V_{CC} = 5$ V, $R_G = R_F = 100$ kΩ, and $R_L = 10$ kΩ. The transfer curve for this circuit is shown in Figure 4–9; a TLV247X serves as the op amp.

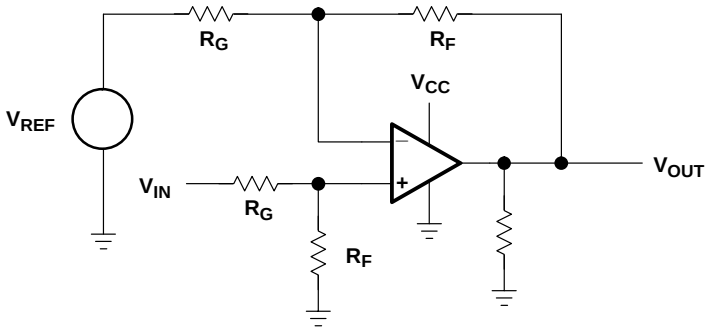


Figure 4–8. Noninverting Op Amp

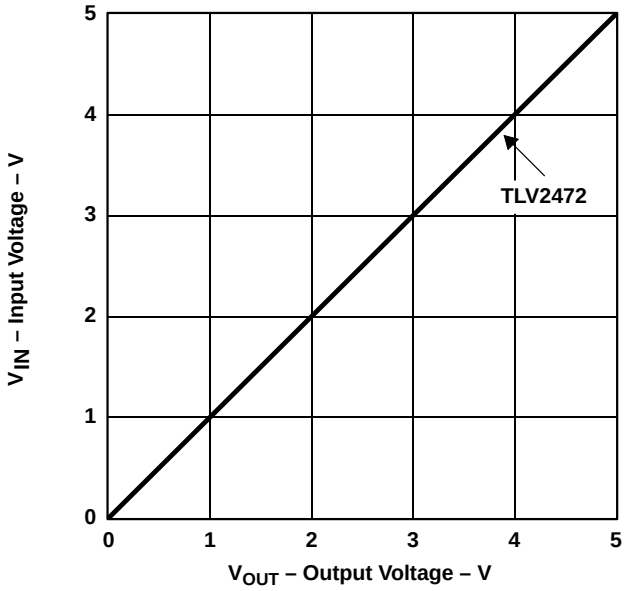


Figure 4–9. Transfer Curve for Noninverting Op Amp

There are many possible variations of inverting and noninverting circuits. At this point many designers analyze these variations hoping to stumble upon the one that solves the circuit problem. Rather than analyze each circuit, it is better to learn how to employ simultaneous equations to render specified data into equation form. When the form of the desired equation is known, a circuit that fits the equation is chosen to solve the problem. The resulting equation must be a straight line, thus there are only four possible solutions.

4.3 Simultaneous Equations

Taking an orderly path to developing a circuit that works the first time starts here; follow these steps until the equation of the op amp is determined. Use the specifications given for the circuit coupled with simultaneous equations to determine what form the op amp equation must have. Go to the section that illustrates that equation form (called a case), solve the equation to determine the resistor values, and you have a working solution.

A linear op amp transfer function is limited to the equation of a straight line (Equation 4–12).

$$y = \pm\ mx \pm b \tag{4–12}$$

The equation of a straight line has four possible solutions depending upon the sign of m, the slope, and b, the intercept; thus simultaneous equations yield solutions in four forms. Four circuits must be developed; one for each form of the equation of a straight line. The four equations, cases, or forms of a straight line are given in Equations 4–13 through 4–16, where electronic terminology has been substituted for math terminology.

$$V_{OUT} = +\ mV_{IN} + b \tag{4–13}$$

$$V_{OUT} = +\ mV_{IN} - b \tag{4–14}$$

$$V_{OUT} = -\ mV_{IN} + b \tag{4–15}$$

$$V_{OUT} = -\ mV_{IN} - b \tag{4–16}$$

Given a set of two data points for V_{OUT} and V_{IN}, simultaneous equations are solved to determine m and b for the equation that satisfies the given data. The sign of m and b determines the type of circuit required to implement the solution. The given data is derived from the specifications; i. e., a sensor output signal ranging from 0.1 V to 0.2 V must be interfaced into an analog-to-digital converter that has an input voltage range of 1 V to 4 V. These data points (V_{OUT} = 1 V @ V_{IN} = 0.1 V, V_{OUT} = 4 V @ V_{IN} = 0.2 V) are inserted into Equation 4–13, as shown in Equations 4–17 and 4–18, to obtain m and b for the specifications.

$$1 = m(0.1) + b \tag{4–17}$$

$$4 = m(0.2) + b \tag{4–18}$$

Multiply Equation 4–17 by 2 and subtract it from Equation 4–18.

$$2 = m(0.2) + 2b \tag{4–19}$$

$$b = -2 \tag{4–20}$$

After algebraic manipulation of Equation 4–17, substitute Equation 4–20 into Equation 4–17 to obtain Equation 4–21.

$$m = \frac{2 + 1}{0.1} = 30 \tag{4–21}$$

Now m and b are substituted back into Equation 4–13 yielding Equation 4–22.

$$V_{OUT} = 30V_{IN} - 2 \tag{4–22}$$

Notice, although Equation 4–13 was the starting point, the form of Equation 4–22 is identical to the format of Equation 4–14. The specifications or given data determine the sign of m and b, and starting with Equation 4–13, the final equation form is discovered after m and b are calculated. The next step required to complete the problem solution is to develop a circuit that has an m = 30 and b = –2. Circuits were developed for Equations 4–13 through 4–16, and they are given under the headings Case 1 through Case 4 respectively. There are different circuits that will yield the same equations, but these circuits were selected because they do not require negative references.

4.3.1 Case 1: $V_{OUT} = +mV_{IN}+b$

The circuit configuration that yields a solution for Case 1 is shown in Figure 4–10. The figure includes two 0.01-µF capacitors. These capacitors are called decoupling capacitors, and they are included to reduce noise and provide increased noise immunity. Sometimes two 0.01-µF capacitors serve this purpose, sometimes more extensive filtering is needed, and sometimes one capacitor serves this purpose. Special attention must be paid to the regulation and noise content of V_{CC} when V_{CC} is used as a reference because some portion of the noise content of V_{CC} will be multiplied by the circuit gain.

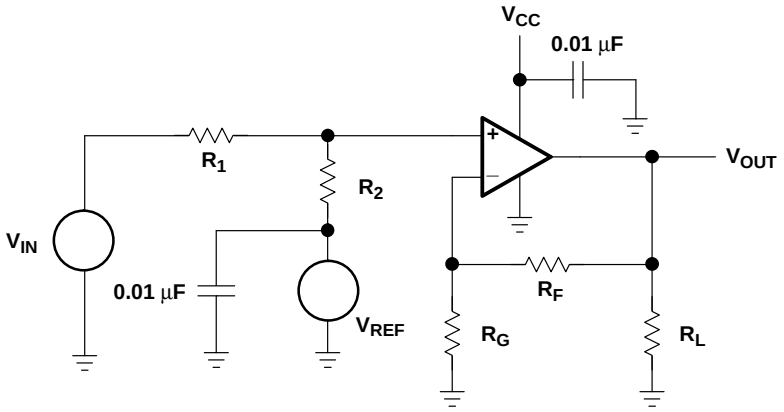


Figure 4–10. Schematic for Case1: $V_{OUT} = +mV_{IN} + b$

The circuit equation is written using the voltage divider rule and superposition.

$$V_{OUT} = V_{IN}\left(\frac{R_2}{R_1 + R_2}\right)\left(\frac{R_F + R_G}{R_G}\right) + V_{REF}\left(\frac{R_1}{R_1 + R_2}\right)\left(\frac{R_F + R_G}{R_G}\right) \quad (4–23)$$

The equation of a straight line (case 1) is repeated in Equation 4–24 below so comparisons can be made between it and Equation 4–23.

$$V_{OUT} = mV_{IN} + b \quad (4–24)$$

Equating coefficients yields Equations 4–25 and 4–26.

$$m = \left(\frac{R_2}{R_1 + R_2}\right)\left(\frac{R_F + R_G}{R_G}\right) \quad (4–25)$$

$$b = V_{REF}\left(\frac{R_1}{R_1 + R_2}\right)\left(\frac{R_F + R_G}{R_G}\right) \quad (4–26)$$

Example; the circuit specifications are V_{OUT} = 1 V at V_{IN} = 0.01 V, V_{OUT} = 4.5 V at V_{IN} = 1 V, R_L = 10 k, five percent resistor tolerances, and V_{CC} = 5 V. No reference voltage is available, thus V_{CC} is used for the reference input, and V_{REF} = 5 V. A reference voltage source is left out of the design as a space and cost savings measure, and it sacrifices noise performance, accuracy, and stability performance. Cost is an important specification, but the V_{CC} supply must be specified well enough to do the job. Each step in the subsequent design procedure is included in this analysis to ease learning and increase boredom. Many steps are skipped when subsequent cases are analyzed.

The data is substituted into simultaneous equations.

$$1 = m(0.01) + b \quad (4–27)$$

$$4.5 = m(1.0) + b \quad (4–28)$$

Equation 4–27 is multiplied by 100 (Equation 4–29) and Equation 4–28 is subtracted from Equation 4–29 to obtain Equation 4–30.

$$100 = m(1.0) + 100b \quad (4–29)$$

$$b = \frac{95.5}{99} = 0.9646 \quad (4–30)$$

The slope of the transfer function, m, is obtained by substituting b into Equation 4–27.

$$m = \frac{1–b}{0.01} = \frac{1–0.9646}{0.01} = 3.535 \quad (4–31)$$

Now that b and m are calculated, the resistor values can be calculated. Equations 4–25 and 4–26 are solved for the quantity $(R_F + R_G)/R_G$, and then they are set equal in Equation 4–32 thus yielding Equation 4–33.

$$\frac{R_F + R_G}{R_G} = m\left(\frac{R_1 + R_2}{R_2}\right) = \frac{b}{V_{CC}}\left(\frac{R_1 + R_2}{R_1}\right) \quad (4–32)$$

$$R_2 = \frac{3.535}{\frac{0.9646}{5}} R_1 = 18.316 R_1 \quad (4–33)$$

Five percent tolerance resistors are specified for this design, so we choose R_1 = 10 kΩ, and that sets the value of R_2 = 183.16 kΩ. The closest 5% resistor value to 183.16 kΩ is 180 kΩ; therefore, select R_1 = 10 kΩ and R_2 = 180 kΩ. Being forced to yield to reality by choosing standard resistor values means that there is an error in the circuit transfer function because m and b are not exactly the same as calculated. The real world constantly forces compromises into circuit design, but the good circuit designer accepts the challenge and throws money or brains at the challenge. Resistor values closer to the calculated values could be selected by using 1% or 0.5% resistors, but that selection increases cost and violates the design specification. The cost increase is hard to justify except in precision circuits. Using ten-cent resistors with a ten-cent op amp usually is false economy.

The left half of Equation 4–32 is used to calculate R_F and R_G.

$$\frac{R_F + R_G}{R_G} = m\left(\frac{R_1 + R_2}{R_2}\right) = 3.535\left(\frac{180 + 10}{180}\right) = 3.73 \quad (4–34)$$

$$R_F = 2.73 R_G \quad (4–35)$$

The resulting circuit equation is given below.

$$V_{OUT} = 3.5 V_{IN} + 0.97 \quad (4–36)$$

The gain setting resistor, R_G, is selected as 10 kΩ, and 27 kΩ, the closest 5% standard value is selected for the feedback resistor, R_F. Again, there is a slight error involved with standard resistor values. This circuit must have an output voltage swing from 1 V to 4.5 V. The older op amps can not be used in this circuit because they lack dynamic range, so the TLV247X family of op amps is selected. The data shown in Figure 4–7 confirms the op amp selection because there is little error. The circuit with the selected component values is shown in Figure 4–11. The circuit was built with the specified components, and the transfer curve is shown in Figure 4–12.

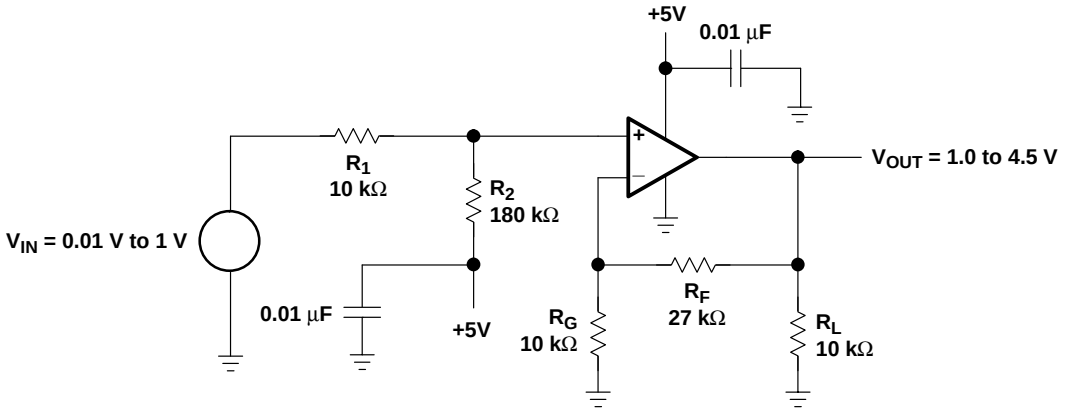


Figure 4–11. Case 1 Example Circuit

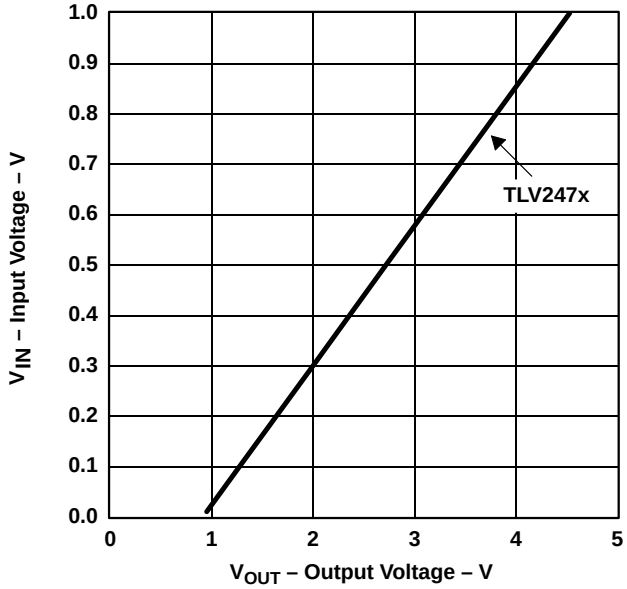


Figure 4–12. Case 1 Example Circuit Measured Transfer Curve

The transfer curve shown is a straight line, and that means that the circuit is linear. The V_{OUT} intercept is about 0.98 V rather than 1 V as specified, and this is excellent performance considering that the components were selected randomly from bins of resistors. Different sets of components would have slightly different slopes because of the resistor tolerances. The TLV247X has input bias currents and input offset voltages, but the effect of these errors is hard to measure on the scale of the output voltage. The output voltage

measured 4.53 V when the input voltage was 1 V. Considering the low and high input voltage errors, it is safe to conclude that the resistor tolerances have skewed the gain slightly, but this is still excellent performance for 5% components. Often lab data similar to that shown here is more accurate than the 5% resistor tolerance, but do not fall into the trap of expecting this performance, because you will be disappointed if you do.

The resistors were selected in the k-Ω range arbitrarily. The gain and offset specifications determine the resistor ratios, but supply current, frequency response, and op amp drive capability determine their absolute values. The resistor value selection in this design is high because modern op amps do not have input current offset problems, and they yield reasonable frequency response. If higher frequency response is demanded, the resistor values must decrease, and resistor value decreases reduce input current errors, while supply current increases. When the resistor values get low enough, it becomes hard for another circuit, or possibly the op amp, to drive the resistors.

4.3.2 Case 2: $V_{OUT} = +mV_{IN} - b$

The circuit shown in Figure 4–13 yields a solution for Case 2. The circuit equation is obtained by taking the Thevenin equivalent circuit looking into the junction of R_1 and R_2. After the R_1, R_2 circuit is replaced with the Thevenin equivalent circuit, the gain is calculated with the ideal gain equation (Equation 4–37).

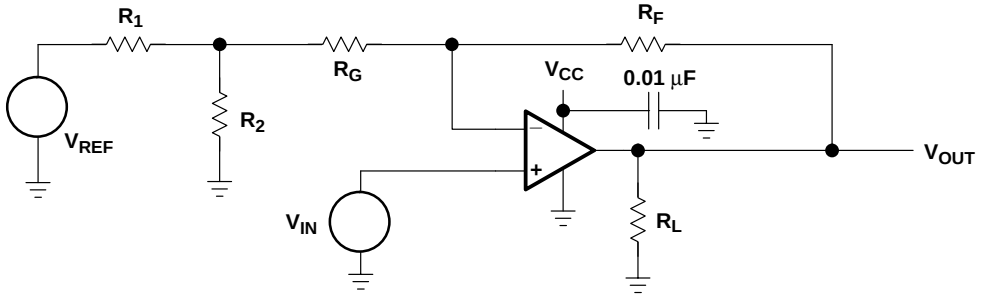


Figure 4–13. Schematic for Case 2: $V_{OUT} = +mV_{IN} - b$

$$V_{OUT} = V_{IN}\left(\frac{R_F + R_G + R_1 \parallel R_2}{R_G + R_1 \parallel R_2}\right) - V_{REF}\left(\frac{R_2}{R_1 + R_2}\right)\left(\frac{R_F}{R_G + R_1 \parallel R_2}\right) \qquad (4\text{–}37)$$

Comparing terms in Equations 4–37 and 4–14 enables the extraction of m and b.

$$m = \frac{R_F + R_G + R_1 \parallel R_2}{R_G + R_1 \parallel R_2} \tag{4–38}$$

$$|b| = V_{REF}\left(\frac{R_2}{R_1 + R_2}\right)\left(\frac{R_F}{R_G + R_1 \parallel R_2}\right) \tag{4–39}$$

The specifications for an example design are: V_{OUT} = 1.5 V @ V_{IN} = 0.2 V, V_{OUT} = 4.5 V @ V_{IN} = 0.5 V, V_{REF} = V_{CC} = 5 V, R_L = 10 kΩ, and 5% resistor tolerances. The simultaneous equations, (Equations 4–40 and 4–41), are written below.

$$1.5 = 0.2m + b \tag{4–40}$$

$$4.5 = 0.5m + b \tag{4–41}$$

From these equations we find that b = -0.5 and m = 10. Making the assumption that $R_1||R_2 << R_G$ simplifies the calculations of the resistor values.

$$m = 10 = \frac{R_F + R_G}{R_G} \tag{4–42}$$

$$R_F = 9R_G \tag{4–43}$$

Let R_G = 20 kΩ, and then R_F = 180 kΩ.

$$b = V_{CC}\left(\frac{R_F}{R_G}\right)\left(\frac{R_2}{R_1 + R_2}\right) = 5\left(\frac{180}{20}\right)\left(\frac{R_2}{R_1 + R_2}\right) \tag{4–44}$$

$$R_1 = \frac{1–0.01111}{0.01111}R_2 = 89R_2 \tag{4–45}$$

Select R_2 = 0.82 kΩ and R_1 equals 72.98 kΩ. Since 72.98 kΩ is not a standard 5% resistor value, R_1 is selected as 75 kΩ. The difference between the selected and calculated value of R_1 has about a 3% effect on b, and this error shows up in the transfer function as an intercept rather than a slope error. The parallel resistance of R_1 and R_2 is approximately 0.82 kΩ and this is much less than R_G, which is 20 kΩ, thus the earlier assumption that $R_G >> R1||R2$ is justified. R_2 could have been selected as a smaller value, but the smaller values yielded poor standard 5% values for R_1. The final circuit is shown in Figure 4–14 and the measured transfer curve for this circuit is shown in Figure 4–15.

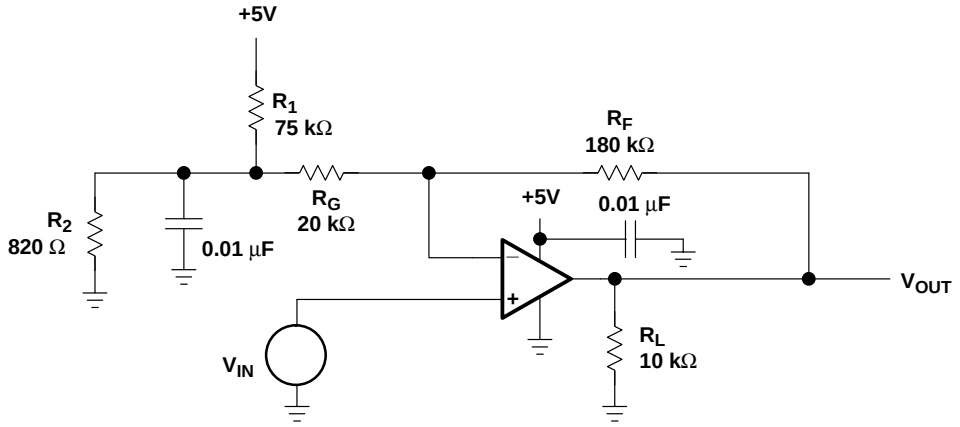


Figure 4–14. Case 2 Example Circuit

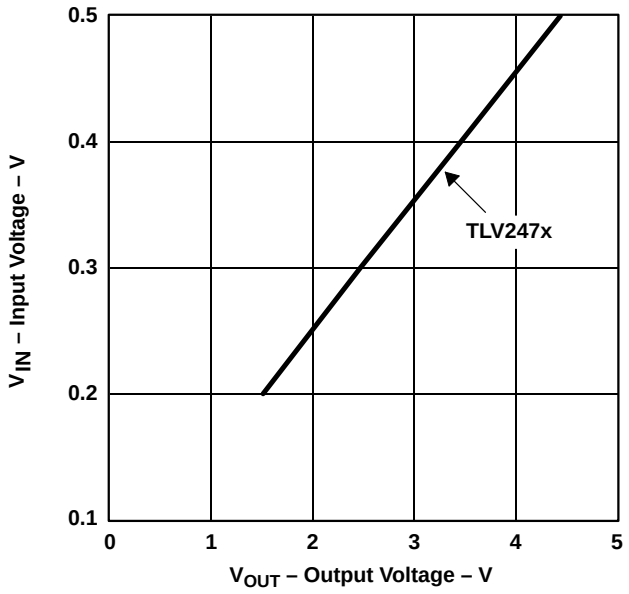


Figure 4–15. Case 2 Example Circuit Measured Transfer Curve

The TLV247X was used to build the test circuit because of its wide dynamic range. The transfer curve plots very close to the theoretical curve; the direct result of using a high performance op amp.

4.3.3 Case 3: $V_{OUT} = -mV_{IN} + b$

The circuit shown in Figure 4–16 yields the transfer function desired for Case 3.

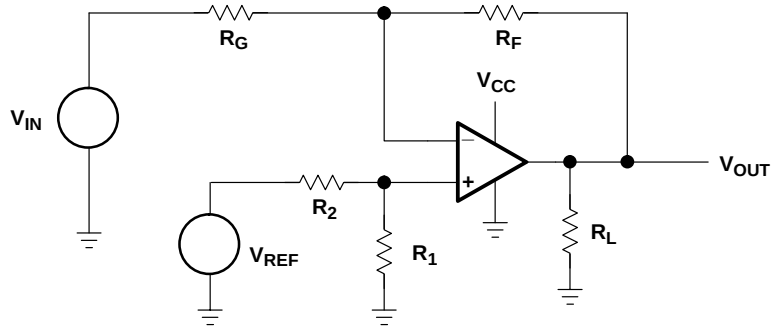


Figure 4–16. Schematic for Case 3: $V_{OUT} = -mV_{IN} + b$

The circuit equation is obtained with superposition.

$$V_{OUT} = -V_{IN}\left(\frac{R_F}{R_G}\right) + V_{REF}\left(\frac{R_1}{R_1 + R_2}\right)\left(\frac{R_F + R_G}{R_G}\right) \quad (4–46)$$

Comparing terms between Equations 4–45 and 4–15 enables the extraction of m and b.

$$|m| = \frac{R_F}{R_G} \quad (4–47)$$

$$b = V_{REF}\left(\frac{R_1}{R_1 + R_2}\right)\left(\frac{R_F + R_G}{R_G}\right) \quad (4–48)$$

The design specifications for an example circuit are: V_{OUT} = 1 V @ V_{IN} = -0.1 V, V_{OUT} = 6 V @ V_{IN} = -1 V, $V_{REF} = V_{CC}$ = 10 V, R_L = 100 Ω, and 5% resistor tolerances. The supply voltage available for this circuit is 10 V, and this exceeds the maximum allowable supply voltage for the TLV247X. Also, this circuit must drive a back-terminated cable that looks like two 50-Ω resistors connected in series, thus the op amp must be able to drive 6/100 = 60 mA. The stringent op amp selection criteria limits the choice to relatively new op amps if ideal op amp equations are going to be used. The TLC07X has excellent single-supply input performance coupled with high output current drive capability, so it is selected for this circuit. The simultaneous equations (Equations 4–49 and 4–50), are written below.

$$1 = (-0.1)m + b \quad (4–49)$$

$$6 = (-1)m + b \quad (4–50)$$

From these equations we find that b = 0.444 and m = –5.6.

$$|m| = 5.56 = \frac{R_F}{R_G} \tag{4–51}$$

$$R_F = 5.56R_G \tag{4–52}$$

Let R_G = 10 kΩ, and then R_F = 56.6 kΩ, which is not a standard 5% value, hence R_F is selected as 56 kΩ.

$$b = V_{CC}\left(\frac{R_F + R_G}{R_G}\right)\left(\frac{R_1}{R_1 + R_2}\right) = 10\left(\frac{56 + 10}{10}\right)\left(\frac{R_1}{R_1 + R_2}\right) \tag{4–53}$$

$$R_2 = \frac{66–0.4444}{0.4444}R_1 = 147.64R_1 \tag{4–54}$$

The final equation for the example is given below

$$V_{OUT} = -5.56V_{IN} + 0.444 \tag{4–55}$$

Select R_1 = 2 kΩ and R_2 = 295.28 kΩ. Since 295.28 kΩ is not a standard 5% resistor value, R_1 is selected as 300 kΩ. The difference between the selected and calculated value of R_1 has a nearly insignificant effect on b. The final circuit is shown in Figure 4–17, and the measured transfer curve for this circuit is shown in Figure 4–18.

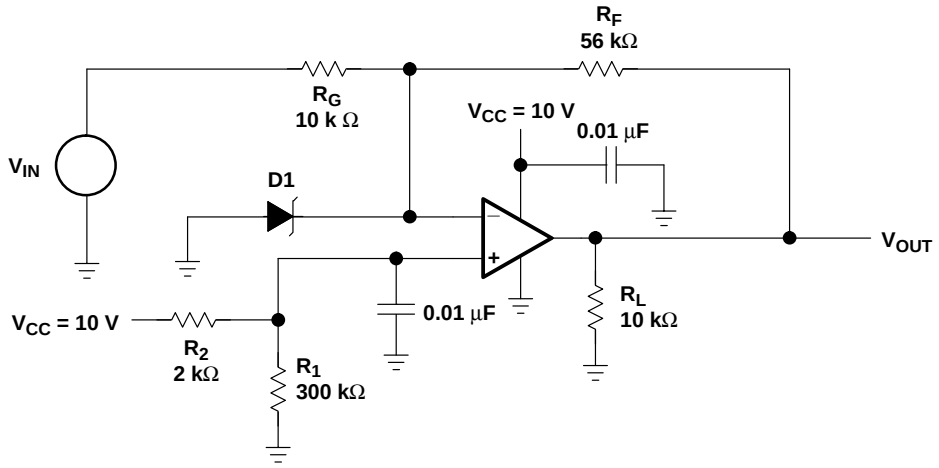


Figure 4–17. Case 3 Example Circuit

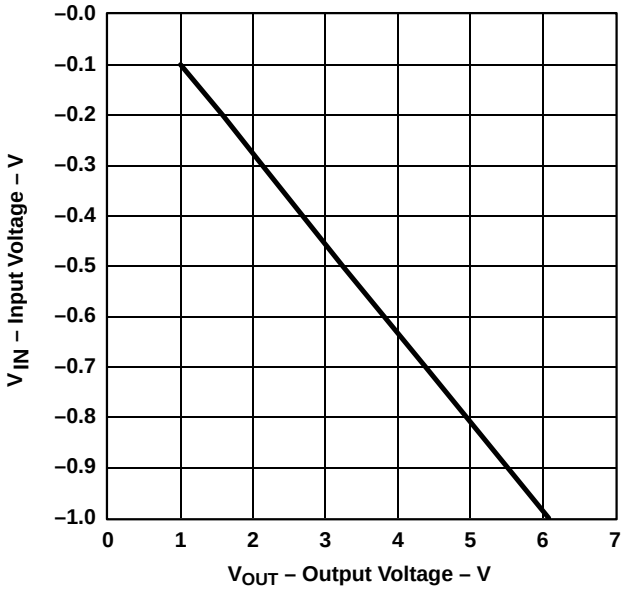


Figure 4–18. Case 3 Example Circuit Measured Transfer Curve

As long as the circuit works normally, there are no problems handling the negative voltage input to the circuit, because the inverting lead of the TLC07X is at a positive voltage. The positive op amp input lead is at a voltage of approximately 65 mV, and normal op amp operation keeps the inverting op amp input lead at the same voltage because of the assumption that the error voltage is zero. When V_{CC} is powered down while there is a negative voltage on the input circuit, most of the negative voltage appears on the inverting op amp input lead.

The most prudent solution is to connect the diode, D_1, with its cathode on the inverting op amp input lead and its anode at ground. If a negative voltage gets on the inverting op amp input lead, it is clamped to ground by the diode. Select the diode type as germanium or Schottky so the voltage drop across the diode is about 200 mV; this small voltage does not harm most op amp inputs. As a further precaution, R_G can be split into two resistors with the diode inserted at the junction of the two resistors. This places a current limiting resistor between the diode and the inverting op amp input lead.

4.3.4 Case 4: $V_{OUT} = -mV_{IN} - b$

The circuit shown in Figure 4–19 yields a solution for Case 4. The circuit equation is obtained by using superposition to calculate the response to each input. The individual responses to V_{IN} and V_{REF} are added to obtain Equation 4–56.

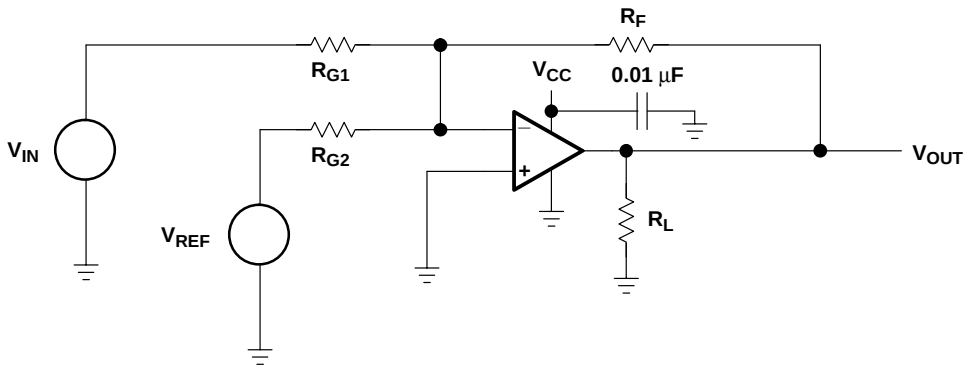


Figure 4–19. Schematic for Case 4: $V_{OUT} = -mV_{IN} - b$

$$V_{OUT} = -V_{IN}\frac{R_F}{R_{G1}} - V_{REF}\frac{R_F}{R_{G2}} \tag{4–56}$$

Comparing terms in Equations 4–56 and 4–16 enables the extraction of m and b.

$$|m| = \frac{R_F}{R_{G1}} \tag{4–57}$$

$$|b| = V_{REF}\frac{R_F}{R_{G2}} \tag{4–58}$$

The design specifications for an example circuit are: V_{OUT} = 1 V @ V_{IN} = –0.1 V, V_{OUT} = 5 V @ V_{IN} =– 0.3 V, $V_{REF} = V_{CC}$ = 5 V, R_L = 10 kΩ, and 5% resistor tolerances. The simultaneous Equations 4–59 and 4–60, are written below.

$$1 = (-0.1)m + b \tag{4–59}$$

$$5 = (-0.3)m + b \tag{4–60}$$

From these equations we find that b = –1 and m = –20. Setting the magnitude of m equal to Equation 4–57 yields Equation 4–61.

$$|m| = 20 = \frac{R_F}{R_{G1}} \tag{4–61}$$

$$R_F = 20R_{G1} \tag{4–62}$$

Let R_{G1} = 1 kΩ, and then R_F = 20 kΩ.

$$|b| = V_{CC}\left(\frac{R_F}{R_{G1}}\right) = 5\left(\frac{R_F}{R_{G2}}\right) = 1 \quad (4–63)$$

$$R_{G2} = \frac{R_F}{0.2} = \frac{20}{0.2} = 100\ k\Omega \quad (4–64)$$

The final equation for this example is given in Equation 4–63.

$$V_{OUT} = -20V_{IN} - 1 \quad (4–65)$$

The final circuit is shown in Figure 4–20 and the measured transfer curve for this circuit is shown in Figure 4–21.

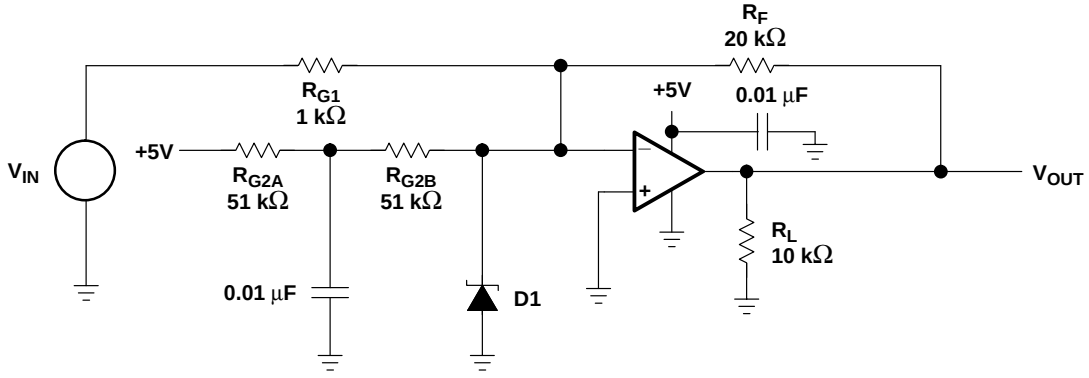


Figure 4–20. Case 4 Example Circuit

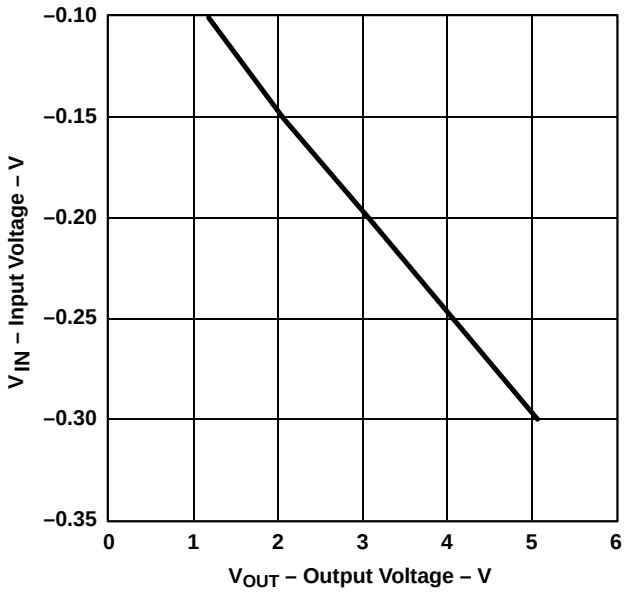


Figure 4–21. Case 4 Example Circuit Measured Transfer Curve

The TLV247X was used to build the test circuit because of its wide dynamic range. The transfer curve plots very close to the theoretical curve, and this results from using a high performance op amp.

As long as the circuit works normally there are no problems handling the negative voltage input to the circuit because the inverting lead of the TLV247X is at a positive voltage. The positive op amp input lead is grounded, and normal op amp operation keeps the inverting op amp input lead at ground because of the assumption that the error voltage is zero. When V_{CC} is powered down while there is a negative voltage on the inverting op amp input lead.

The most prudent solution is to connect the diode, D_1, with its cathode on the inverting op amp input lead and its anode at ground. If a negative voltage gets on the inverting op amp input lead it is clamped to ground by the diode. Select the diode type as germanium or Schottky so the voltage drop across the diode is about 200 mV; this small voltage does not harm most op amp inputs. R_{G2} is split into two resistors ($R_{G2A} = R_{G2B} = 51\ k\Omega$) with a capacitor inserted at the junction of the two resistors. This places a power supply filter in series with V_{CC}.

4.4 Summary

Single-supply op amp design is more complicated than split-supply op amp design, but with a logical design approach excellent results are achieved. Single-supply design used to be considered technically limiting because older op amps had limited capability. The new op amps, such as the TLC247X, TLC07X, and TLC08X have excellent single-supply parameters; thus when used in the correct applications these op amps yield rail-to-rail performance equal to their split-supply counterparts.

Single-supply op amp design usually involves some form of biasing, and this requires more thought, so single-supply op amp design needs discipline and a procedure. The recommended design procedure for single-supply op amp design is:

- Substitute the specification data into simultaneous equations to obtain m and b (the slope and intercept of a straight line).
- Let m and b determine the form of the circuit.
- Choose the circuit configuration that fits the form.
- Using the circuit equations for the circuit configuration selected, calculate the resistor values.
- Build the circuit, take data, and verify performance.
- Test the circuit for nonstandard operating conditions (circuit power off while interface power is on, over/under range inputs, etc.).
- Add protection components as required.
- Retest.

When this procedure is followed, good results follow. As single-supply circuit designers expand their horizon, new challenges require new solutions. Remember, the only equation a linear op amp can produce is the equation of a straight line. That equation only has four forms. The new challenges may consist of multiple inputs, common-mode voltage rejection, or something different, but this method can be expanded to meet these challenges.

Chapter 5

Feedback and Stability Theory

Ron Mancini

5.1 Why Study Feedback Theory?

The gain of all op amps decreases as frequency increases, and the decreasing gain results in decreasing accuracy as the ideal op amp assumption ($a \Rightarrow \infty$) breaks down. In most real op amps the open loop gain starts to decrease before 10 Hz, so an understanding of feedback is required to predict the closed loop performance of the op amp. The real world application of op amps is feedback controlled, and depends on op amp open loop gain at a given frequency. A designer must know theory to be able to predict the circuit response regardless of frequency or open loop gain.

Analysis tools have something in common with medicine because they both can be distasteful but necessary. Medicine often tastes bad or has undesirable side effects, and analysis tools involve lots of hard learning work before they can be applied to yield results. Medicine assists the body in fighting an illness; analysis tools assist the brain in learning/designing feedback circuits.

The analysis tools given here are a synopsis of salient points; thus they are detailed enough to get you where you are going without any extras. The references, along with thousands of their counterparts, must be consulted when making an in-depth study of the field. Aspirin, home remedies, and good health practice handle the majority of health problems, and these analysis tools solve the majority of circuit problems.

Ideal op amp circuits can be designed without knowledge of feedback analysis tools, but these circuits are limited to low frequencies. Also, an understanding of feedback analysis tools is required to understand AC effects like ringing and oscillations.

5.2 Block Diagram Math and Manipulations

Electronic systems and circuits are often represented by block diagrams, and block diagrams have a unique algebra and set of transformations[1]. Block diagrams are used because they are a shorthand pictorial representation of the cause-and-effect relationship between the input and output in a real system. They are a convenient method for characterizing the functional relationships between components. It is not necessary to understand the functional details of a block to manipulate a block diagram.

The input impedance of each block is assumed to be infinite to preclude loading. Also, the output impedance of each block is assumed to be zero to enable high fan-out. The systems designer sets the actual impedance levels, but the fan-out assumption is valid because the block designers adhere to the system designer's specifications. All blocks multiply the input times the block quantity (see Figure 5–1) unless otherwise specified within the block. The quantity within the block can be a constant as shown in Figure 5–1(c), or it can be a complex math function involving Laplace transforms. The blocks can perform time-based operations such as differentiation and integration.

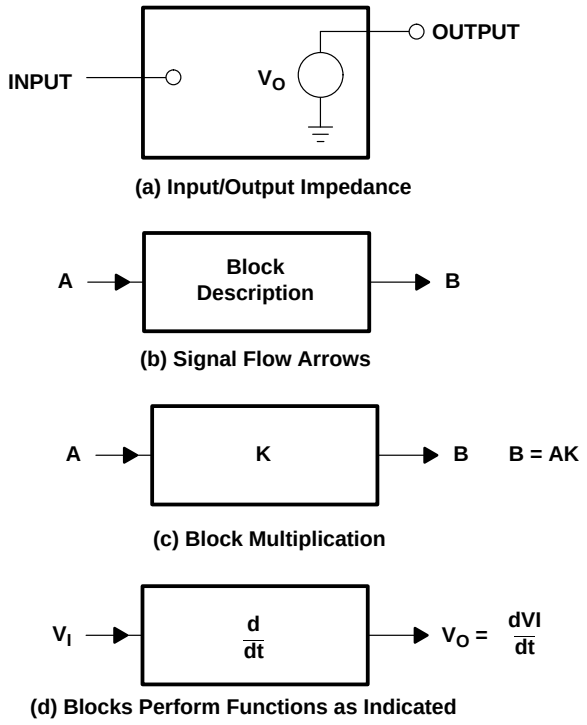


Figure 5–1. Definition of Blocks

Adding and subtracting are done in special blocks called summing points. Figure 5–2 gives several examples of summing points. Summing points can have unlimited inputs, can add or subtract, and can have mixed signs yielding addition and subtraction within a single summing point. Figure 5–3 defines the terms in a typical control system, and Figure 5–4 defines the terms in a typical electronic feedback system. Multiloop feedback systems (Figure 5–5) are intimidating, but they can be reduced to a single loop feedback system, as shown in the figure, by writing equations and solving for V_{OUT}/V_{IN}. An easier method for reducing multiloop feedback systems to single loop feedback systems is to follow the rules and use the transforms given in Figure 5–6.

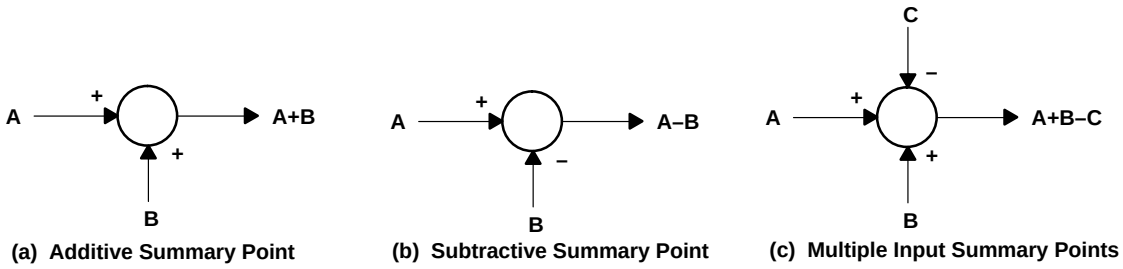


Figure 5–2. Summary Points

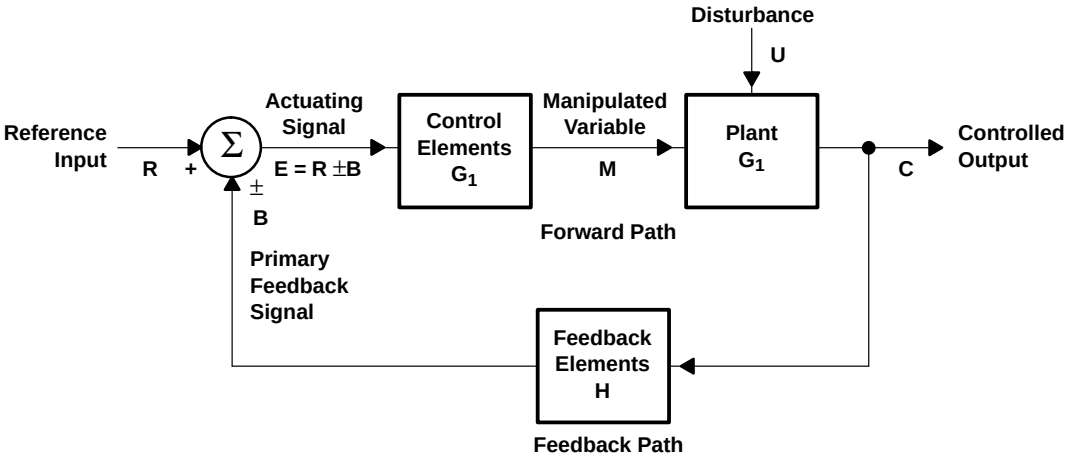


Figure 5–3. Definition of Control System Terms

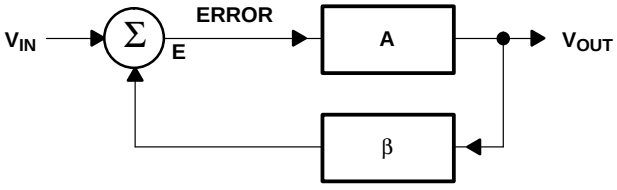


Figure 5–4. Definition of an Electronic Feedback Circuit

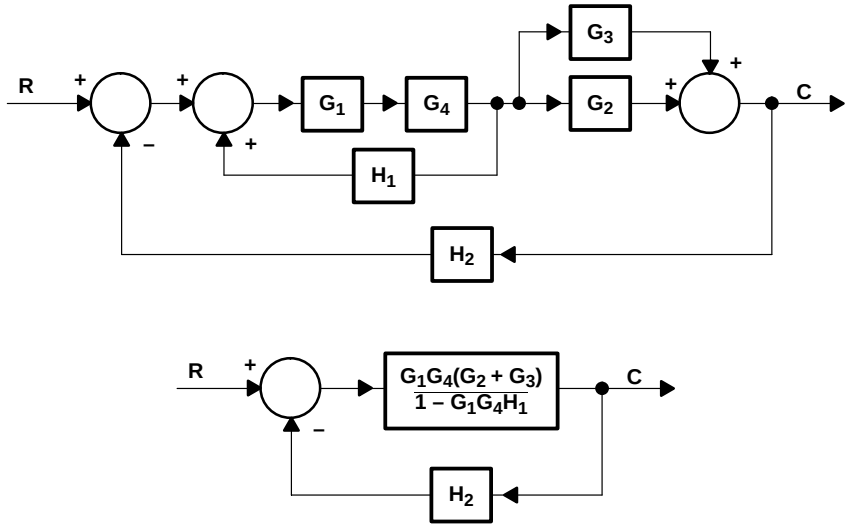


Figure 5–5. Multiloop Feedback System

Block diagram reduction rules:

- Combine cascade blocks.
- Combine parallel blocks.
- Eliminate interior feedback loops.
- Shift summing points to the left.
- Shift takeoff points to the right.
- Repeat until canonical form is obtained.

Figure 5–6 gives the block diagram transforms. The idea is to reduce the diagram to its canonical form because the canonical feedback loop is the simplest form of a feedback loop, and its analysis is well documented. All feedback systems can be reduced to the canonical form, so all feedback systems can be analyzed with the same math. A canonical loop exists for each input to a feedback system; although the stability dynamics are independent of the input, the output results are input dependent. The response of each input of a multiple input feedback system can be analyzed separately and added through superposition.

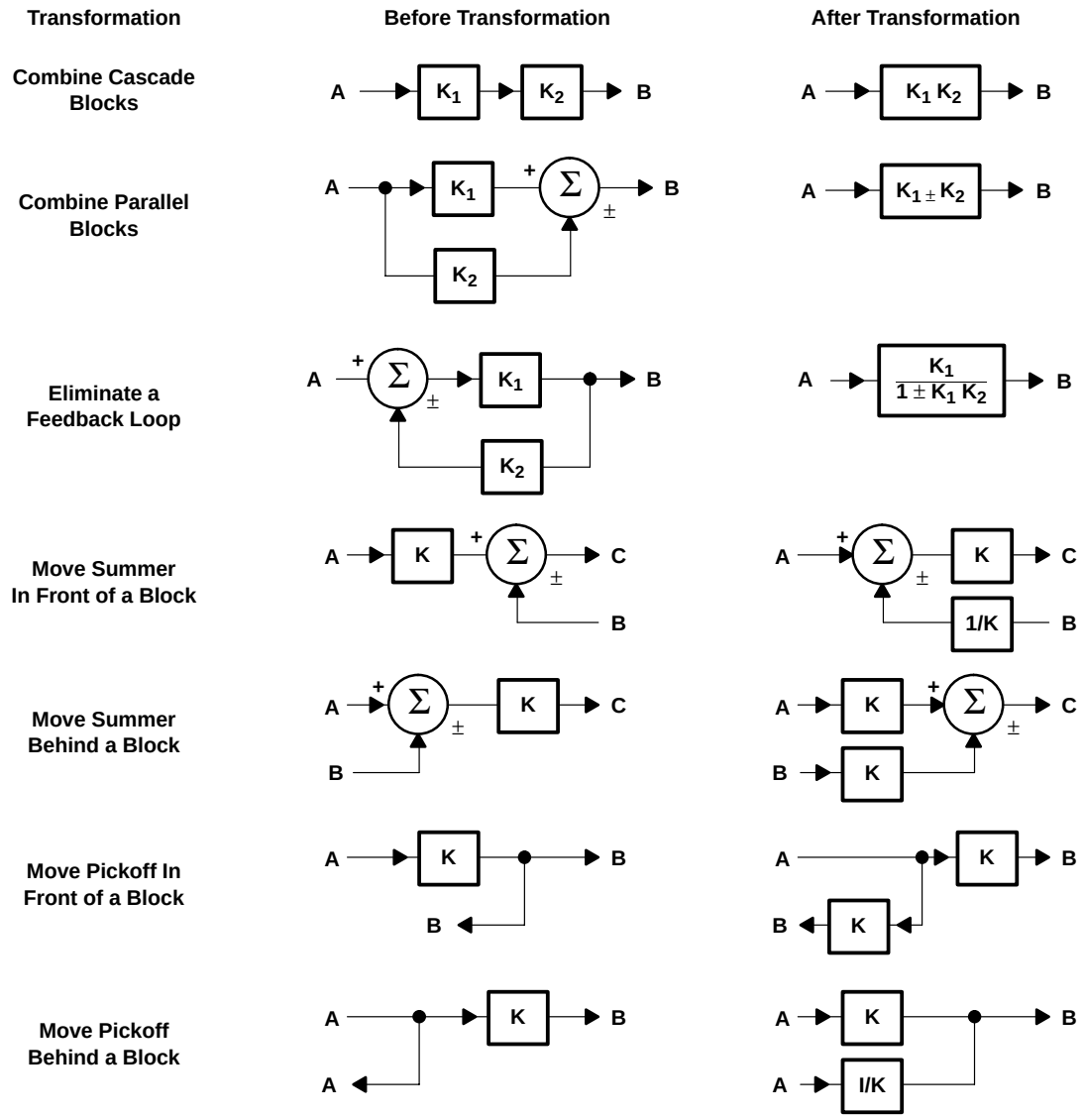


Figure 5–6. Block Diagram Transforms

5.3 Feedback Equation and Stability

Figure 5–7 shows the canonical form of a feedback loop with control system and electronic system terms. The terms make no difference except that they have meaning to the system engineers, but the math does have meaning, and it is identical for both types of terms. The electronic terms and negative feedback sign are used in this analysis, because subsequent chapters deal with electronic applications. The output equation is written in Equation 5–1.

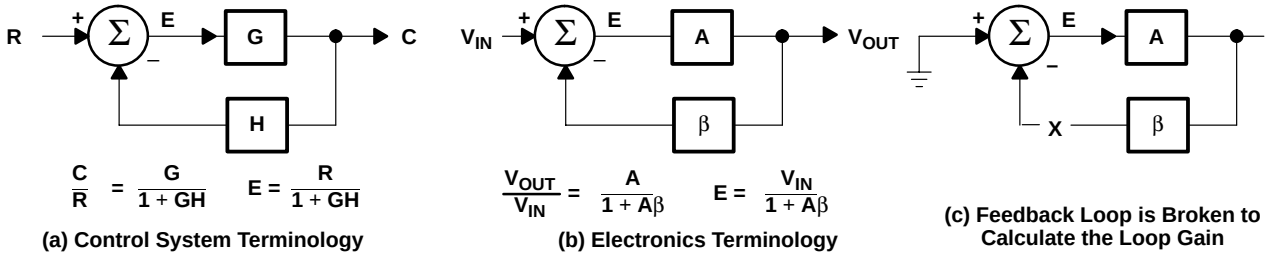


Figure 5–7. Comparison of Control and Electronic Canonical Feedback Systems

$$V_{OUT} = EA \tag{5–1}$$

The error equation is written in Equation 5–2.

$$E = V_{IN} - \beta V_{OUT} \tag{5–2}$$

Combining Equations 5–1 and 5–2 yields Equation 5–3.

$$\frac{V_{OUT}}{A} = V_{IN} - \beta V_{OUT} \tag{5–3}$$

Collecting terms yields Equation 5–4.

$$V_{OUT}\left(\frac{1}{A} + \beta\right) = V_{IN} \tag{5–4}$$

Rearranging terms yields the classic form of the feedback Equation 5–5.

$$\frac{V_{OUT}}{V_{IN}} = \frac{A}{1 + A\beta} \tag{5–5}$$

When the quantity $A\beta$ in Equation 5–5 becomes very large with respect to one, the one can be neglected, and Equation 5–5 reduces to Equation 5–6, which is the ideal feedback equation. Under the conditions that $A\beta >> 1$, the system gain is determined by the feedback factor β. Stable passive circuit components are used to implement the feedback factor, thus in the ideal situation, the closed loop gain is predictable and stable because β is predictable and stable.

$$\frac{V_{OUT}}{V_{IN}} = \frac{1}{\beta} \tag{5–6}$$

The quantity Aβ is so important that it has been given a special name: loop gain. In Figure 5–7, when the voltage inputs are grounded (current inputs are opened) and the loop is broken, the calculated gain is the loop gain, Aβ. Now, keep in mind that we are using complex numbers, which have magnitude and direction. When the loop gain approaches minus one, or to express it mathematically $1\angle -180°$, Equation 5–5 approaches $1/0 \Rightarrow \propto$. The circuit output heads for infinity as fast as it can using the equation of a straight line. If the output were not energy limited, the circuit would explode the world, but happily, it is energy limited, so somewhere it comes up against the limit.

Active devices in electronic circuits exhibit nonlinear phenomena when their output approaches a power supply rail, and the nonlinearity reduces the gain to the point where the loop gain no longer equals $1\angle -180°$. Now the circuit can do two things: first it can become stable at the power supply limit, or second, it can reverse direction (because stored charge keeps the output voltage changing) and head for the negative power supply rail.

The first state where the circuit becomes stable at a power supply limit is named lockup; the circuit will remain in the locked up state until power is removed and reapplied. The second state where the circuit bounces between power supply limits is named oscillatory. Remember, the loop gain, Aβ, is the sole factor determining stability of the circuit or system. Inputs are grounded or disconnected, so they have no bearing on stability.

Equations 5–1 and 5–2 are combined and rearranged to yield Equation 5–7, which is the system or circuit error equation.

$$E = \frac{V_{IN}}{1 + A\beta} \tag{5–7}$$

First, notice that the error is proportional to the input signal. This is the expected result because a bigger input signal results in a bigger output signal, and bigger output signals require more drive voltage. As the loop gain increases, the error decreases, thus large loop gains are attractive for minimizing errors.

5.4 Bode Analysis of Feedback Circuits

H. W. Bode developed a quick, accurate, and easy method of analyzing feedback amplifiers, and he published a book about his techniques in 1945.[2] Operational amplifiers had not been developed when Bode published his book, but they fall under the general classification of feedback amplifiers, so they are easily analyzed with Bode techniques. The mathematical manipulations required to analyze a feedback circuit are complicated because they involve multiplication and division. Bode developed the Bode plot, which simplifies the analysis through the use of graphical techniques.

The Bode equations are log equations that take the form 20LOG(F(t)) = 20LOG(|F(t)|) + phase angle. Terms that are normally multiplied and divided can now be added and subtracted because they are log equations. The addition and subtraction is done graphically, thus easing the calculations and giving the designer a pictorial representation of circuit performance. Equation 5–8 is written for the low pass filter shown in Figure 5–8.

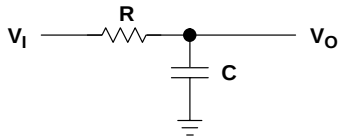


Figure 5–8. Low-Pass Filter

$$\frac{V_{OUT}}{V_{IN}} = \frac{\frac{1}{C\,s}}{R + \frac{1}{C\,s}} = \frac{1}{1 + RCs} = \frac{1}{1 + \tau s} \qquad (5\text{–}8)$$

Where: $s = j\omega$, $j = \sqrt{(-1)}$, and $RC = \tau$

The magnitude of this transfer function is $|V_{OUT}/V_{IN}| = 1/\sqrt{1^2 + (\tau\omega)^2}$. This magnitude, $|V_{OUT}/V_{IN}| \cong 1$ when $\omega = 0.1/\tau$, it equals 0.707 when $\omega = 1/\tau$, and it is approximately = 0.1 when $\omega = 10/\tau$. These points are plotted in Figure 5–9 using straight line approximations. The negative slope is –20 dB/decade or –6 dB/octave. The magnitude curve is plotted as a horizontal line until it intersects the breakpoint where $\omega = 1/\tau$. The negative slope begins at the breakpoint because the magnitude starts decreasing at that point. The gain is equal to 1 or 0 dB at very low frequencies, equal to 0.707 or –3 dB at the break frequency, and it keeps falling with a –20 dB/decade slope for higher frequencies.

The phase shift for the low pass filter or any other transfer function is calculated with the aid of Equation 5–9.

$$\phi = \text{tangent}^{-1}\left(\frac{\text{Real}}{\text{Imaginary}}\right) = -\text{tangent}^{-1}\left(\frac{\omega\tau}{1}\right) \qquad (5\text{–}9)$$

The phase shift is much harder to approximate because the tangent function is nonlinear. Normally the phase information is only required around the 0 dB intercept point for an active circuit, so the calculations are minimized. The phase is shown in Figure 5–9, and it is approximated by remembering that the tangent of 90° is 1, the tangent of 60° is $\sqrt{3}$, and the tangent of 30° is $\sqrt{3}/3$.

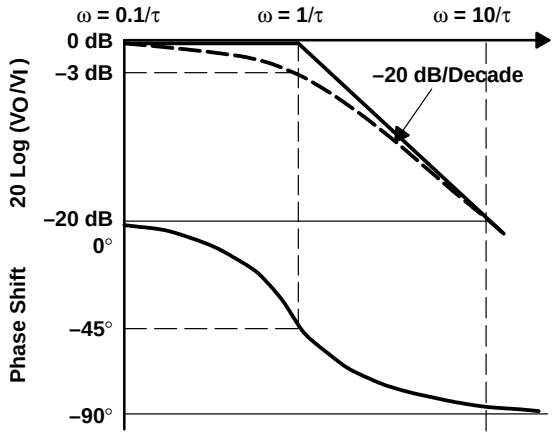


Figure 5–9. Bode Plot of Low-Pass Filter Transfer Function

A breakpoint occurring in the denominator is called a pole, and it slopes down. Conversely, a breakpoint occurring in the numerator is called a zero, and it slopes up. When the transfer function has multiple poles and zeros, each pole or zero is plotted independently, and the individual poles/zeros are added graphically. If multiple poles, zeros, or a pole/zero combination have the same breakpoint, they are plotted on top of each other. Multiple poles or zeros cause the slope to change by multiples of 20 dB/decade.

An example of a transfer function with multiple poles and zeros is a band reject filter (see Figure 5–10). The transfer function of the band reject filter is given in Equation 5–10.

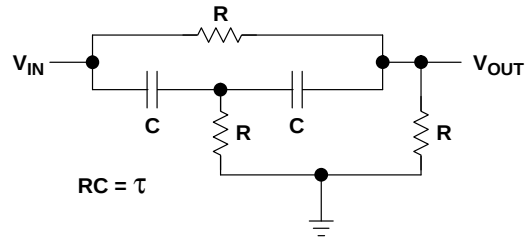


Figure 5–10. Band Reject Filter

$$G = \frac{V_{OUT}}{V_{IN}} = \frac{(1 + \tau s)(1 + \tau s)}{2\left(1 + \frac{\tau s}{0.44}\right)\left(1 + \frac{\tau s}{4.56}\right)} \qquad (5–10)$$

The pole zero plot for each individual pole and zero is shown in Figure 5–11, and the combined pole zero plot is shown in Figure 5–12.

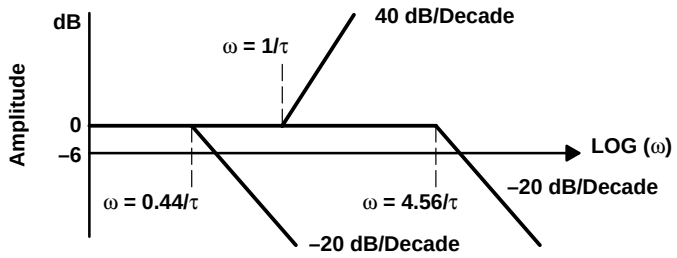


Figure 5–11. Individual Pole Zero Plot of Band Reject Filter

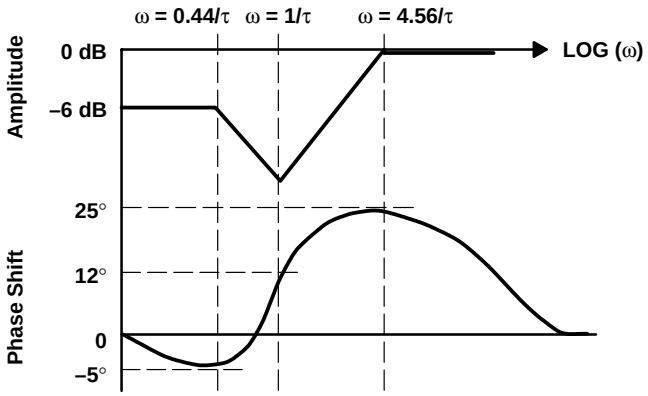


Figure 5–12. Combined Pole Zero Plot of Band Reject Filter

The individual pole zero plots show the dc gain of 1/2 plotting as a straight line from the –6 dB intercept. The two zeros occur at the same break frequency, thus they add to a 40-dB/decade slope. The two poles are plotted at their breakpoints of $\omega = 0.44/\tau$ and $\omega = 4.56/\tau$. The combined amplitude plot intercepts the amplitude axis at –6 dB because of the dc gain, and then breaks down at the first pole. When the amplitude function gets to the double zero, the first zero cancels out the first pole, and the second zero breaks up. The upward slope continues until the second pole cancels out the second zero, and the amplitude is flat from that point out in frequency.

When the separation between all the poles and zeros is great, a decade or more in frequency, it is easy to draw the Bode plot. As the poles and zeros get closer together, the plot gets harder to make. The phase is especially hard to plot because of the tangent function, but picking a few salient points and sketching them in first gets a pretty good approximation.[3] The Bode plot enables the designer to get a good idea of pole zero placement, and it is valuable for fast evaluation of possible compensation techniques. When the situation gets critical, accurate calculations must be made and plotted to get an accurate result.

Consider Equation 5–11.

$$\frac{V_{OUT}}{V_{IN}} = \frac{A}{1 + A\beta} \tag{5–11}$$

Taking the log of Equation 5–11 yields Equation 5–12.

$$20\text{Log}\left(\frac{V_{OUT}}{V_{IN}}\right) = 20\text{Log}(A) - 20\text{Log}(1 + A\beta) \tag{5–12}$$

If A and β do not contain any poles or zeros there will be no break points. Then the Bode plot of Equation 5–12 looks like that shown in Figure 5–13, and because there are no poles to contribute negative phase shift, the circuit cannot oscillate.

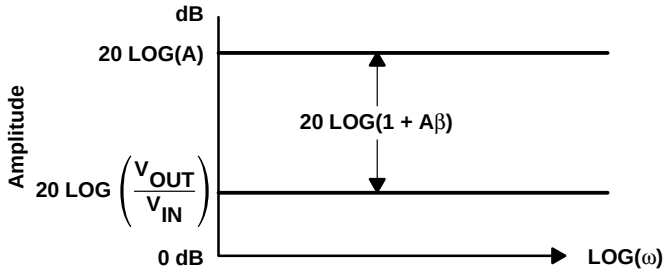


Figure 5–13. When No Pole Exists in Equation (5–12)

All real amplifiers have many poles, but they are normally internally compensated so that they appear to have a single pole. Such an amplifier would have an equation similar to that given in Equation 5–13.

$$A = \frac{a}{1 + j\frac{\omega}{\omega_a}} \tag{5–13}$$

The plot for the single pole amplifier is shown in Figure 5–14.

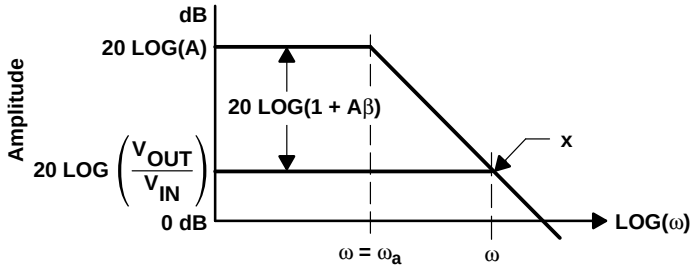


Figure 5–14. When Equation 5–12 has a Single Pole

The amplifier gain, A, intercepts the amplitude axis at 20Log(A), and it breaks down at a slope of –20 dB/decade at $\omega = \omega_a$. The negative slope continues for all frequencies greater than the breakpoint, $\omega = \omega_a$. The closed loop circuit gain intercepts the amplitude axis at $20\text{Log}(V_{OUT}/V_{IN})$, and because β does not have any poles or zeros, it is constant until its projection intersects the amplifier gain at point X. After intersection with the amplifier gain curve, the closed loop gain follows the amplifier gain because the amplifier is the controlling factor.

Actually, the closed loop gain starts to roll off earlier, and it is down 3 dB at point X. At point X the difference between the closed loop gain and the amplifier gain is –3 dB, thus according to Equation 5–12 the term $-20\text{Log}(1+A\beta) = -3$ dB. The magnitude of 3 dB is $\sqrt{2}$, hence $\sqrt{1 + (A\beta)^2} = \sqrt{2}$, and elimination of the radicals shows that $A\beta = 1$. There is a method [4] of relating phase shift and stability to the slope of the closed loop gain curves, but only the Bode method is covered here. An excellent discussion of poles, zeros, and their interaction is given by M. E Van Valkenberg,[5] and he also includes some excellent prose to liven the discussion.

5.5 Loop Gain Plots are the Key to Understanding Stability

Stability is determined by the loop gain, and when $A\beta = -1 = |1| \angle -180^\circ$ instability or oscillation occurs. If the magnitude of the gain exceeds one, it is usually reduced to one by circuit nonlinearities, so oscillation generally results for situations where the gain magnitude exceeds one.

Consider oscillator design, which depends on nonlinearities to decrease the gain magnitude; if the engineer designed for a gain magnitude of one at nominal circuit conditions, the gain magnitude would fall below one under worst case circuit conditions causing oscillation to cease. Thus, the prudent engineer designs for a gain magnitude of one under worst case conditions knowing that the gain magnitude is much more than one under optimistic conditions. The prudent engineer depends on circuit nonlinearities to reduce the gain magnitude to the appropriate value, but this same engineer pays a price of poorer distortion performance. Sometimes a design compromise is reached by putting a nonlinear component, such as a lamp, in the feedback loop to control the gain without introducing distortion.

Some high gain control systems always have a gain magnitude greater than one, but they avoid oscillation by manipulating the phase shift. The amplifier designer who pushes the amplifier for superior frequency performance has to be careful not to let the loop gain phase shift accumulate to 180°. Problems with overshoot and ringing pop up before the loop gain reaches 180° phase shift, thus the amplifier designer must keep a close eye on loop dynamics. Ringing and overshoot are handled in the next section, so preventing oscillation is emphasized in this section. Equation 5–14 has the form of many loop gain transfer functions or circuits, so it is analyzed in detail.

$$(A)\beta = \frac{(K)}{\left(1 + \tau_1(s)\right)\left(1 + \tau_2(s)\right)} \tag{5–14}$$

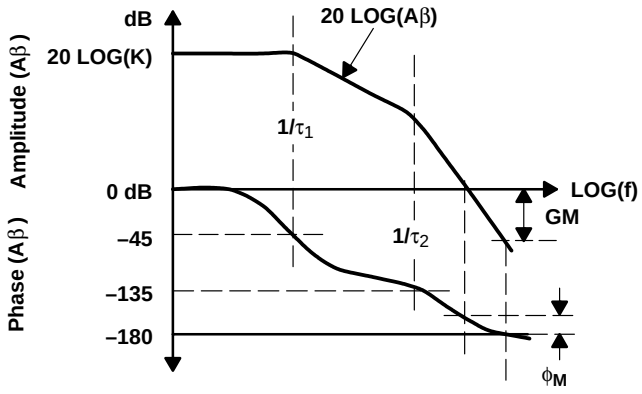


Figure 5–15. Magnitude and Phase Plot of Equation 5–14

The quantity, K, is the dc gain, and it plots as a straight line with an intercept of 20Log(K). The Bode plot of Equation 5–14 is shown in Figure 5–15. The two break points, $\omega = \omega_1 = 1/\tau_1$ and $\omega = \omega_2 = 1/\tau_2$, are plotted in the Bode plot. Each breakpoint adds –20 dB/decade slope to the plot, and 45° phase shift accumulates at each breakpoint. This transfer function is referred to as a two slope because of the two breakpoints. The slope of the curve when it crosses the 0 dB intercept indicates phase shift and the ability to oscillate. Notice that a one slope can only accumulate 90° phase shift, so when a transfer function passes through 0 dB with a one slope, it cannot oscillate. Furthermore, a two-slope system can accumulate 180° phase shift, therefore a transfer function with a two or greater slope is capable of oscillation.

A one slope crossing the 0 dB intercept is stable, whereas a two or greater slope crossing the 0 dB intercept may be stable or unstable depending upon the accumulated phase shift. Figure 5–15 defines two stability terms; the phase margin, ϕ_M, and the gain margin, G_M. Of these two terms the phase margin is much more important because phase shift is critical for stability. Phase margin is a measure of the difference in the actual phase shift and the theoretical 180° required for oscillation, and the phase margin measurement or calculation is made at the 0 dB crossover point. The gain margin is measured or calculated at the 180° phase crossover point. Phase margin is expressed mathematically in Equation 5–15.

$$\phi_M = 180 - \text{tangent}^{-1}(A\beta) \tag{5–15}$$

The phase margin in Figure 5–15 is very small, 20°, so it is hard to measure or predict from the Bode plot. A designer probably doesn't want a 20° phase margin because the system overshoots and rings badly, but this case points out the need to calculate small phase margins carefully. The circuit is stable, and it does not oscillate because the phase margin is positive. Also, the circuit with the smallest phase margin has the highest frequency response and bandwidth.

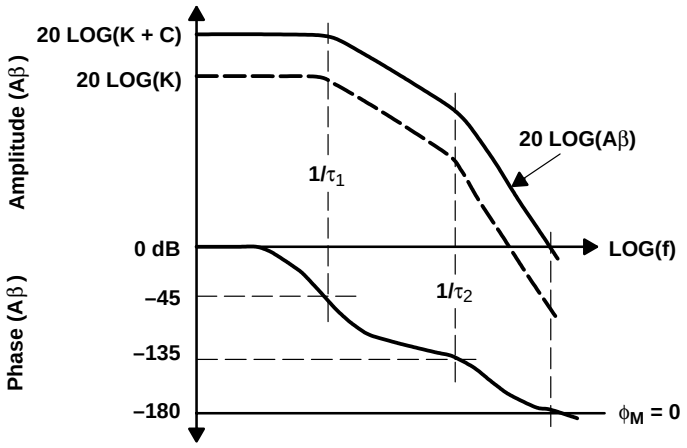


Figure 5–16. Magnitude and Phase Plot of the Loop Gain Increased to (K+C)

Increasing the loop gain to (K+C) as shown in Figure 5–16 shifts the magnitude plot up. If the pole locations are kept constant, the phase margin reduces to zero as shown, and the circuit will oscillate. The circuit is not good for much in this condition because production tolerances and worst case conditions ensure that the circuit will oscillate when you want it to amplify, and vice versa.

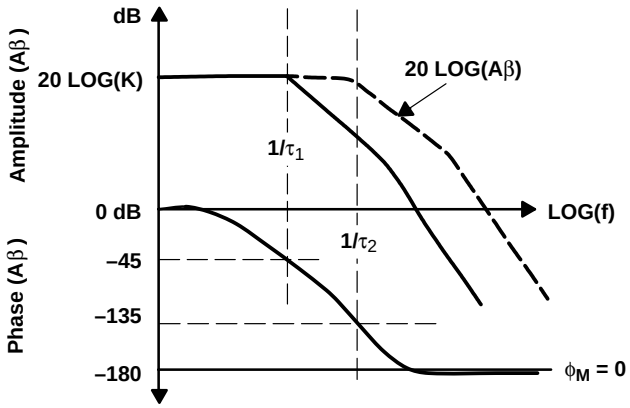


Figure 5–17. Magnitude and Phase Plot of the Loop Gain With Pole Spacing Reduced

The circuit poles are spaced closer in Figure 5–17, and this results in a faster accumulation of phase shift. The phase margin is zero because the loop gain phase shift reaches 180° before the magnitude passes through 0 dB. This circuit oscillates, but it is not a very stable oscillator because the transition to 180° phase shift is very slow. Stable oscillators have a very sharp transition through 180°.

When the closed loop gain is increased the feedback factor, β, is decreased because $V_{OUT}/V_{IN} = 1/\beta$ for the ideal case. This in turn decreases the loop gain, $A\beta$, thus the stability increases. In other words, increasing the closed loop gain makes the circuit more stable. Stability is not important except to oscillator designers because overshoot and ringing become intolerable to linear amplifiers long before oscillation occurs. The overshoot and ringing situation is investigated next.

5.6 The Second Order Equation and Ringing/Overshoot Predictions

The second order equation is a common approximation used for feedback system analysis because it describes a two-pole circuit, which is the most common approximation used. All real circuits are more complex than two poles, but except for a small fraction, they can be represented by a two-pole equivalent. The second order equation is extensively described in electronic and control literature [6].

$$(1 + A\beta) = 1 + \frac{K}{(1 + \tau_1 s)(1 + \tau_2 s)} \tag{5–16}$$

After algebraic manipulation Equation 5–16 is presented in the form of Equation 5–17.

$$s^2 + S\frac{\tau_1 + \tau_2}{\tau_1 \tau_2} + \frac{1 + K}{\tau_1 \tau_2} = 0 \tag{5–17}$$

Equation 5–17 is compared to the second order control Equation 5–18, and the damping ratio, ζ, and natural frequency, w_N are obtained through like term comparisons.

$$s^2 + 2\zeta\omega_N s + \omega_N^2 \tag{5–18}$$

Comparing these equations yields formulas for the phase margin and per cent overshoot as a function of damping ratio.

$$\omega_N = \sqrt{\frac{1 + K}{\tau_1 \tau_2}} \tag{5–19}$$

$$\xi = \frac{\tau_1 + \tau_2}{2\omega_N \tau_1 \tau_2} \tag{5–20}$$

When the two poles are well separated, Equation 5–21 is valid.

$$\phi_M = \text{tangent}^{-1}(2\xi) \tag{5–21}$$

The salient equations are plotted in Figure 5–18, which enables a designer to determine the phase margin and overshoot when the gain and pole locations are known.

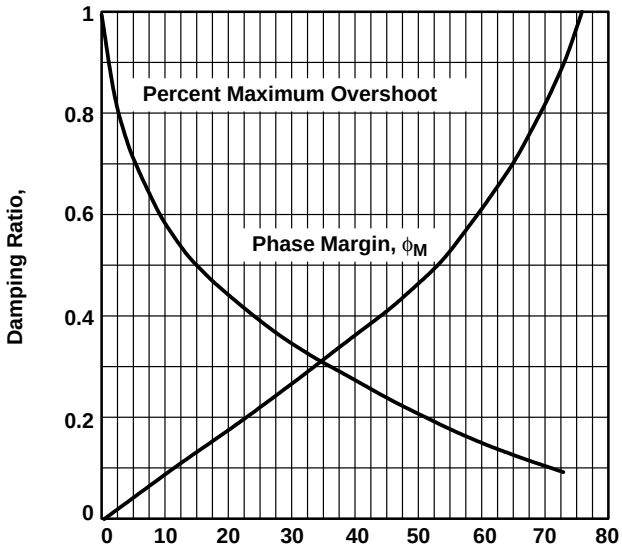


Figure 5–18. Phase Margin and Overshoot vs Damping Ratio

Enter Figure 5–18 at the calculated damping ratio, say 0.4, and read the overshoot at 25% and the phase margin at 42°. If a designer had a circuit specification of 5% maximum overshoot, then the damping ratio must be 0.78 with a phase margin of 62°.

5.7 References

1. DiStefano, Stubberud, and Williams, *Theory and Problems of Feedback and Control Systems, Schaum's Outline Series*, Mc Graw Hill Book Company, 1967
2. Bode, H. W., Network *Analysis And Feedback Amplifier Design*, D. Van Nostrand, Inc., 1945
3. Frederickson, Thomas, *Intuitive Operational Amplifiers*, McGraw Hill Book Company, 1988
4. Bower, J. L. and Schultheis, P. M., *Introduction To The Design Of Servomechanisms*, Wiley, 1961
5. Van Valkenberg, M. E., *Network Analysis*, Prentice-Hall, 1964
6. Del Toro, V., and Parker, S., Principles of Control Systems Engineering, McGraw–Hill, 1960.

Chapter 6

Development of the Non Ideal Op Amp Equations

Ron Mancini

6.1 Introduction

There are two types of error sources in op amps, and they fall under the general classification of dc and ac errors. Examples of dc errors are input offset voltage and input bias current. The dc errors stay constant over the usable op amp frequency range; therefore, the input bias current is 10 pA at 1 kHz and it is 10 pA at 10 kHz. Because of their constant and controlled behavior, dc errors are not considered until later chapters.

AC errors are flighty, so we address them here by developing a set of nonideal equations that account for ac errors. The ac errors may show up under dc conditions, but they get worse as the operating frequency increases. A good example of an ac error is common–mode rejection ration (CMRR). Most op amps have a guaranteed CMRR specification, but this specification is only valid at dc or very low frequencies. Further inspection of the data sheet reveals that CMRR decreases as operating frequency increases. Several other specifications that fall into the category of ac specifications are output impedance, power-supply rejection-ratio, peak-to-peak output voltage, differential gain, differential phase, and phase margin.

Differential gain is the most important ac specification because the other ac specifications are derived from the differential gain. Until now, differential gain has been called op amp gain or op amp open loop gain, and we shall continue with that terminology. Let the data sheet call it differential gain.

As shown in prior chapters, when frequency increases, the op amp gain decreases and errors increase. This chapter develops the equations that illustrate the effects of the gain changes. We start with a review of the basic canonical feedback system stability because the op amp equations are developed using the same techniques.

Amplifiers are built with active components such as transistors. Pertinent transistor parameters like transistor gain are subject to drift and initial inaccuracies from many sources, so amplifiers being built from these components are subject to drift and inaccura-

cies. The drift and inaccuracy is minimized or eliminated by using negative feedback. The op amp circuit configuration employs feedback to make the transfer equation of the circuit independent of the amplifier parameters (well almost), and while doing this, the circuit transfer function is made dependent on external passive components. The external passive components can be purchased to meet almost any drift or accuracy specification; only the cost and size of the passive components limit their use.

Once feedback is applied to the op amp it is possible for the op amp circuit to become unstable. Certain amplifiers belong to a family called internally compensated op amps; they contain internal capacitors that are sometimes advertised as precluding instabilities. Although internally compensated op amps should not oscillate when operated under specified conditions, many have relative stability problems that manifest themselves as poor phase response, ringing, and overshoot. The only absolutely stable internally compensated op amp is the one lying on the workbench without power applied! All other internally compensated op amps oscillate under some external circuit conditions.

Noninternally compensated or *externally* compensated op amps are unstable without the addition of external stabilizing components. This situation is a disadvantage in many cases because they require additional components, but the lack of internal compensation enables the top-drawer circuit designer to squeeze the last drop of performance from the op amp. You have two options: op amps internally compensated by the IC manufacturer, or op amps externally compensated by you. Compensation, except that done by the op amp manufacturer, must be done external to the IC. Surprisingly enough, internally compensated op amps require external compensation for demanding applications.

Compensation is achieved by adding external components that modify the circuit transfer function so that it becomes unconditionally stable. There are several different methods of compensating an op amp, and as you might suspect, there are pros and cons associated with each method of compensation. After the op amp circuit is compensated, it must be analyzed to determine the effects of compensation. The modifications that compensation have on the closed loop transfer function often determine which compensation scheme is most profitably employed.

6.2 Review of the Canonical Equations

A block diagram for a generalized feedback system is repeated in Figure 6–1. This simple block diagram is sufficient to determine the stability of any system.

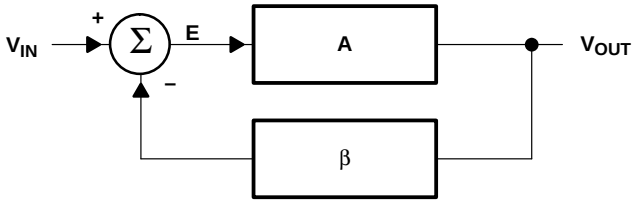


Figure 6–1. Feedback System Block Diagram

The output and error equation development is repeated below.

$$V_{OUT} = EA \tag{6–1}$$

$$E = V_{IN} - \beta V_{OUT} \tag{6–2}$$

Combining Equations 6–1 and 6–2 yields Equation 6–3:

$$\frac{V_{OUT}}{A} = V_{IN} - \beta V_{OUT} \tag{6–3}$$

Collecting terms yields Equation 6–4:

$$V_{OUT}\left(\frac{1}{A} + \beta\right) = V_{IN} \tag{6–4}$$

Rearranging terms yields the classic form of the feedback equation.

$$\frac{V_{OUT}}{V_{IN}} = \frac{A}{1 + A\beta} \tag{6–5}$$

Notice that Equation 6–5 reduces to Equation 6–6 when the quantity Aβ in Equation 6–5 becomes very large with respect to one. Equation 6–6 is called the ideal feedback equation because it depends on the assumption that Aβ >> 1, and it finds extensive use when amplifiers are assumed to have ideal qualities. Under the conditions that Aβ >>1, the system gain is determined by the feedback factor β. Stable passive circuit components are used to implement the feedback factor, thus the ideal closed loop gain is predictable and stable because β is predictable and stable.

$$\frac{V_{OUT}}{V_{IN}} = \frac{1}{\beta} \tag{6–6}$$

The quantity Aβ is so important that it has been given a special name, loop gain. Consider Figure 6–2; when the voltage inputs are grounded (current inputs are opened) and the loop is broken, the calculated gain is the loop gain, Aβ. Now, keep in mind that this is a mathematics of complex numbers, which have magnitude and direction. When the loop gain approaches minus one, or to express it mathematically $1 \angle -180°$, Equation 6–5 approaches infinity because $1/0 \Rightarrow \infty$. The circuit output heads for infinity as fast as it can

using the equation of a straight line. If the output were not energy limited the circuit would explode the world, but it is energy limited by the power supplies so the world stays intact.

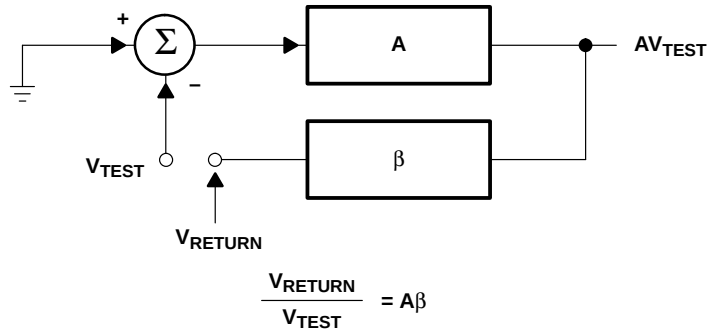


Figure 6–2. Feedback Loop Broken to Calculate Loop Gain

Active devices in electronic circuits exhibit nonlinear behavior when their output approaches a power supply rail, and the nonlinearity reduces the amplifier gain until the loop gain no longer equals $1\angle -180°$. Now the circuit can do two things: first, it could become stable at the power supply limit, or second, it can reverse direction (because stored charge keeps the output voltage changing) and head for the negative power supply rail.

The first state where the circuit becomes stable at a power supply limit is named lockup; the circuit will remain in the locked up state until power is removed. The second state where the circuit bounces between power supply limits is named oscillatory. Remember, the loop gain, $A\beta$, is the sole factor that determines stability for a circuit or system. Inputs are grounded or disconnected when the loop gain is calculated, so they have no effect on stability. The loop gain criteria is analyzed in depth later.

Equations 6–1 and 6–2 are combined and rearranged to yield Equation 6–7, which gives an indication of system or circuit error.

$$E = \frac{V_{IN}}{1 + A\beta} \tag{6–7}$$

First, notice that the error is proportional to the input signal. This is the expected result because a bigger input signal results in a bigger output signal, and bigger output signals require more drive voltage. Second, the loop gain is inversely proportional to the error. As the loop gain increases the error decreases, thus large loop gains are attractive for minimizing errors. Large loop gains also decrease stability, thus there is always a tradeoff between error and stability.

6.3 Noninverting Op Amps

A noninverting op amp is shown in Figure 6–3. The dummy variable, V_B, is inserted to make the calculations easier and a is the op amp gain.

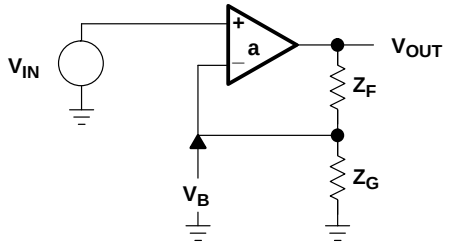


Figure 6–3. Noninverting Op Amp

Equation 6–8 is the amplifier transfer equation.

$$V_{OUT} = a\left(V_{IN} - V_B\right) \tag{6–8}$$

The output equation is developed with the aid of the voltage divider rule. Using the voltage divider rule assumes that the op amp impedance is low.

$$V_B = \frac{V_{OUT} Z_G}{Z_F + Z_G} \text{ for } I_B = 0 \tag{6–9}$$

Combining Equations 6–8 and 6–9 yields Equation 6–10.

$$V_{OUT} = aV_{IN} - \frac{aZ_G\, V_{OUT}}{Z_G + Z_F} \tag{6–10}$$

Rearranging terms in Equation 6–10 yields Equation 6–11, which describes the transfer function of the circuit.

$$\frac{V_{OUT}}{V_{IN}} = \frac{a}{1 + \frac{aZ_G}{Z_G + Z_F}} \tag{6–11}$$

Equation 6–5 is repeated as Equation 6–12 to make a term by term comparison of the equations easy.

$$\frac{V_{OUT}}{V_{IN}} = \frac{A}{1 + A\beta} \tag{6–12}$$

By virtue of the comparison we get Equation 6–13, which is the loop-gain equation for the noninverting op amp. The loop-gain equation determines the stability of the circuit. The

comparison also shows that the open loop gain, A, is equal to the op amp open loop gain, a, for the noninverting circuit.

$$A\beta = \frac{aZ_G}{Z_G + Z_F} \tag{6–13}$$

Equation 6–13 is also derived with the aid of Figure 6–4, which shows the open loop noninverting op amp.

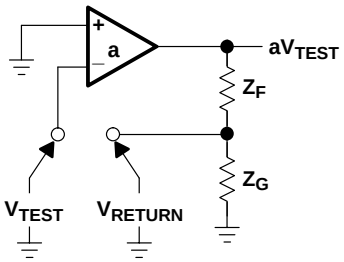


Figure 6–4. Open Loop Noninverting Op Amp

The test voltage, V_{TEST}, is multiplied by the op amp open loop gain to obtain the op amp output voltage, aV_{TEST}. The voltage divider rule is used to calculate Equation 6–15, which is identical to Equation 6–14 after some algebraic manipulation.

$$V_{RETURN} = \frac{aV_{TEST}\,Z_G}{Z_F + Z_G} \tag{6–14}$$

$$\frac{V_{RETURN}}{V_{TEST}} = A\beta = \frac{aZ_G}{Z_F + Z_G} \tag{6–15}$$

6.4 Inverting Op Amps

The inverting op amp circuit is shown in Figure 6–5. The dummy variable (V_A) is inserted to make the calculations easier, and a is the op amp open loop gain.

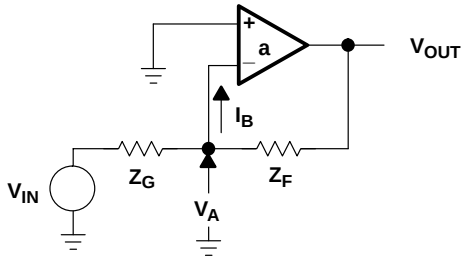


Figure 6–5. Inverting Op Amp

The transfer equation is given in Equation 6–16:

$$V_{OUT} = -aV_A \tag{6–16}$$

The node voltage (Equation 6–17) is obtained with the aid of superposition and the voltage divider rule. Equation 6–18 is obtained by combining Equations 6–16 and 6–17.

$$V_A = \frac{V_{IN} Z_F}{Z_G + Z_F} + \frac{V_{OUT} Z_G}{Z_G + Z_F} \quad \text{for } I_B = 0 \tag{6–17}$$

$$\frac{V_{OUT}}{V_{IN}} = \frac{\dfrac{-aZ_F}{Z_G + Z_F}}{1 + \dfrac{aZ_G}{Z_G + Z_F}} \tag{6–18}$$

Equation 6–16 is the transfer function of the inverting op amp. By virtue of the comparison between Equations 6–18 and 6–14, we get Equation 6–15 again, which is also the loop gain equation for the inverting op amp circuit. The comparison also shows that the open loop gain (A) is different from the op amp open loop gain (a) for the noninverting circuit.

The inverting op amp with the feedback loop broken is shown in Figure 6–6, and this circuit is used to calculate the loop-gain given in Equation 6–19.

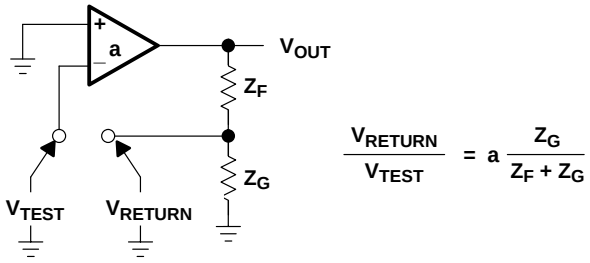


Figure 6–6. Inverting Op Amp: Feedback Loop Broken for Loop Gain Calculation

$$\frac{V_{RETURN}}{V_{TEST}} = \frac{aZ_G}{Z_G + Z_F} = A\beta \tag{6–19}$$

Several things must be mentioned at this point in the analysis. First, the transfer functions for the noninverting and inverting Equations, 6–13 and 6–18, are different. For a common set of Z_G and Z_F values, the magnitude and polarity of the gains are different. Second, the loop gain of both circuits, as given by Equations 6–15 and 6–19, is identical. Thus, the stability performance of both circuits is identical although their transfer equations are different. This makes the important point that *stability is not dependent on the circuit inputs*. Third, the A gain block shown in Figure 6–1 is different for each op amp circuit. By comparison of Equations 6–5, 6–11, and 6–18 we see that $A_{NON–INV} = a$ and $A_{INV} = aZ_F \div (Z_G + Z_F)$.

6.5 Differential Op Amps

The differential amplifier circuit is shown in Figure 6–7. The dummy variable, V_E, is inserted to make the calculations easier, and a is the open loop gain.

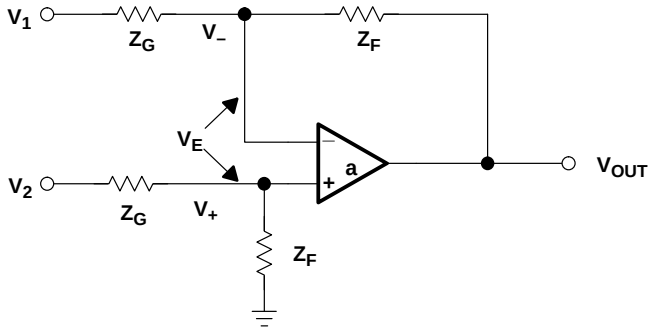


Figure 6–7. Differential Amplifier Circuit

Equation 6–20 is the circuit transfer equation.

$$V_{OUT} = a\,V_E = V_+ - V_- \tag{6–20}$$

The positive input voltage, V_+, is written in Equation 6–21 with the aid of superposition and the voltage divider rule.

$$V_+ = V_2 \frac{Z_F}{Z_F + Z_G} \tag{6–21}$$

The negative input voltage, V_-, is written in Equation 6–22 with the aid of superposition and the voltage divider rule.

$$V_- = V_1 \frac{Z_F}{Z_F + Z_G} - V_{OUT} \frac{Z_G}{Z_F + Z_G} \tag{6–22}$$

Combining Equations 6–20, 6–21, and 6–22 yields Equation 6–23.

$$V_{OUT} = a\left[\frac{V_2 Z_F}{Z_F + Z_G} - \frac{V_1 Z_F}{Z_F + Z_G} - \frac{V_{OUT} Z_G}{Z_F + Z_G}\right] \tag{6–23}$$

After algebraic manipulation, Equation 6–23 reduces to Equation 6–24.

$$\frac{V_{OUT}}{V_2 - V_1} = \frac{\frac{aZ_F}{Z_F + Z_G}}{1 + \frac{aZ_G}{Z_F + Z_G}} \tag{6–24}$$

The comparison method reveals that the loop gain as shown in Equation 6–25 is identical to that shown in Equations 6–13 and 6–19.

$$A\beta = \frac{aZ_G}{Z_G + Z_F} \tag{6–25}$$

Again, the loop gain, which determines stability, is only a function of the closed loop, and independent of the inputs.

Chapter 7

Voltage-Feedback Op Amp Compensation

Ron Mancini

7.1 Introduction

Voltage-feedback amplifiers (VFA) have been with us for about 60 years, and they have been problems for circuit designers since the first day. You see, the feedback that makes them versatile and accurate also has a tendency to make them unstable. The operational amplifier (op amp) circuit configuration uses a high-gain amplifier whose parameters are determined by external feedback components. The amplifier gain is so high that without these external feedback components, the slightest input signal would saturate the amplifier output. The op amp is in common usage, so this configuration is examined in detail, but the results are applicable to many other voltage-feedback circuits. Current-feedback amplifiers (CFA) are similar to VFAs, but the differences are important enough to warrant CFAs being handled separately.

Stability as used in electronic circuit terminology is often defined as achieving a nonoscillatory state. This is a poor, inaccurate definition of the word. Stability is a relative term, and this situation makes people uneasy because relative judgments are exhaustive. It is easy to draw the line between a circuit that oscillates and one that does not oscillate, so we can understand why some people believe that oscillation is a natural boundary between stability and instability.

Feedback circuits exhibit poor phase response, overshoot, and ringing long before oscillation occurs, and these effects are considered undesirable by circuit designers. This chapter is not concerned with oscillators; thus, relative stability is defined in terms of performance. By definition, when designers decide what tradeoffs are acceptable, they determine what the relative stability of the circuit is. A relative stability measurement is the damping ratio (ζ) and the damping ratio is discussed in detail in Reference 1. The damping ratio is related to phase margin, hence phase margin is another measure of relative stability. The most stable circuits have the longest response times, lowest bandwidth, highest accuracy, and least overshoot. The least stable circuits have the fastest response times, highest bandwidth, lowest accuracy, and some overshoot.

Op Amps left in their native state oscillate without some form of compensation. The first IC op amps were very hard to stabilize, but there were a lot of good analog designers

around in the ’60s, so we used them. Internally compensated op amps were introduced in the late ’60s in an attempt to make op amps easy for everyone to use. Unfortunately, internally compensated op amps sacrifice a lot of bandwidth and still oscillate under some conditions, so an understanding of compensation is required to apply op amps.

Internal compensation provides a worst-case trade-off between stability and performance. Uncompensated op amps require more attention, but they can do more work. Both are covered here.

Compensation is a process of applying a judicious patch in the form of an RC network to make up for a less than perfect op amp or circuit. There are many different problems that can introduce instability, thus there are many different compensation schemes.

7.2 Internal Compensation

Op amps are internally compensated to save external components and to enable their use by less knowledgeable people. It takes some measure of analog knowledge to compensate an analog circuit. Internally compensated op amps normally are stable when they are used in accordance with the applications instructions. Internally compensated op amps are not unconditionally stable. They are multiple pole systems, but they are internally compensated such that they appear as a single pole system over much of the frequency range. Internal compensation severely decreases the possible closed-loop bandwidth of the op amp.

Internal compensation is accomplished in several ways, but the most common method is to connect a capacitor across the collector-base junction of a voltage gain transistor (see Figure 7–1). The Miller effect multiplies the capacitor value by an amount approximately equal to the stage gain, thus the Miller effect uses small value capacitors for compensation.

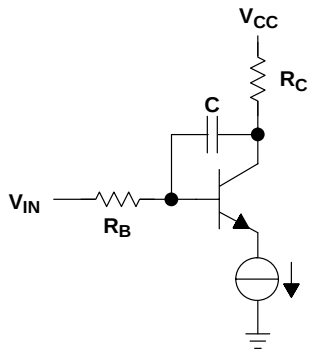


Figure 7–1. Miller Effect Compensation

Figure 7–2 shows the gain/phase diagram for an older op amp (TL03X). When the gain crosses the 0-dB axis (gain equal to one) the phase shift is approximately 108°, thus the op amp must be modeled as a second-order system because the phase shift is more than 90°.

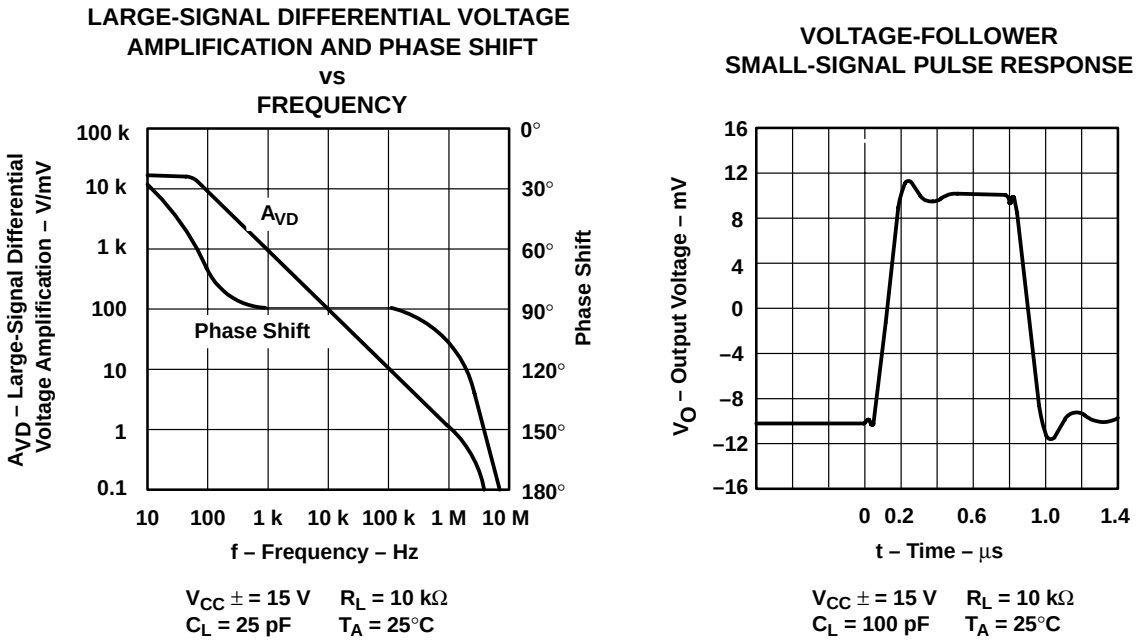


Figure 7–2. TL03X Frequency and Time Response Plots

This yields a phase margin of $\phi = 180° - 108° = 72°$, thus the circuit should be very stable. Referring to Figure 7–3, the damping ratio is one and the expected overshoot is zero. Figure 7–2 shows approximately 10% overshoot, which is unexpected, but inspecting Figure 7–2 further reveals that the loading capacitance for the two plots is different. The pulse response is loaded with 100 pF rather than 25 pF shown for the gain/phase plot, and this extra loading capacitance accounts for the loss of phase margin.

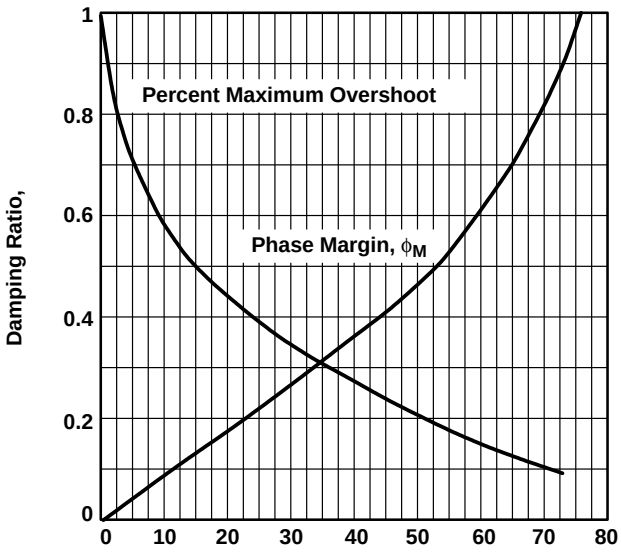


Figure 7–3. Phase Margin and Percent Overshoot Versus Damping Ratio

Why does the loading capacitance make the op amp unstable? Look closely at the gain/phase response between 1 MHz and 9 MHz, and observe that the gain curve changes slope drastically while the rate of phase change approaches 120°/decade. The radical gain/phase slope change proves that several poles are located in this area. The loading capacitance works with the op amp output impedance to form another pole, and the new pole reacts with the internal op amp poles. As the loading capacitor value is increased, its pole migrates down in frequency, causing more phase shift at the 0-dB crossover frequency. The proof of this is given in the TL03X data sheet where plots of ringing and oscillation versus loading capacitance are shown.

Figure 7–4 shows similar plots for the TL07X, which is the newer family of op amps. Notice that the phase shift is approximately 100° when the gain crosses the 0-dB axis. This yields a phase margin of 80°, which is close to unconditionally stable. The slope of the phase curve changes to 180°/decade about one decade from the 0-dB crossover point. The radical slope change causes suspicion about the 90° phase margin, furthermore the gain curve must be changing radically when the phase is changing radically. The gain/phase plot may not be totally false, but it sure is overly optimistic.

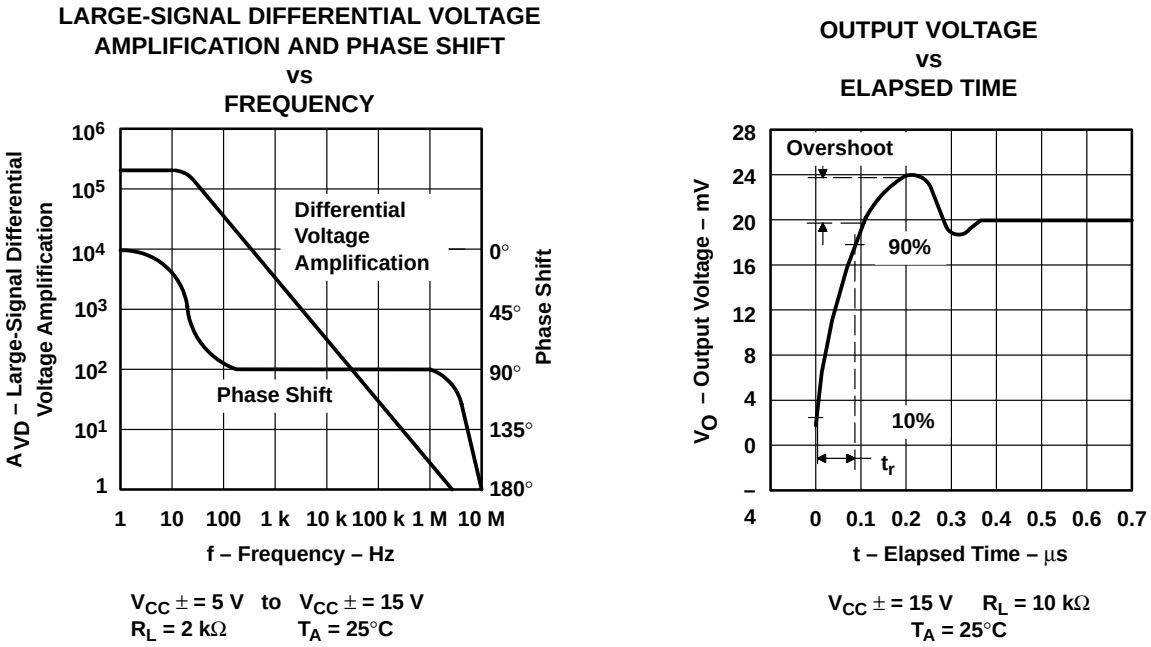


Figure 7–4. TL07X Frequency and Time Response Plots

The TL07X pulse response plot shows approximately 20% overshoot. There is no loading capacitance indicated on the plot to account for a seemingly unconditionally stable op amp exhibiting this large an overshoot. Something is wrong here: the analysis is wrong, the plots are wrong, or the parameters are wrong. Figure 7–5 shows the plots for the TL08X family of op amps, which are sisters to the TL07X family. The gain/phase curve and pulse response is virtually identical, but the pulse response lists a 100 pF loading capacitor. This little exercise illustrates three valuable points: first, if the data seems wrong it probably is wrong, second, even the factory people make mistakes, and third, the loading capacitor makes op amps ring, overshoot, or oscillate.

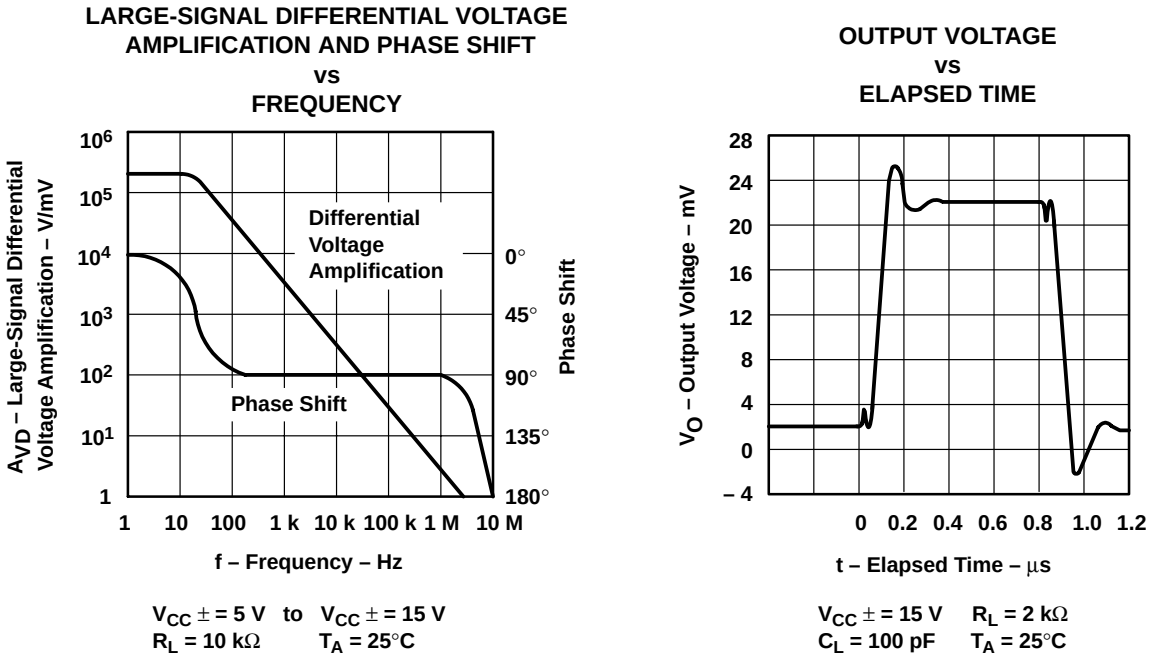


Figure 7–5. TL08X Frequency and Time Response Plots

The frequency and time-response plots for the TLV277X family of op amps is shown in Figures 7–6 and 7–7. First, notice that the information is more sophisticated because the phase response is given in degrees of phase margin; second, both gain/phase plots are done with substantial loading capacitors (600 pF), so they have some practical value; and third, the phase margin is a function of power supply voltage.

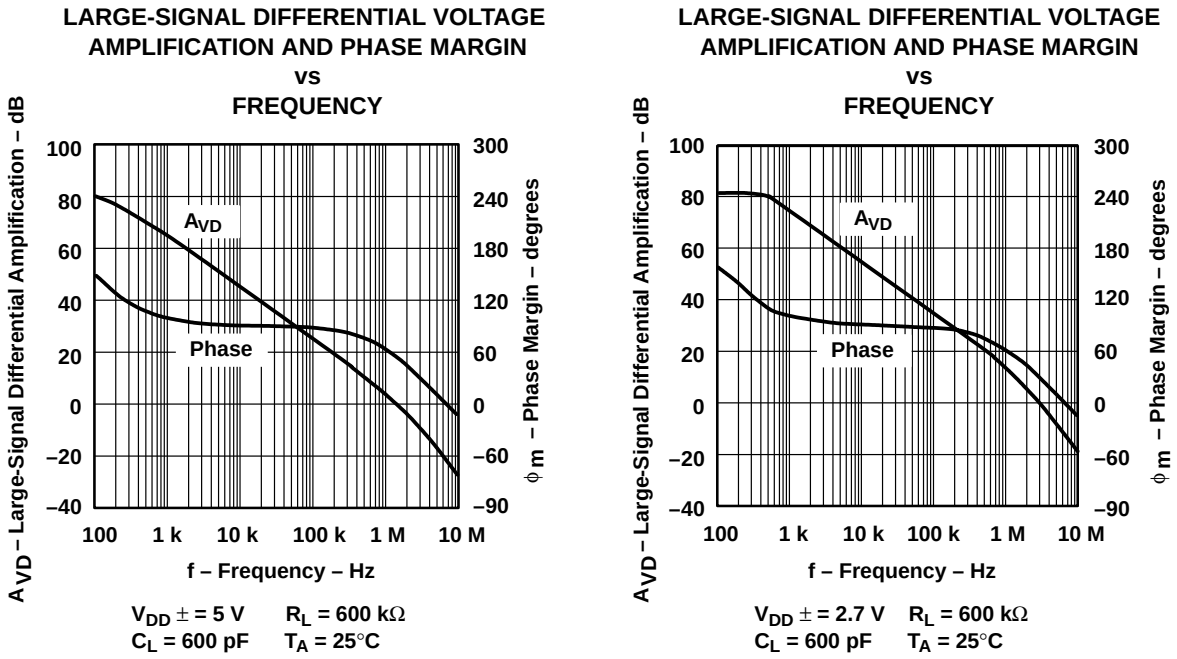


Figure 7–6. TLV277X Frequency Response Plots

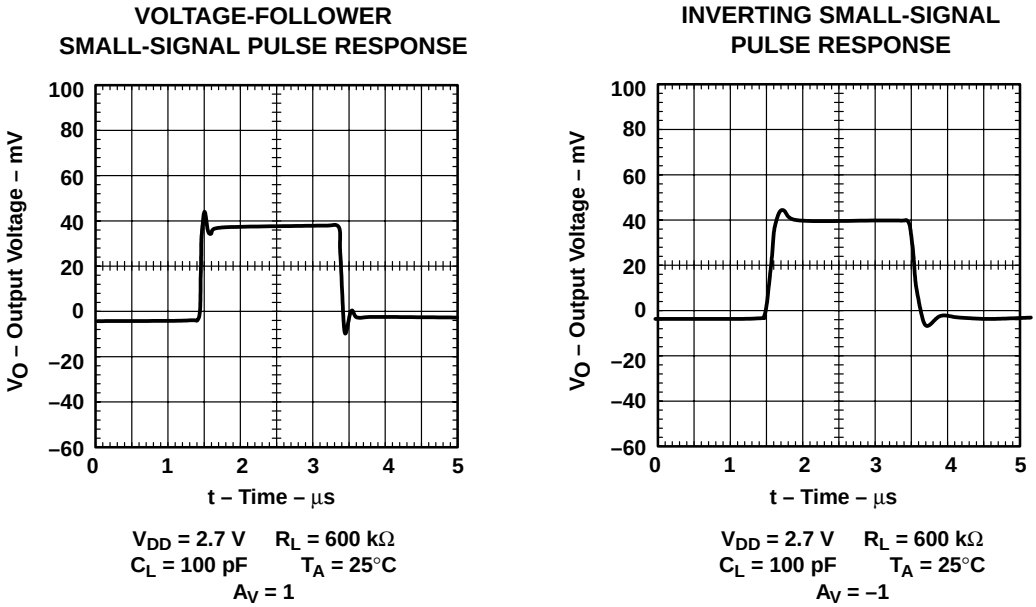


Figure 7–7. TLV227X Time Response Plots

At V_{CC} = 5 V, the phase margin at the 0-dB crossover point is 60°, while it is 30° at V_{CC} = 2.7 V. This translates into an expected overshoot of 18% at V_{CC} = 5 V, and 28% at V_{CC} = 2.7 V. Unfortunately the time response plots are done with 100-pF loading capacitance, hence we can not check our figures very well. The V_{CC} = 2.7 V overshoot is approximately 2%, and it is almost impossible to figure out what the overshoot would have been with a 600 pF loading capacitor. The small-signal pulse response is done with mV-signals, and that is a more realistic measurement than using the full signal swing.

Internally compensated op amps are very desirable because they are easy to use, and they do not require external compensation components. Their drawback is that the bandwidth is limited by the internal compensation scheme. The op amp open-loop gain eventually (when it shows up in the loop gain) determines the error in an op amp circuit. In a noninverting buffer configuration, the TL277X is limited to 1% error at 50 kHz (V_{CC} = 2.7 V) because the op amp gain is 40 dB at that point. Circuit designers can play tricks such as bypassing the op amp with a capacitor to emphasize the high-frequency gain, but the error is still 1%. Keep Equation 7–1 in mind because it defines the error. If the TLV277X were not internally compensated, it could be externally compensated for a lower error at 50 kHz because the gain would be much higher.

$$E = \frac{V_{IN}}{1 + A\beta} \tag{7–1}$$

7.3 External Compensation, Stability, and Performance

Nobody compensates an op amp just because it is there; they have a reason to compensate the op amp, and that reason is usually stability. They want the op amp to perform a function in a circuit where it is potentially unstable. Internally and noninternally compensated op amps are compensated externally because certain circuit configurations do cause oscillations. Several potentially unstable circuit configurations are analyzed in this section, and the reader can extend the external compensation techniques as required.

Other reasons for externally compensating op amps are noise reduction, flat amplitude response, and obtaining the highest bandwidth possible from an op amp. An op amp generates noise, and noise is generated by the system. The noise contains many frequency components, and when a high-pass filter is incorporated in the signal path, it reduces high frequency noise. Compensation can be employed to roll off the op amp's high frequency, closed-loop response, thus causing the op amp to act as a noise filter. Internally compensated op amps are modeled with a second order equation, and this means that the output voltage can overshoot in response to a step input. When this overshoot (or peaking) is undesirable, external compensation can increase the phase margin to 90° where there is no peaking. An uncompensated op amp has the highest bandwidth possible. External compensation is required to stabilize uncompensated op amps, but the compensation can be tailored to the specific circuit, thus yielding the highest possible bandwidth consistent with the pulse response requirements.

7.4 Dominant-Pole Compensation

We saw that capacitive loading caused potential instabilities, thus an op amp loaded with an output capacitor is a circuit configuration that must be analyzed. This circuit is called dominant pole compensation because if the pole formed by the op amp output impedance and the loading capacitor is located close to the zero frequency axis, it becomes dominant. The op amp circuit is shown in Figure 7–8, and the open loop circuit used to calculate the loop gain (Aβ) is shown in Figure 7–9.

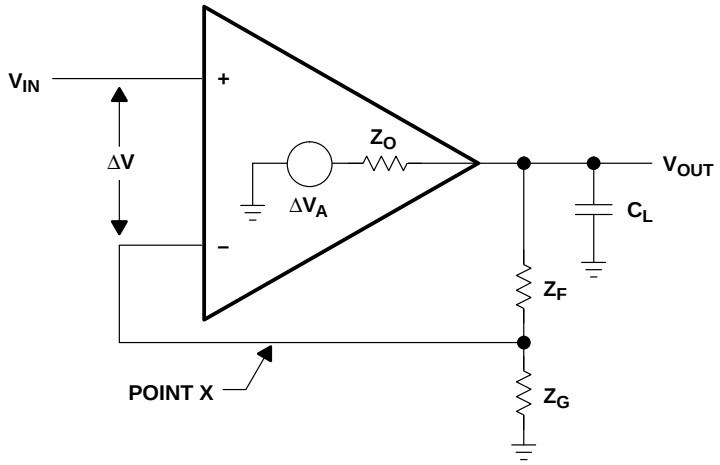


Figure 7–8. Capacitively-Loaded Op Amp

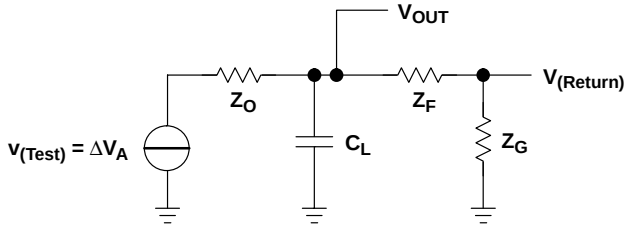


Figure 7–9. Capacitively-Loaded Op Amp With Loop Broken for Loop Gain (Aβ) Calculation

The analysis starts by looking into the capacitor and taking the Thevenin equivalent circuit.

$$V_{TH} = \frac{\Delta Va}{Z_O C_L s + 1} \tag{7–2}$$

$$Z_{TH} = \frac{Z_O}{Z_O C_L s + 1} \tag{7–3}$$

Then the output equation is written.

$$V_{RETURN} = \frac{V_{TH}Z_G}{Z_G + Z_F + Z_{TH}} = \frac{\Delta Va}{Z_O C_L s + 1}\left(\frac{Z_G}{Z_F + Z_G + \frac{Z_O}{Z_O C_L s + 1}}\right) \quad (7\text{–}4)$$

Rearranging terms yields Equation 7–5.

$$\frac{V_{RETURN}}{V_{TEST}} = A\beta = \frac{\frac{aZ_G}{Z_F + Z_G + Z_O}}{\frac{(Z_F + Z_G)Z_O C_L s}{Z_F + Z_G + Z_O} + 1} \quad (7\text{–}5)$$

When the assumption is made that $(Z_F + Z_G) >> Z_O$, Equation 7–5 reduces to Equation 7–6.

$$A\beta = \frac{aZ_G}{Z_F + Z_G}\left(\frac{1}{Z_O C_L s + 1}\right) \quad (7\text{–}6)$$

Equation 7–7 models the op amp as a second-order system. Hence, substituting the second-order model for *a* in Equation 7–6 yields Equation 7–8, which is the stability equation for the dominant-pole compensation circuit.

$$a = \frac{K}{(s + \tau_1)(s + \tau_2)} \quad (7\text{–}7)$$

$$A\beta = \frac{K}{(s + \tau_1)(s + \tau_2)} \frac{Z_G}{Z_F + Z_G} \frac{1}{Z_O C_L s + 1} \quad (7\text{–}8)$$

Several conclusions can be drawn from Equation 7–8 depending on the location of the poles. If the Bode plot of Equation 7–7, the op amp transfer function, looks like that shown in Figure 7–10, it only has 25° phase margin, and there is approximately 48% overshoot. When the pole introduced by Z_O and C_L moves towards the zero frequency axis it comes close to the τ_2 pole, and it adds phase shift to the system. Increased phase shift increases peaking and decreases stability. In the real world, many loads, especially cables, are capacitive, and an op amp like the one pictured in Figure 7–10 would ring while driving a capacitive load. The load capacitance causes peaking and instability in internally compensated op amps when the op amps do not have enough phase margin to allow for the phase shift introduced by the load.

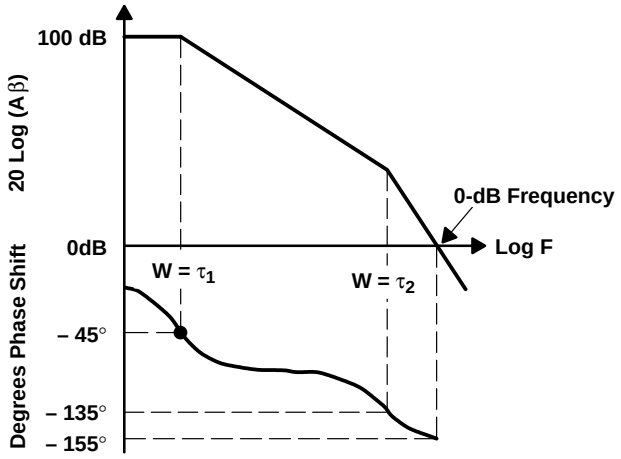


Figure 7–10. Possible Bode Plot of the Op Amp Described in Equation 7–7

Prior to compensation, the Bode plot of an uncompensated op amp looks like that shown in Figure 7–11. Notice that the break points are located close together thus accumulating about 180° of phase shift before the 0 dB crossover point; the op amp is not usable and probably unstable. Dominant pole compensation is often used to stabilize these op amps. If a dominant pole, in this case ω_D, is properly placed it rolls off the gain so that τ_1 introduces 45° phase at the 0-dB crossover point. After the dominant pole is introduced the op amp is stable with 45° phase margin, but the op amp gain is drastically reduced for frequencies higher than ω_D. This procedure works well for internally compensated op amps, but is seldom used for externally compensated op amps because inexpensive discrete capacitors are readily available.

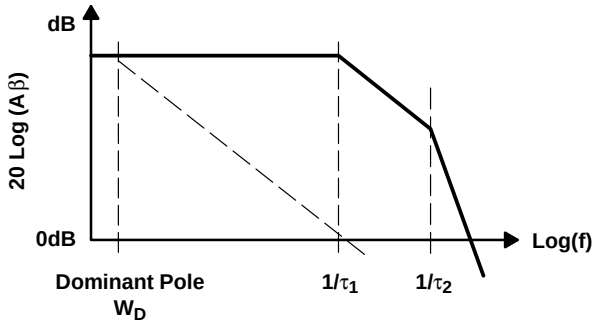


Figure 7–11. Dominant-Pole Compensation Plot

Assuming that $Z_O << Z_F$, the closed-loop transfer function is easy to calculate because C_L is enclosed in the feedback loop. The ideal closed-loop transfer equation is the same as Equation 6–11 for the noninverting op amp, and is repeated below as Equation 7–9.

$$\frac{V_{OUT}}{V_{IN}} = \frac{a}{1 + \frac{aZ_G}{Z_G + Z_F}} \tag{7–9}$$

When a $\Rightarrow \infty$ Equation 7–9 reduces to Equation 7–10.

$$\frac{V_{OUT}}{V_{IN}} = \frac{Z_F + Z_G}{Z_G} \tag{7–10}$$

As long as the op amp has enough compliance and current to drive the capacitive load, and Z_O is small, the circuit functions as though the capacitor was not there. When the capacitor becomes large enough, its pole interacts with the op amp pole causing instability. When the capacitor is huge, it completely kills the op amp's bandwidth, thus lowering the noise while retaining a large low-frequency gain.

7.5 Gain Compensation

When the closed-loop gain of an op amp circuit is related to the loop gain, as it is in voltage-feedback op amps, the closed-loop gain can be used to stabilize the circuit. This type of compensation can not be used in current-feedback op amps because the mathematical relationship between the loop gain and ideal closed-loop gain does not exist. The loop gain equation is repeated as Equation 7–11. Notice that the closed-loop gain parameters Z_G and Z_F are contained in Equation 7–11, hence the stability can be controlled by manipulating the closed-loop gain parameters.

$$A\beta = \frac{aZ_G}{Z_G + Z_F} \tag{7–11}$$

The original loop gain curve for a closed-loop gain of one is shown in Figure 7–12, and it is or comes very close to being unstable. If the closed-loop noninverting gain is changed to 9, then K changes from K/2 to K/10. The loop gain intercept on the Bode plot (Figure 7–12) moves down 14 dB, and the circuit is stabilized.

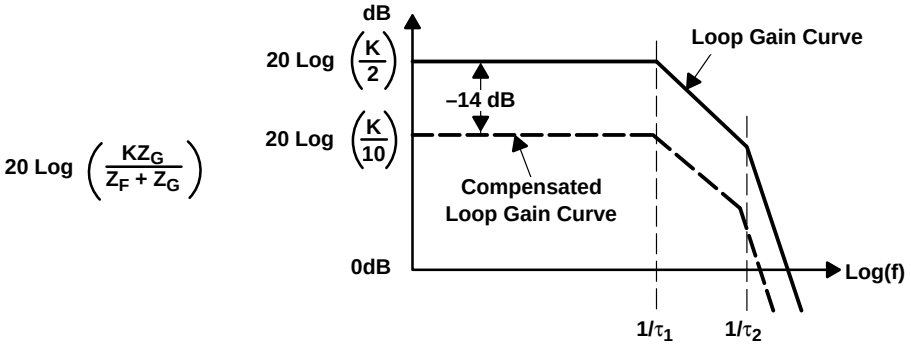


Figure 7–12. Gain Compensation

Gain compensation works for inverting or noninverting op amp circuits because the loop gain equation contains the closed-loop gain parameters in both cases. When the closed-loop gain is increased, the accuracy and the bandwidth decrease. As long as the application can stand the higher gain, gain compensation is the best type of compensation to use. Uncompensated versions of normally internally compensated op amps are offered for sale as stable op amps with minimum gain restrictions. As long as gain in the circuit you design exceeds the gain specified on the data sheet, this is economical and a safe mode of operation.

7.6 Lead Compensation

Sometimes lead compensation is forced on the circuit designer because of the parasitic capacitance associated with packaging and wiring op amps. Figure 7–13 shows the circuit for lead compensation; notice the capacitor in parallel with R_F. That capacitor is often made by parasitic wiring and the ground plane, and high frequency circuit designers go to great lengths to minimize or eliminate it. What is good in one sense is bad in another, because adding the parallel capacitor is a good way to stabilize the op amp and reduce noise. Let us analyze the stability first, and then we will analyze the closed-loop performance.

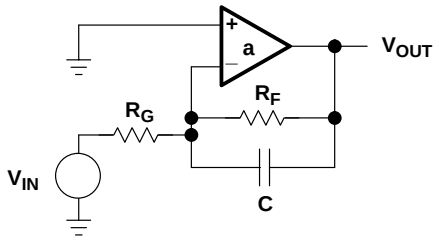


Figure 7–13. Lead-Compensation Circuit

The loop equation for the lead-compensation circuit is given by Equation 7–12.

$$A\beta = \left(\frac{R_G}{R_G + R_F}\right)\left(\frac{R_F Cs + 1}{R_G \parallel R_F Cs + 1}\right)\left[\frac{K}{(s + \tau_1)(s + \tau_2)}\right] \qquad (7\text{–}12)$$

The compensation capacitor introduces a pole and zero into the loop equation. The zero always occurs before the pole because $R_F > R_F || R_G$. When the zero is properly placed it cancels out the τ_2 pole along with its associated phase shift. The original transfer function is shown in Figure 7–14 drawn in solid lines. When the $R_F C$ zero is placed at $\omega = 1/\tau_2$, it cancels out the τ_2 pole causing the bode plot to continue on a slope of –20 dB/decade. When the frequency gets to $\omega = 1/(R_F||RG)C$, this pole changes the slope to –40 dB/decade. Properly placed, the capacitor aids stability, but what does it do to the closed-loop

transfer function? The equation for the inverting op amp closed-loop gain is repeated below.

$$\frac{V_{OUT}}{V_{IN}} = \frac{\frac{-aZ_F}{Z_G + Z_F}}{1 + \frac{aZ_G}{Z_G + Z_F}} \tag{7–13}$$

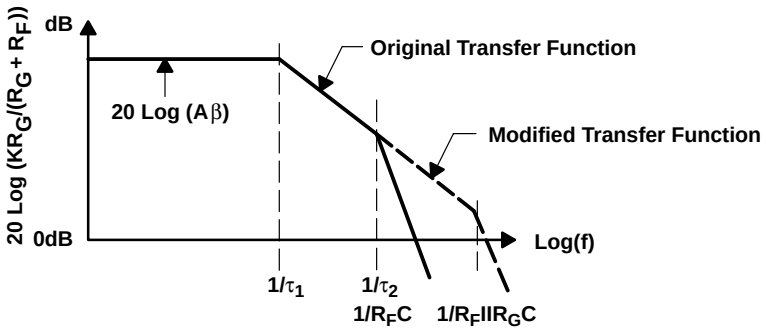


Figure 7–14. Lead-Compensation Bode Plot

When *a* approaches infinity, Equation 7–13 reduces to Equation 7–14.

$$\frac{V_{OUT}}{V_{IN}} = -\frac{Z_F}{Z_{IN}} \tag{7–14}$$

Substituting $R_F||C$ for Z_F and R_G for Z_G in Equation 7–14 yields Equation 7–15, which is the ideal closed-loop gain equation for the lead compensation circuit.

$$\frac{V_{OUT}}{V_{IN}} = -\frac{R_F}{R_G}\left(\frac{1}{R_F Cs + 1}\right) \tag{7–15}$$

The forward gain for the inverting amplifier is given by Equation 7–16. Compare Equation 7–13 with Equation 6–5 to determine A.

$$A = \frac{aZ_F}{Z_G + A_F} = \left(\frac{aR_F}{R_G + R_F}\right)\left(\frac{1}{R_F \parallel R_G Cs + 1}\right) \tag{7–16}$$

The op amp gain (a), the forward gain (A), and the ideal closed-loop gain are plotted in Figure 7–15. The op amp gain is plotted for reference only. The forward gain for the inverting op amp is not the op amp gain. Notice that the forward gain is reduced by the factor $R_F/(R_G + R_F)$, and it contains a high frequency pole. The ideal closed-loop gain follows the ideal curve until the $1/R_FC$ breakpoint (same location as $1/\tau_2$ breakpoint), and then it

slopes down at –20 dB/decade. Lead compensation sacrifices the bandwidth between the $1/R_FC$ breakpoint and the forward gain curve. The location of the $1/R_FC$ pole determines the bandwidth sacrifice, and it can be much greater than shown here. The pole caused by R_F, R_G, and C does not appear until the op amp's gain has crossed the 0-dB axis, thus it does not affect the ideal closed-loop transfer function.

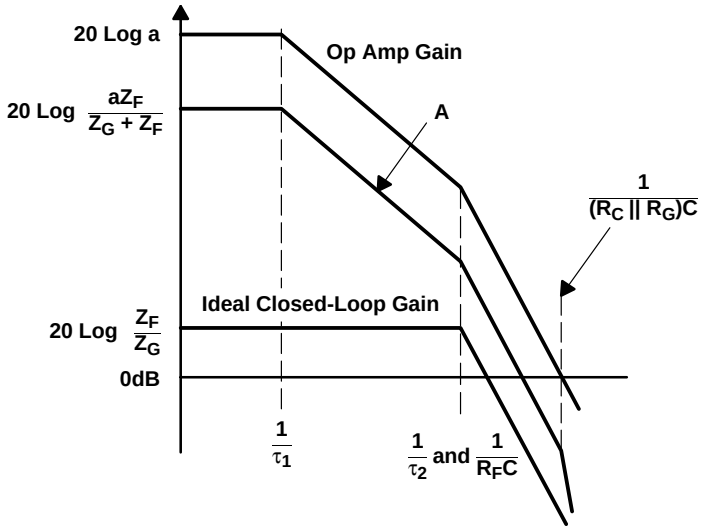


Figure 7–15. Inverting Op Amp With Lead Compensation

The forward gain for the noninverting op amp is *a*; compare Equation 6–11 to Equation 6–5. The ideal closed-loop gain is given by Equation 7–17.

$$\frac{V_{OUT}}{V_{IN}} = \frac{Z_F + Z_G}{Z_G} = \left(\frac{R_F + R_G}{R_G}\right)\left(\frac{R_F \parallel R_G Cs + 1}{R_F Cs + 1}\right) \qquad (7\text{–}17)$$

The plot of the noninverting op amp with lead compensation is shown in Figure 7–16. There is only one plot for both the op amp gain (a) and the forward gain (A), because they are identical in the noninverting circuit configuration. The ideal starts out as a flat line, but it slopes down because its closed-loop gain contains a pole and a zero. The pole always occurs closer to the low frequency axis because $R_F > R_F||R_G$. The zero flattens the ideal closed-loop gain curve, but it never does any good because it cannot fall on the pole. The pole causes a loss in the closed-loop bandwidth by the amount separating the closed-loop and forward gain curves.

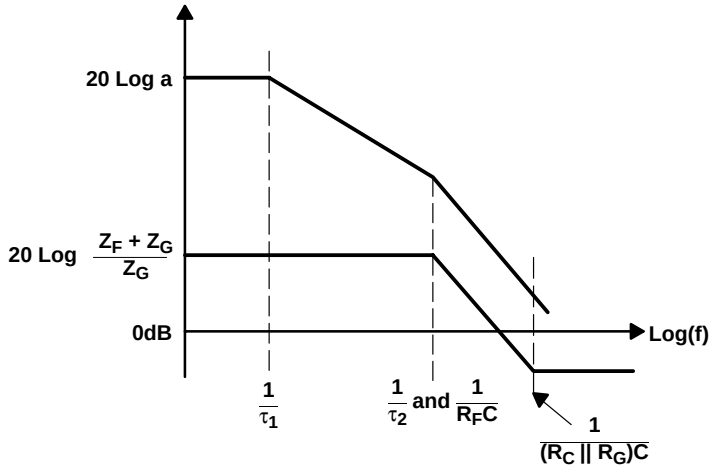


Figure 7–16. Noninverting Op Amp With Lead Compensation

Although the forward gain is different in the inverting and noninverting circuits, the closed-loop transfer functions take very similar shapes. This becomes truer as the closed-loop gain increases because the noninverting forward gain approaches the op amp gain. This relationship cannot be relied on in every situation, and each circuit must be checked to determine the closed-loop effects of the compensation scheme.

7.7 Compensated Attenuator Applied to Op Amp

Stray capacitance on op amp inputs is a problem that circuit designers are always trying to get away from because it decreases stability and causes peaking. The circuit shown in Figure 7–17 has some stray capacitance (C_G,) connected from the inverting input to ground. Equation 7–18 is the loop gain equation for the circuit with input capacitance.

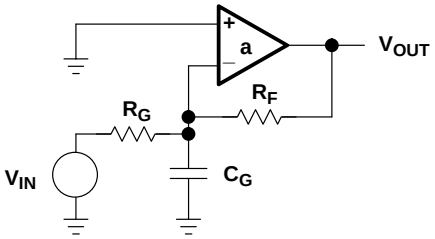


Figure 7–17. Op Amp With Stray Capacitance on the Inverting Input

$$A\beta = \left(\frac{R_G}{R_G + R_F}\right)\left(\frac{1}{R_G \| R_F Cs + 1}\right)\left[\frac{K}{(\tau_1 s + 1)(\tau_2 s + 1)}\right] \quad (7\text{–}18)$$

Op amps having high input and feedback resistors are subject to instability caused by stray capacitance on the inverting input. Referring to Equation 7–18, when the

$1/(R_F||R_GC_G)$ pole moves close to τ_2 the stage is set for instability. Reasonable component values for a CMOS op amp are R_F = 1 MΩ, R_G = 1 MΩ, and C_G = 10 pF. The resulting pole occurs at 318 kHz, and this frequency is lower than the breakpoint of τ_2 for many op amps. There is 90° of phase shift resulting from τ_1, the $1/(R_F||R_GC)$ pole adds 45° phase shift at 318 kHz, and τ_2 starts to add another 45° phase shift at about 600 kHz. This circuit is unstable because of the stray input capacitance. The circuit is compensated by adding a feedback capacitor as shown in Figure 7–18.

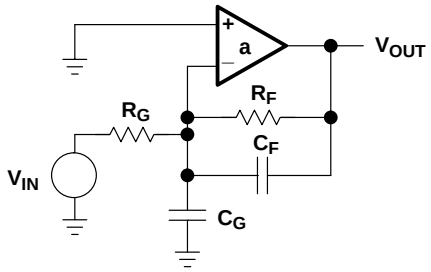


Figure 7–18. Compensated Attenuator Circuit

The loop gain with C_F added is given by Equation 7–19.

$$A\beta = \left[\frac{\frac{R_G}{R_GC_Gs+1}}{\frac{R_G}{R_GC_Gs+1}+\frac{R_F}{R_FC_Fs+1}}\right]\left[\frac{K}{(\tau_1 s+1)(\tau_2 s+1)}\right] \qquad (7\text{–}19)$$

If $R_GC_G = R_FC_F$ Equation 7–19 reduces to Equation 7–20.

$$A\beta = \left(\frac{R_G}{R_G+R_F}\right)\left[\frac{K}{(\tau_1 s+1)(\tau_2 s+1)}\right] \qquad (7\text{–}20)$$

The compensated attenuator Bode plot is shown in Figure 7–19. Adding the correct $1/R_FC_F$ breakpoint cancels out the $1/R_GC_G$ breakpoint; the loop gain is independent of the capacitors. Now is the time to take advantage of the stray capacitance. C_F can be formed by running a wide copper strip from the output of the op amp over the ground plane under R_F; do not connect the other end of this copper strip. The circuit is tuned by removing some copper (a razor works well) until all peaking is eliminated. Then measure the copper, and have an identical trace put on the printed-circuit board.

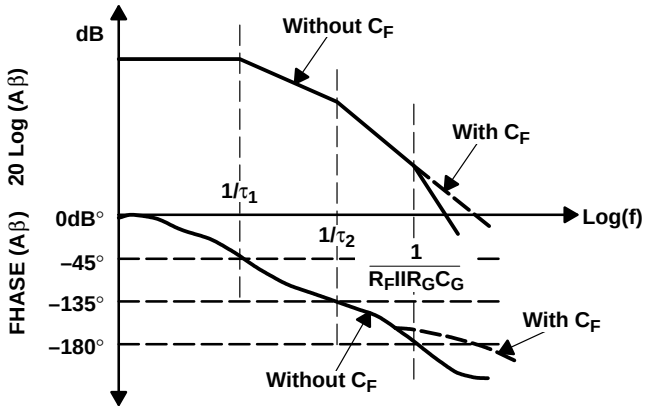


Figure 7–19. Compensated Attenuator Bode Plot

The inverting and noninverting closed-loop gain equations are a function of frequency. Equation 7–21 is the closed-loop gain equation for the inverting op amp. When $R_FC_F = R_GC_G$, Equation 7–21 reduces to Equation 7–22, which is independent of the breakpoint. This also happens to the noninverting op amp circuit. This is one of the few occasions when the compensation does not affect the closed-loop gain frequency response.

$$\frac{V_{OUT}}{V_{IN}} = - \frac{\frac{R_F}{R_FC_Fs+1}}{\frac{R_G}{R_GC_Gs+1}} \qquad (7–21)$$

When $R_FC_F = R_GC_G$ $\qquad \frac{V_{OUT}}{V_{IN}} = - \left(\frac{R_F}{R_G}\right)$ (7–22)

7.8 Lead-Lag Compensation

Lead-lag compensation stabilizes the circuit without sacrificing the closed-loop gain performance. It is often used with uncompensated op amps. This type of compensation provides excellent high-frequency performance. The circuit schematic is shown in Figure 7–20, and the loop gain is given by Equation 7–23.

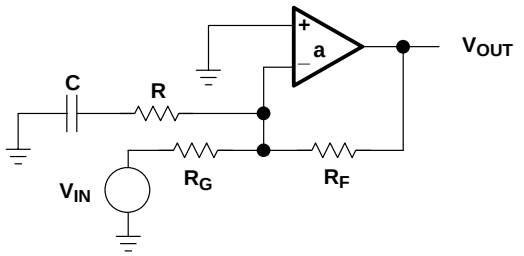


Figure 7–20. Lead-Lag Compensated Op Amp

$$A\beta = \frac{K}{(\tau_1 s + 1)(\tau_2 s + 1)} \frac{R_G}{R_G + R_F} \frac{RCs + 1}{\frac{(RR_G + RR_F + R_G R_F)}{(R_G + R_F)} Cs + 1} \quad (7\text{–}23)$$

Referring to Figure 7–21, a pole is introduced at $\omega = 1/RC$, and this pole reduces the gain 3 dB at the breakpoint. When the zero occurs prior to the first op amp pole it cancels out the phase shift caused by the $\omega = 1/RC$ pole. The phase shift is completely canceled before the second op amp pole occurs, and the circuit reacts as if the pole was never introduced. Nevertheless, $A\beta$ is reduced by 3 dB or more, so the loop gain crosses the 0-dB axis at a lower frequency. The beauty of lead lag compensation is that the closed-loop ideal gain is not affected as is shown below. The Thevenin equivalent of the input circuit is calculated in Equation 7–24, the circuit gain in terms of Thevenin equivalents is calculated in Equation 7–25, and the ideal closed-loop gain is calculated in Equation 7–26.

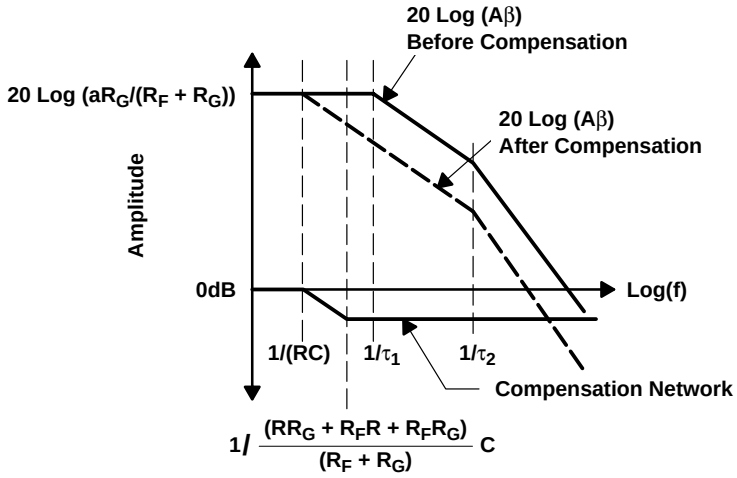


Figure 7–21. Bode Plot of Lead-Lag Compensated Op Amp

$$V_{TH} = V_{IN}\frac{R + \frac{1}{Cs}}{R + R_G + \frac{1}{Cs}} \qquad R_{TH} = \frac{R_G\left(R + \frac{1}{Cs}\right)}{R + R_G + \frac{1}{Cs}} \tag{7–24}$$

$$V_{OUT} = -V_{TH}\frac{R_F}{R_{TH}} \tag{7–25}$$

$$-\frac{V_{OUT}}{V_{IN}} = \frac{R + \frac{1}{Cs}}{R + R_G + \frac{1}{Cs}}\frac{R_F}{\frac{R_G\left(R+\frac{1}{Cs}\right)}{R+R_G+\frac{1}{Cs}}} = \frac{R_F}{R_G} \tag{7–26}$$

Equation 7–26 is intuitively obvious because the RC network is placed across a virtual ground. As long as the loop gain, Aβ, is large, the feedback will null out the closed-loop effect of RC, and the circuit will function as if it were not there. The closed-loop log plot of the lead-lag-compensated op amp is given in Figure 7–22. Notice that the pole and zero resulting from the compensation occur and are gone before the first amplifier poles come on the scene. This prevents interaction, but it is not required for stability.

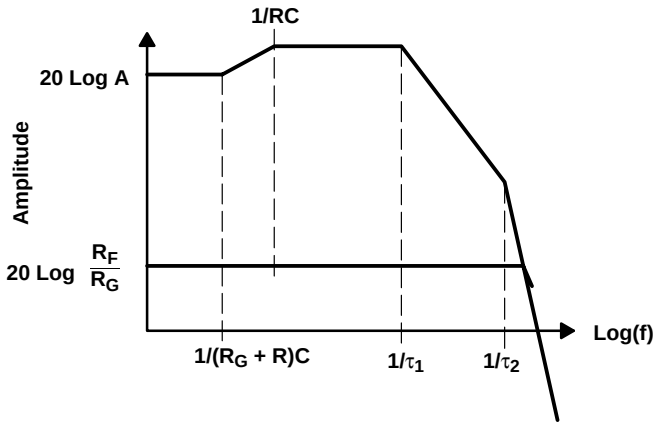


Figure 7–22. Closed-Loop Plot of Lead-Lag Compensated Op Amp

7.9 Comparison of Compensation Schemes

Internally compensated op amps can, and often do, oscillate under some circuit conditions. Internally compensated op amps need an external pole to get the oscillation or ringing started, and circuit stray capacitances often supply the phase shift required for instability. Loads, such as cables, often cause internally compensated op amps to ring severely.

Dominant pole compensation is often used in IC design because it is easy to implement. It rolls off the closed-loop gain early; thus, it is seldom used as an external form of compensation unless filtering is required. Load capacitance, depending on its pole location, usually causes the op amp to ring. Large load capacitance can stabilize the op amp because it acts as dominant pole compensation.

The simplest form of compensation is gain compensation. High closed-loop gains are reflected in lower loop gains, and in turn, lower loop gains increase stability. If an op amp circuit can be stabilized by increasing the closed-loop gain, do it.

Stray capacitance across the feedback resistor tends to stabilize the op amp because it is a form of lead compensation. This compensation scheme is useful for limiting the circuit bandwidth, but it decreases the closed-loop gain.

Stray capacitance on the inverting input works with the parallel combination of the feedback and gain setting resistors to form a pole in the Bode plot, and this pole decreases the circuit's stability. This effect is normally observed in high-impedance circuits built with CMOS op amps. Adding a feedback capacitor forms a compensated attenuator scheme that cancels out the input pole. The cancellation occurs when the input and feedback RC time constants are equal. Under the conditions of equal time constants, the op amp functions as though the stray input capacitance was not there. An excellent method of implementing a compensated attenuator is to build a stray feedback capacitor using the ground plane and a trace off the output node.

Lead-lag compensation stabilizes the op amp, and it yields the best closed-loop frequency performance. Contrary to some published opinions, no compensation scheme will increase the bandwidth beyond that of the op amp. Lead-lag compensation just gives the best bandwidth for the compensation.

7.10 Conclusions

The stability criteria often is not oscillation, rather it is circuit performance as exhibited by peaking and ringing.

The circuit bandwidth can often be increased by connecting an external capacitor in parallel with the op amp. Some op amps have hooks that enable a parallel capacitor to be connected in parallel with a portion of the input stages. This increases bandwidth because it shunts high frequencies past the low bandwidth g_m stages, but this method of compensation depends on the op amp type and manufacturer.

The compensation techniques given here are adequate for the majority of applications. When the new and challenging application presents itself, use the procedure outlined here to invent your own compensation technique.

Chapter 8

Current-Feedback Op Amp Analysis

Ron Mancini

8.1 Introduction

Current-feedback amplifiers (CFA) do not have the traditional differential amplifier input structure, thus they sacrifice the parameter matching inherent to that structure. The CFA circuit configuration prevents them from obtaining the precision of voltage-feedback amplifiers (VFA), but the circuit configuration that sacrifices precision results in increased bandwidth and slew rate. The higher bandwidth is relatively independent of closed-loop gain, so the constant gain-bandwidth restriction applied to VFAs is removed for CFAs. The slew rate of CFAs is much improved from their counterpart VFAs because their structure enables the output stage to supply slewing current until the output reaches its final value. In general, VFAs are used for precision and general purpose applications, while CFAs are restricted to high frequency applications above 100 MHz.

Although CFAs do not have the precision of their VFA counterparts, they are precise enough to be dc-coupled in video applications where dynamic range requirements are not severe. CFAs, unlike previous generation high-frequency amplifiers, have eliminated the ac coupling requirement; they are usually dc-coupled while they operate in the GHz range. CFAs have much faster slew rates than VFAs, so they have faster rise/fall times and less intermodulation distortion.

8.2 CFA Model

The CFA model is shown in Figure 8–1. The noninverting input of a CFA connects to the input of the input buffer, so it has very high impedance similar to that of a bipolar transistor noninverting VFA input. The inverting input connects to the input buffer's output, so the inverting input impedance is equivalent to a buffer's output impedance, which is very low. Z_B models the input buffer's output impedance, and it is usually less than 50 Ω. The input buffer gain, G_B, is as close to one as IC design methods can achieve, and it is small enough to neglect in the calculations.

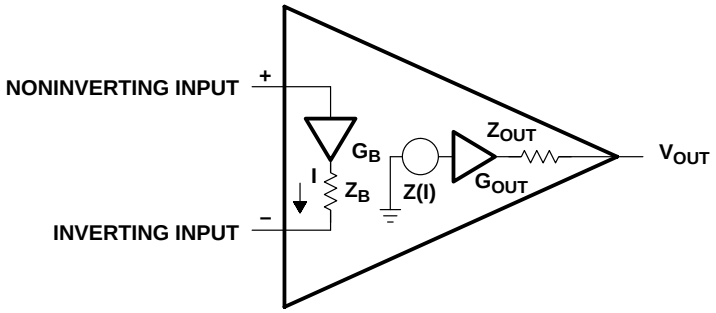


Figure 8–1. Current-Feedback Amplifier Model

The output buffer provides low output impedance for the amplifier. Again, the output buffer gain, G_{OUT}, is very close to one, so it is neglected in the analysis. The output impedance of the output buffer is ignored during the calculations. This parameter may influence the circuit performance when driving very low impedance or capacitive loads, but this is usually not the case. The input buffer's output impedance can't be ignored because affects stability at high frequencies.

The current-controlled current source, Z, is a transimpedance. The transimpedance in a CFA serves the same function as gain in a VFA; it is the parameter that makes the performance of the op amp dependent only on the passive parameter values. Usually the transimpedance is very high, in the MΩ range, so the CFA gains accuracy by closing a feedback loop in the same manner that the VFA does.

8.3 Development of the Stability Equation

The stability equation is developed with the aid of Figure 8–2. Remember, stability is independent of the input, and stability depends solely on the loop gain, Aβ. Breaking the loop at point *X*, inserting a test signal, V_{TI}, and calculating the return signal V_{TO} develops the stability equation.

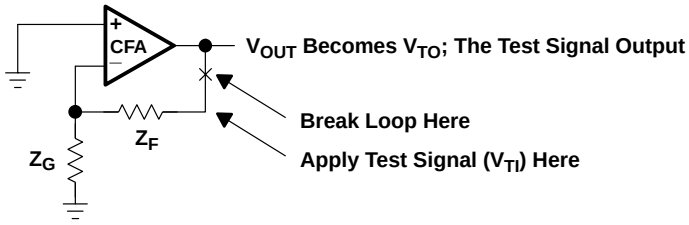


Figure 8–2. Stability Analysis Circuit

The circuit used for stability calculations is shown in Figure 8–3 where the model of Figure 8–1 is substituted for the CFA symbol. The input and output buffer gain, and output buffer

output impedance have been deleted from the circuit to simplify calculations. This approximation is valid for almost all applications.

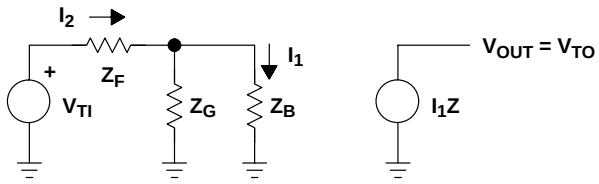


Figure 8–3. Stability Analysis Circuit

The transfer equation is given in Equation 8–1, and the Kirchoff"s law is used to write Equations 8–2 and 8–3.

$$V_{TO} = I_1 Z \tag{8–1}$$

$$V_{TI} = I_2\left(Z_F + Z_G \parallel Z_B\right) \tag{8–2}$$

$$I_2\left(Z_G \parallel Z_B\right) = I_1 Z_B \tag{8–3}$$

Equations 8–2 and 8–3 are combined to yield Equation 8–4.

$$V_{TI} = I_1\left(Z_F + Z_G \parallel Z_B\right)\left(1 + \frac{Z_B}{Z_G}\right) = I_1 Z_F\left(1 + \frac{Z_B}{Z_F \parallel Z_G}\right) \tag{8–4}$$

Dividing Equation 8–1 by Equation 8–4 yields Equation 8–5, and this is the open loop transfer equation. This equation is commonly known as the loop gain.

$$A\beta = \frac{V_{TO}}{V_{TI}} = \frac{Z}{\left(Z_F\left(1 + \frac{Z_B}{Z_F \parallel Z_G}\right)\right)} \tag{8–5}$$

8.4 The Noninverting CFA

The closed-loop gain equation for the noninverting CFA is developed with the aid of Figure 8–4, where external gain setting resistors have been added to the circuit. The buffers are shown in Figure 8–4, but because their gains equal one and they are included within the feedback loop, the buffer gain does not enter into the calculations.

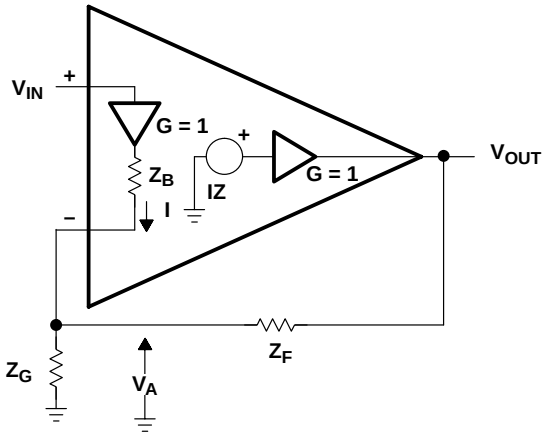


Figure 8–4. Noninverting CFA

Equation 8–6 is the transfer equation, Equation 8–7 is the current equation at the inverting node, and Equation 8–8 is the input loop equation. These equations are combined to yield the closed-loop gain equation, Equation 8–9.

$$V_{OUT} = IZ \tag{8–6}$$

$$I = \left(\frac{V_A}{Z_G}\right) - \left(\frac{V_{OUT} - V_A}{Z_F}\right) \tag{8–7}$$

$$V_A = V_{IN} - IZ_B \tag{8–8}$$

$$\frac{V_{OUT}}{V_{IN}} = \frac{\dfrac{Z\left(1 + \dfrac{Z_F}{Z_G}\right)}{Z_F\left(1 + \dfrac{Z_B}{Z_F \| Z_G}\right)}}{1 + \dfrac{Z}{Z_F\left(1 + \dfrac{Z_B}{Z_F \| Z_G}\right)}} \tag{8–9}$$

When the input buffer output impedance, Z_B, approaches zero, Equation 8–9 reduces to Equation 8–10.

$$\frac{V_{OUT}}{V_{IN}} = \frac{\frac{Z\left(1+\frac{Z_F}{Z_G}\right)}{Z_F}}{1+\frac{Z}{Z_F}} = \frac{1+\frac{Z_F}{Z_G}}{1+\frac{Z_F}{Z}} \qquad (8\text{–}10)$$

When the transimpedance, Z, is very high, the term Z_F/Z in Equation 8–10 approaches zero, and Equation 8–10 reduces to Equation 8–11; the ideal closed-loop gain equation for the CFA. The ideal closed-loop gain equations for the CFA and VFA are identical, and the degree to which they depart from ideal is dependent on the validity of the assumptions. The VFA has one assumption that the direct gain is very high, while the CFA has two assumptions, that the transimpedance is very high and that the input buffer output impedance is very low. As would be expected, two assumptions are much harder to meet than one, thus the CFA departs from the ideal more than the VFA does.

$$\frac{V_{OUT}}{V_{IN}} = 1 + \frac{Z_F}{Z_G} \qquad (8\text{–}11)$$

8.5 The Inverting CFA

The inverting CFA configuration is seldom used because the inverting input impedance is very low ($Z_B||Z_F + Z_G$). When Z_G is made dominant by selecting it as a high resistance value it overrides the effect of Z_B. Z_F must also be selected as a high value to achieve at least unity gain, and high values for Z_F result in poor bandwidth performance, as we will see in the next section. If Z_G is selected as a low value the frequency sensitive Z_B causes the gain to increase as frequency increases. These limitations restrict inverting applications of the inverting CFA.

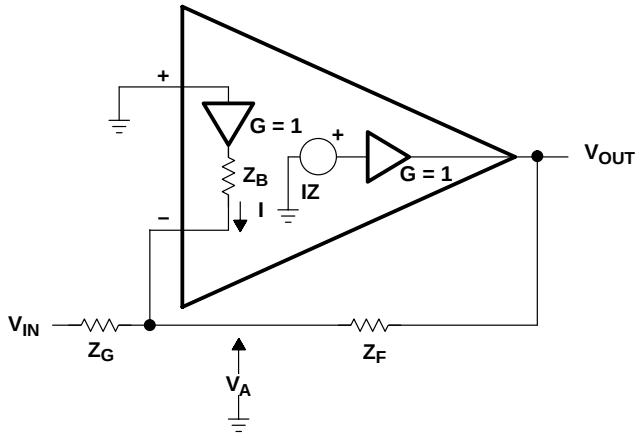


Figure 8–5. Inverting CFA

The current equation for the input node is written as Equation 8–12. Equation 8–13 defines the dummy variable, V_A, and Equation 8–14 is the transfer equation for the CFA. These equations are combined and simplified leading to Equation 8–15, which is the closed-loop gain equation for the inverting CFA.

$$I + \frac{V_{IN} - V_A}{Z_G} = \frac{V_A - V_{OUT}}{Z_F} \tag{8–12}$$

$$IZ_B = -V_A \tag{8–13}$$

$$IZ = V_{OUT} \tag{8–14}$$

$$\frac{V_{OUT}}{V_{IN}} = -\frac{\dfrac{Z}{Z_G\left(1 + \dfrac{Z_B}{Z_F \| Z_G}\right)}}{1 + \dfrac{Z}{Z_F\left(1 + \dfrac{Z_B}{Z_F \| Z_G}\right)}} \tag{8–15}$$

When Z_B approaches zero, Equation 8–15 reduces to Equation 8–16.

$$\frac{V_{OUT}}{V_{IN}} = -\frac{\dfrac{1}{Z_G}}{\dfrac{1}{Z} + \dfrac{1}{Z_F}} \tag{8–16}$$

When Z is very large, Equation 8–16 becomes Equation 8–17, which is the ideal closed-loop gain equation for the inverting CFA.

$$\frac{V_{OUT}}{V_{IN}} = -\frac{Z_F}{Z_G} \tag{8–17}$$

The ideal closed-loop gain equation for the inverting VFA and CFA op amps are identical. Both configurations have lower input impedance than the noninverting configuration has, but the VFA has one assumption while the CFA has two assumptions. Again, as was the case with the noninverting counterparts, the CFA is less ideal than the VFA because of the two assumptions. The zero Z_B assumption always breaks down in bipolar junction transistors as is shown later. The CFA is almost never used in the differential amplifier configuration because of the CFA's gross input impedance mismatch.

8.6 Stability Analysis

The stability equation is repeated as Equation 8–18.

$$A\beta = \frac{V_{TO}}{V_{TI}} = \frac{Z}{\left(Z_F\left(1 + \frac{Z_B}{Z_F \| Z_G}\right)\right)} \tag{8–18}$$

Comparing Equations 8–9 and 8–15 to Equation 8–18 reveals that the inverting and noninverting CFA op amps have identical stability equations. This is the expected result because stability of any feedback circuit is a function of the loop gain, and the input signals have no affect on stability. The two op amp parameters affecting stability are the transimpedance, Z, and the input buffer's output impedance, Z_B. The external components affecting stability are Z_G and Z_F. The designer controls the external impedance, although stray capacitance that is a part of the external impedance sometimes seems to be uncontrollable. Stray capacitance is the primary cause of ringing and overshoot in CFAs. Z and Z_B are CFA op amp parameters that can't be controlled by the circuit designer, so he has to live with them.

Prior to determining stability with a Bode plot, we take the log of Equation 8–18, and plot the logs (Equations 8–19 and 8–20) in Figure 8–6.

$$20 \text{ LOG } |A\beta| = 20 \text{ LOG } |Z| - 20 \text{ LOG } \left|Z_F\left(1 + \frac{Z_B}{Z_F \| Z_B}\right)\right| \tag{8–19}$$

$$\phi = \text{TANGENT}^{-1}(A\beta) \tag{8–20}$$

This enables the designer to add and subtract components of the stability equation graphically.

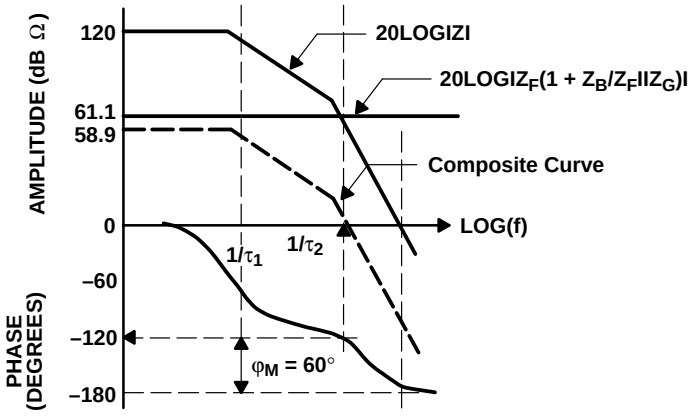


Figure 8–6. Bode Plot of Stability Equation

The plot in Figure 8–6 assumes typical values for the parameters:

$$Z = \frac{1M\Omega}{(1 + \tau_1 S)(1 + \tau_2 S)} \tag{8–21}$$

$$Z_B = 70\Omega \tag{8–22}$$

$$Z_G = Z_F = 1k\Omega \tag{8–23}$$

The transimpedance has two poles and the plot shows that the op amp will be unstable without the addition of external components because 20 LOG|Z| crosses the 0-dB axis after the phase shift is 180°. Z_F, Z_B, and Z_G reduce the loop gain 61.1 dB, so the circuit is stable because it has 60°-phase margin. Z_F is the component that stabilizes the circuit. The parallel combination of Z_F and Z_G contribute little to the phase margin because Z_B is very small, so Z_B and Z_G have little effect on stability.

The manufacturer determines the optimum value of R_F during the characterization of the IC. Referring to Figure 8–6, it is seen that when R_F exceeds the optimum value recommended by the IC manufacturer, stability increases. The increased stability has a price called decreased bandwidth. Conversely, when R_F is less than the optimum value recommended by the IC manufacturer, stability decreases, and the circuit response to step inputs is overshoot or possibly ringing. Sometimes the overshoot associated with less than optimum R_F is tolerated because the bandwidth increases as R_F decreases. The peaked response associated with less than optimum values of R_F can be used to compensate for cable droop caused by cable capacitance.

When $Z_B = 0\ \Omega$ and $Z_F = R_F$ the loop gain equation is; $A\beta = Z/R_F$. Under these conditions Z and R_F determine stability, and a value of R_F can always be found to stabilize the circuit. The transimpedance and feedback resistor have a major impact on stability, and the input buffer's output impedance has a minor effect on stability. Since Z_B increases with an increase in frequency, it tends to increase stability at higher frequencies. Equation 8–18 is rewritten as Equation 8–24, but it has been manipulated so that the ideal closed-loop gain is readily apparent.

$$A\beta = \frac{Z}{Z_F + Z_B\left(1 + \frac{R_F}{R_G}\right)} \tag{8–24}$$

The closed-loop ideal gain equation (inverting and noninverting) shows up in the denominator of Equation 8–24, so the closed-loop gain influences the stability of the op amp. When Z_B approaches zero, the closed-loop gain term also approaches zero, and the op amp becomes independent of the ideal closed-loop gain. Under these conditions R_F determines stability, and the bandwidth is independent of the closed-loop gain. Many people claim that the CFA bandwidth is independent of the gain, and that claim's validity is dependent on the ratios Z_B/Z_F being very low.

Z_B is important enough to warrant further investigation, so the equation for Z_B is given below.

$$Z_B \cong h_{ib} + \frac{R_B}{\beta_0 + 1}\left(\frac{1 + \frac{s\beta_0}{\omega_T}}{1 + \frac{S\beta_0}{(\beta_0+1)\omega_T}}\right) \quad (8\text{–}25)$$

At low frequencies $h_{ib} = 50\ \Omega$ and $R_B/(\beta_0+1) = 25\ \Omega$, so $Z_B = 75\ \Omega$. Z_B varies in accordance with Equation 8–25 at high frequencies. Also, the transistor parameters in Equation 8–25 vary with transistor type; they are different for NPN and PNP transistors. Because Z_B is dependent on the output transistors being used, and this is a function of the quadrant the output signal is in, Z_B has an extremely wide variation. Z_B is a small factor in the equation, but it adds a lot of variability to the current-feedback op amp.

8.7 Selection of the Feedback Resistor

The feedback resistor determines stability, and it affects closed-loop bandwidth, so it must be selected very carefully. Most CFA IC manufacturers employ applications and product engineers who spend a great deal of time and effort selecting R_F. They measure each noninverting gain with several different feedback resistors to gather data. Then they pick a compromise value of R_F that yields stable operation with acceptable peaking, and that value of R_F is recommended on the data sheet for that specific gain. This procedure is repeated for several different gains in anticipation of the various gains their customer applications require (often G = 1, 2, or 5). When the value of R_F or the gain is changed from the values recommended on the data sheet, bandwidth and/or stability is affected.

When the circuit designer must select a different R_F value from that recommended on the data sheet he gets into stability or low bandwidth problems. Lowering R_F decreases stability, and increasing R_F decreases bandwidth. What happens when the designer needs to operate at a gain not specified on the data sheet? The designer must select a new value of R_F for the new gain, but there is no guarantee that new value of R_F is an optimum value. One solution to the R_F selection problem is to assume that the loop gain, $A\beta$, is a linear function. Then the assumption can be made that $(A\beta)_1$ for a gain of one equals $(A\beta)_N$ for a gain of N, and that this is a linear relationship between stability and gain. Equations 8–26 and 8–27 are based on the linearity assumption.

$$\frac{Z}{Z_{F1} + Z_B\left(1 + \frac{Z_{F1}}{Z_{G1}}\right)} = \frac{Z}{Z_{FN} + Z_B\left(1 + \frac{Z_{FN}}{Z_{GN}}\right)} \quad (8\text{–}26)$$

$$Z_{FN} = Z_{F1} + Z_B\left(\left(1 + \frac{Z_{F1}}{Z_{G1}}\right) - \left(1 + \frac{Z_{FN}}{Z_{GN}}\right)\right) \quad (8\text{–}27)$$

Equation 8–27 leads one to believe that a new value for Z_F can easily be chosen for each new gain. This is not the case in the real world; the assumptions don't hold up well enough to rely on them. When you change to a new gain not specified on the data sheet, Equation 8–27, at best, supplies a starting point for R_F, but you must test to determine the final value of R_F.

When the R_F value recommended on the data sheet can't be used, an alternate method of selecting a starting value for R_F is to use graphical techniques. The graph shown in Figure 8–7 is a plot of the typical 300-MHz CFA data given in Table 8–1.

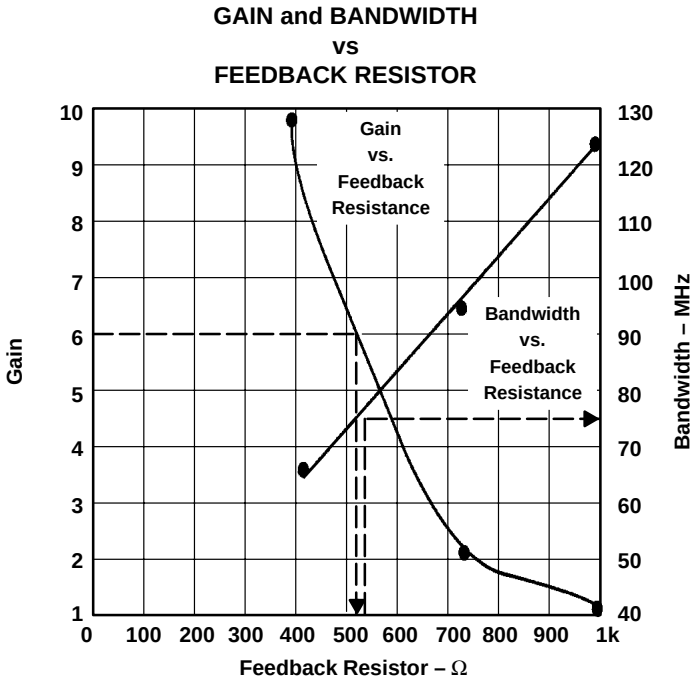


Figure 8–7. Plot of CFA R_F, G, and BW

Table 8–1. Data Set for Curves in Figure 8–7

GAIN (A_{CL})	R_F (Ω)	BANDWIDTH (MHz)
+ 1	1000	125
+ 2	681	95
+ 10	383	65

Enter the graph at the new gain, say A_{CL} = 6, and move horizontally until you reach the intersection of the gain versus feedback resistance curve. Then drop vertically to the resistance axis and read the new value of R_F (500 Ω in this example). Enter the graph at the new value of R_F, and travel vertically until you intersect the bandwidth versus feedback resistance curve. Now move to the bandwidth axis to read the new bandwidth (75 MHz in this example). As a starting point you should expect to get approximately 75 MHz BW with a gain of 6 and R_F = 500 Ω. Although this technique yields more reliable solutions than Equation 8–27 does, op amp peculiarities, circuit board stray capacitances, and wiring make extensive testing mandatory. The circuit must be tested for performance and stability at each new operating point.

8.8 Stability and Input Capacitance

When designer lets the circuit board introduce stray capacitance on the inverting input node to ground, it causes the impedance Z_G to become reactive. The new impedance, Z_G, is given in Equation 8–28, and Equation 8–29 is the stability equation that describes the situation.

$$Z_G = \frac{R_G}{1 + R_G C_G s} \tag{8–28}$$

$$A\beta = \frac{Z}{Z_B + \dfrac{Z_F}{Z_G^2 + Z_B Z_G}} \tag{8–29}$$

$$A\beta = \frac{Z}{R_F\left(1 + \dfrac{R_B}{R_F \| R_G}\right)\left(1 + R_B \parallel R_F \parallel R_G C_G s\right)} \tag{8–30}$$

Equation 8–29 is the stability equation when Z_G consists of a resistor in parallel with stray capacitance between the inverting input node and ground. The stray capacitance, C_G, is a fixed value because it is dependent on the circuit layout. The pole created by the stray capacitance is dependent on R_B because it dominates R_F and R_G. R_B fluctuates with manufacturing tolerances, so the $R_B C_G$ pole placement is subject to IC manufacturing tolerances. As the $R_B C_G$ combination becomes larger, the pole moves towards the zero fre-

quency axis, lowering the circuit stability. Eventually it interacts with the pole contained in Z, $1/\tau_2$, and instability results.

The effects of stray capacitance on CFA closed-loop performance are shown in Figure 8–8.

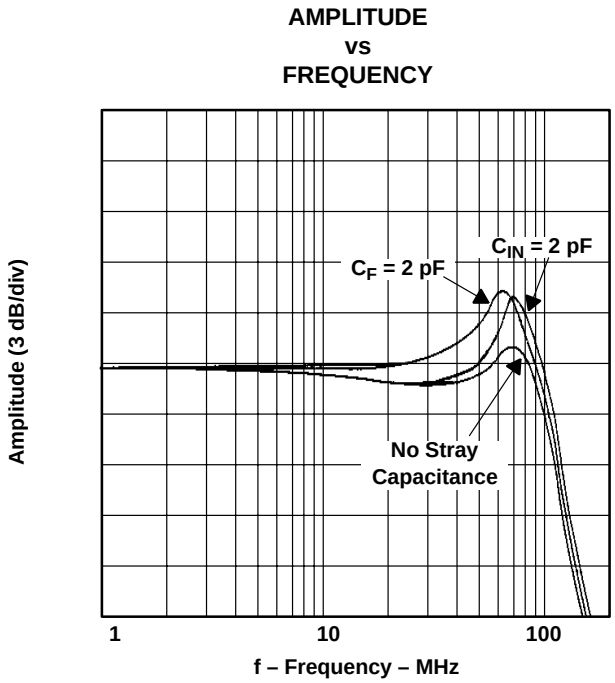


Figure 8–8. Effects of Stray Capacitance on CFAs

Notice that the introduction of C_G causes more than 3 dB peaking in the CFA frequency response plot, and it increases the bandwidth about 18 MHz. Two picofarads are not a lot of capacitance because a sloppy layout can easily add 4 or more picofarads to the circuit.

8.9 Stability and Feedback Capacitance

When a stray capacitor is formed across the feedback resistor, the feedback impedance is given by Equation 8–31. Equation 8–32 gives the loop gain when a feedback capacitor has been added to the circuit.

$$Z_F = \frac{R_F}{1 + R_F C_F s} \tag{8–31}$$

$$A\beta = \frac{Z\left(1 + R_F C_F s\right)}{R_F\left(1 + \frac{R_B}{R_F \| R_G}\right)\left(1 + R_B \| R_F \| R_G C_F s\right)} \tag{8–32}$$

This loop gain transfer function contains a pole and zero, thus, depending on the pole/zero placement, oscillation can result. The Bode plot for this case is shown in Figure 8–9. The original and composite curves cross the 0-dB axis with a slope of –40 dB/decade, so either curve can indicate instability. The composite curve crosses the 0-dB axis at a higher frequency than the original curve, hence the stray capacitance has added more phase shift to the system. The composite curve is surely less stable than the original curve. Adding capacitance to the inverting input node or across the feedback resistor usually results in instability. R_B largely influences the location of the pole introduced by C_F, thus here is another case where stray capacitance leads to instability.

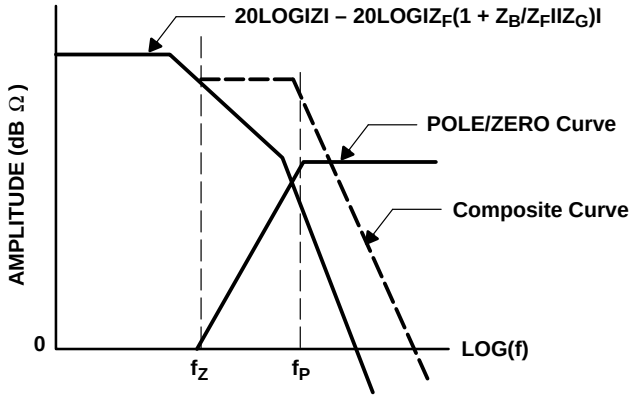


Figure 8–9. Bode Plot with C_F

Figure 8–8 shows that C_F = 2 pF adds about 4 dB of peaking to the frequency response plot. The bandwidth increases about 10 MHz because of the peaking. C_F and C_G are the major causes of overshoot, ringing, and oscillation in CFAs, and the circuit board layout must be carefully done to eliminate these stray capacitances.

8.10 Compensation of C_F and C_G

When C_F and C_G both are present in the circuit they may be adjusted to cancel each other out. The stability equation for a circuit with C_F and C_G is Equation 8–33.

$$A\beta = \frac{Z\left(1 + R_F C_F s\right)}{R_F\left(1 + \frac{R_B}{R_F \| R_G}\right)\left(R_B \| R_F \| R_G \left(C_F + C_G\right) s + 1\right)} \quad (8\text{–}33)$$

If the zero and pole in Equation 8–33 are made to cancel each other, the only poles remaining are in Z. Setting the pole and zero in Equation 8–33 equal yields Equation 8–34 after some algebraic manipulation.

$$R_F C_F = C_G\left(R_G \| R_B\right) \quad (8\text{–}34)$$

R_B dominates the parallel combination of R_B and R_G, so Equation 8–34 is reduced to Equation 8–35.

$$R_F C_F = R_B C_G \quad (8\text{–}35)$$

R_B is an IC parameter, so it is dependent on the IC process. R_B it is an important IC parameter, but it is not important enough to be monitored as a control variable during the manufacturing process. R_B has widely spread, unspecified parameters, thus depending on R_B for compensation is risky. Rather, the prudent design engineer assures that the circuit will be stable for any reasonable value of R_B, and that the resulting frequency response peaking is acceptable.

8.11 Summary

Constant gain-bandwidth is not a limiting criterion for the CFA, so the feedback resistor is adjusted for maximum performance. Stability is dependent on the feedback resistor; as R_F is decreased, stability is decreased, and when R_F goes to zero the circuit becomes unstable. As R_F is increased stability increases, but the bandwidth decreases.

The inverting input impedance is very high, but the noninverting input impedance is very low. This situation precludes CFAs from operation in the differential amplifier configuration. Stray capacitance on the inverting input node or across the feedback resistor always leads to peaking, usually to ringing, and sometimes to oscillations. A prudent circuit designer scans the PC board layout for stray capacitances, and he eliminates them. Breadboarding and lab testing are a must with CFAs. The CFA performance can be improved immeasurably with a good layout, good decoupling capacitors, and low inductance components.

Chapter 9

Voltage- and Current-Feedback Op Amp Comparison

Ron Mancini and James Karki

9.1 Introduction

The name, operational amplifier, was given to voltage-feedback amplifiers (VFA) when they were the only op amps in existence. These new (they were new in the late ’40s) amplifiers could be programmed with external components to perform various math operations on a signal; thus, they were nicknamed op amps. Current-feedback amplifiers (CFA) have been around approximately twenty years, but their popularity has only increased in the last several years. Two factors limiting the popularity of CFAs is their application difficulty and lack of precision.

The VFA is familiar component, and there are several variations of internally compensated VFAs that can be used with little applications work. Because of its long history, the VFA comes in many varieties and packages, so there are VFAs applicable to almost any job. VFA bandwidth is limited, so it can’t function as well at high signal frequencies as the CFA can. For now, the signal frequency and precision separates the applications of the two op amp configurations.

The VFA has some other redeeming virtues, such as excellent precision, that makes it the desirable amplifier in low frequency applications. Many functions other than signal amplification are accomplished at low frequencies, and functions like level-shifting a signal require precision. Fortunately, precision is not required in most high frequency applications where amplification or filtering of a signal is predominant, so CFAs are suitable to high frequency applications. The lack of precision coupled with the application difficulties prevents the CFA from replacing the VFA.

9.2 Precision

The long-tailed pair input structure gives the VFA its precision; the long-tailed pair is shown in Figure 9–1

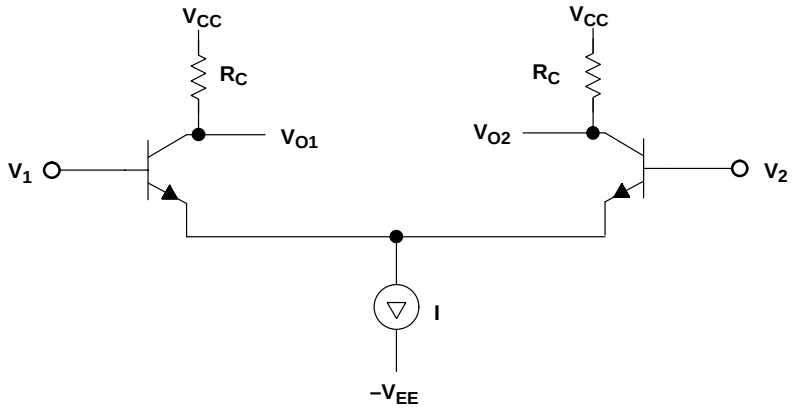


Figure 9–1. Long-Tailed Pair

The transistors, Q_1 and Q_2, are very carefully matched for initial and drift tolerances. Careful attention is paid to detail in the transistor design to insure that parameters like current gain, β, and base-emitter voltage, V_{BE}, are matched between the input transistors, Q_1 and Q_2. When $V_{B1} = V_{B2,}$ the current, I, splits equally between the transistors, and $V_{O1} = V_{O2}$. As long as the transistor parameters are matched, the collector currents stay equal. The slightest change of V_{B1} with respect to V_{B2} causes a mismatch in the collector currents and a differential output voltage $|V_{B1}–V_{B2}|$.

When temperature or other outside influences change transistor parameters like current gain or base-emitter voltage, as long as the change is equal, it causes no change in the differential output voltage. IC designers go to great lengths to ensure that transistor parameter changes due to external influences do not cause a differential output voltage change. Now, the slightest change in either base voltage causes a differential output voltage change, and gross changes in external conditions do not cause a differential output voltage change. This is the formula for a precision amplifier because it can amplify small input changes while ignoring changes in the parameters or ambient conditions.

This is a simplified explanation, and there are many different techniques used to ensure transistor matching. Some of the techniques used to match input transistors are parameter trimming, special layout techniques, thermal balancing, and symmetrical layouts. The long-tailed pair is an excellent circuit configuration for obtaining precision in the input circuit, but the output circuit has one fault. The output circuit collector impedance has to be high to achieve high gain in the first stage. High impedance coupled with the Miller capacitance discussed in Chapter 7 forms a quasidominant pole compensation circuit that has poor high frequency response.

The noninverting input of the CFA (see Figure 9–2) connects to a buffer input inside the op amp. The inverting input of the CFA connects to a buffer output inside the CFA. Buffer inputs and outputs have dramatically different impedance levels, so any matching becomes a moot point. The buffer can't reject common-mode voltages introduced by parameter drifts because it has no common-mode rejection capability. The input current causes a voltage drop across the input buffer's output impedance, R_B, and there is no way that this voltage drop can be distinguished from an input signal.

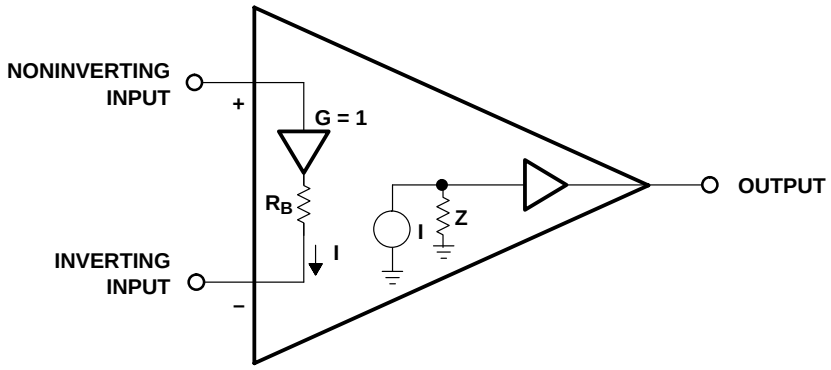


Figure 9–2. Ideal CFA

The CFA circuit configuration was selected for high frequency amplification because it has current-controlled gain and a current-dominant input. Being a current device, the CFA does not have the Miller-effect problem that the VFA has. The input structure of the CFA sacrifices precision for bandwidth, but CFAs achieve usable bandwidths ten times the usable VFA bandwidth.

9.3 Bandwidth

The bandwidth of a circuit is defined by high frequency errors. When the gain falls off at high frequencies unequal frequency amplification causes the signal to become distorted. The signal loses its high frequency components; an example of high frequency signal degradation is a square wave with sharp corners that is amplified and turned into slump cornered semi sine wave. The error equation for any feedback circuit is repeated in Equation 9–1.

$$E = \frac{V_{IN}}{1 + A\beta} \tag{9–1}$$

This equation is valid for any feedback circuit, so it applies equally to a VFA or a CFA. The loop gain equation for any VFA is repeated as Equation 9–2.

$$A\beta = \frac{aR_G}{R_F + R_G} \tag{9–2}$$

Equation 9–2 is rewritten below as Equations 9–3 and 9–4 for the noninverting and inverting circuits respectively. In each case, the symbol G_{CLNI} and G_{CLI} represent the closed loop gain for the noninverting and inverting circuits respectively.

$$A\beta = \frac{a}{\frac{R_F + R_G}{R_G}} = \frac{a}{G_{CLNI}} \tag{9–3}$$

$$A\beta = \frac{a}{\frac{R_F + R_G}{R_G}} = \frac{a}{G_{CLI} + 1} \tag{9–4}$$

In both cases the loop gain decreases as the closed loop gain increases, thus all VFA errors increase as the closed loop gain increases. The error increase is mathematically coupled to the closed loop gain equation, so there is no working around this fact. For the VFA, effective bandwidth decreases as the closed loop gain increases because the loop gain decreases as the closed loop gain increases.

A plot of the VFA loop gain, closed loop gain, and error is given in Figure 9–3. Referring to Figure 9–3, the direct gain, A, is the op amp open loop gain, a, for a noninverting op amp. The direct gain for an inverting op amp is $(a(Z_F/(Z_G + Z_F)))$. The Miller effect causes the direct gain to fall off at high frequencies, thus error increases as frequency increases because the effective loop gain decreases. At a given frequency, the error also increases when the closed loop gain is increased.

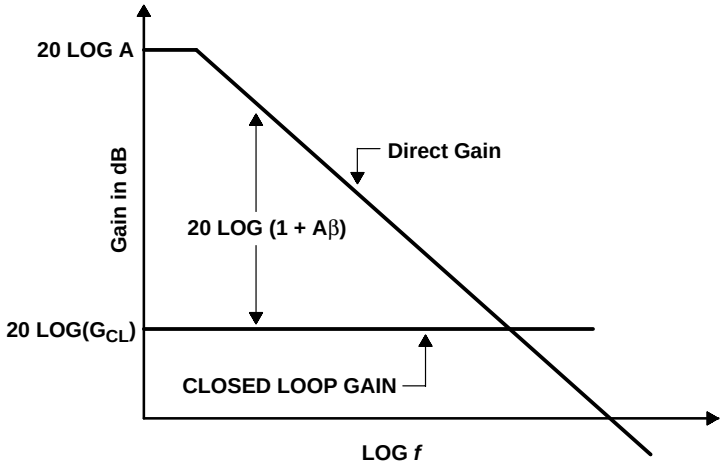


Figure 9–3. VFA Gain versus Frequency

The CFA is a current operated device; hence, it not nearly as subject to the Miller effect resulting from stray capacitance as the VFA is. The absence of the Miller effect enables the CFA's frequency response to hold up far better than the VFA's does. A plot of the CFA loop gain, transimpedance, and error is given in Figure 9–4. Notice that the transimpedance stays at the large low frequency intercept value until much higher frequencies than the VFA does.

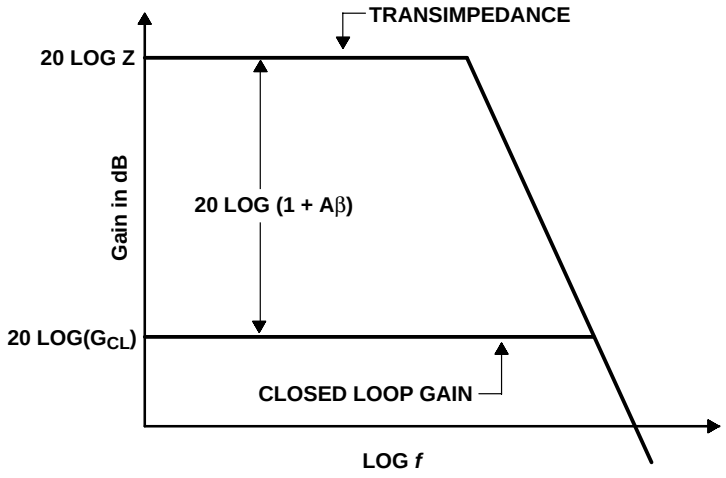


Figure 9–4. CFA Gain vs Frequency

The loop gain equation for the CFA is repeated here as Equation 9–5.

$$A\beta = \frac{Z}{R_F\left(1 + \frac{R_B}{R_F \| R_G}\right)} \tag{9–5}$$

When the input buffer output resistance approaches zero, Equation 9–5 reduces to Equation 9–6.

$$A\beta = \frac{Z}{R_F} \tag{9–6}$$

Equation 9–6 shows that the closed-loop gain has no effect on the loop gain when R_B = 0, so under ideal conditions one would expect the transimpedance to fall off with a zero slope. Figure 9–4 shows that there is a finite slope, but much less than that of a VFA, and the slope is caused by R_B not being equal to zero. For example, R_B is usually 50 Ω when R_F = 1000 Ω at A_{CL} = 1. If we let $R_F = R_G$, then $R_F||R_G$ = 500 Ω, and $R_B/R_F||R_G$ = 50/500 = 0.1.

Substituting this value into Equation 9–6 yields Equation 9–7, and Equation 9–7 is almost identical to Equation 9–6. R_B does cause some interaction between the loop gain and the transimpedance, but because the interaction is secondary the CFA gain falls off with a faster slope.

$$A\beta = \frac{Z}{1.1\ R_F} \tag{9–7}$$

The direct gain of a VFA starts falling off early, often at 10 Hz or 100 Hz, but the transimpedance of a CFA does not start falling off until much higher frequencies. The VFA is constrained by the gain-bandwidth limitation imposed by the closed loop gain being incorporated within the loop gain. The CFA, with the exception of the effects of R_B, does not have this constraint. This adds up to the CFA being the superior high frequency amplifier.

9.4 Stability

Stability in a feedback system is defined by the loop gain, and no other factor, including the inputs or type of inputs, affects stability. The loop gain for a VFA is given in Equation 9–2. Examining Equation 9–2 we see that the stability of a VFA is depends on two items; the op amp transfer function, a, and the gain setting components, Z_F/Z_G.

The op amp contains many poles, and if it is not internally compensated, it requires external compensation. The op amp always has at least one dominant pole, and the most phase margin that an op amp has is 45°. Phase margins beyond 60° are a waste of op amp bandwidth. When poles and zeros are contained in Z_F and Z_G, they can compensate for the op amp phase shift or add to its instability. In any case, the gain setting components always affect stability. When the closed-loop gain is high, the loop gain is low, and low loop gain circuits are more stable than high loop gain circuits.

Wiring the op amp to a printed circuit board always introduces components formed from stray capacitance and inductance. Stray inductance becomes dominant at very high frequencies, hence, in VFAs, it does not interfere with stability as much as it does with signal handling properties. Stray capacitance causes stability to increase or decrease depending on its location. Stray capacitance from the input or output lead to ground induces instability, while the same stray capacitance in parallel with the feedback resistor increases stability.

The loop gain for a CFA with no input buffer output impedance, R_B, is given in Equation 9–6. Examining Equation 9–6 we see that the stability of a CFA depends on two items: the op amp transfer function, Z, and the gain setting component, Z_F. The op amp contains many poles, thus they require external compensation. Fortunately, the external compensation for a CFA is done with Z_F. The factory applications engineer does extensive testing to determine the optimum value of R_F for a given gain. This value should be used in all applications at that gain, but increased stability and less peaking can be obtained by increasing R_F. Essentially this is sacrificing bandwidth for lower frequency performance, but in applications not requiring the full bandwidth, it is a wise tradeoff.

The CFA stability is not constrained by the closed-loop gain, thus a stable operating point can be found for any gain, and the CFA is not limited by the gain-bandwidth constraint. If the optimum feedback resistor value is not given for a specific gain, one must test to find the optimum feedback resistor value.

Stray capacitance from any node to ground adversely affects the CFA performance. Stray capacitance of just a couple of pico Farads from any node to ground causes 3 dB or more of peaking in the frequency response. Stray capacitance across the CFA feedback resistor, quite unlike that across the VFA feedback resistor, always causes some form of instability. CFAs are applied at very high frequencies, so the printed circuit board inductance associated with the trace length and pins adds another variable to the stability equation. Inductance cancels out capacitance at some frequency, but this usually seems to happen in an adverse manner. The wiring of VFAs is critical, but the wiring of CFAs is a science. Stay with the layout recommended by the manufacturer whenever possible.

9.5 Impedance

The input impedance of a VFA and CFA differ dramatically because their circuit configurations are very different. The VFA input circuit is a long-tailed pair, and this configuration gives the advantages that both input impedances match. Also, the input signal looks into an emitter-follower circuit that has high input impedance. The emitter-follower input impedance is $\beta(r_e + R_E)$ where R_E is a discrete emitter resistor. At low input currents, R_E is very high and the input impedance is very high. If a higher input impedance is required, the op amp uses a Darlington circuit that has an input impedance of $\beta^2(r_e + R_E)$.

So far, the implicit assumption is that the VFA is made with a bipolar semiconductor process. Applications requiring very high input impedances often use a FET process. Both BIFET and CMOS processes offer very high input impedance in any long-tailed pair configuration. It is easy to get matched and high input impedances at the amplifier inputs. Do not confuse the matched input impedance at the op amp leads with the overall circuit input impedance. The input impedance looking into the inverting input is R_G, and the impedance looking into the noninverting input is the input impedance of the op amp. While these are two different impedances, they are mismatched because of the circuit not the op amp.

The CFA has a radically different input structure that causes it to have mismatched input impedances. The noninverting input lead of the CFA is the input of a buffer that has very high input impedance. The inverting input lead is the output of a buffer that has very low impedance. There is no possibility that these two input impedances can be matched.

Again, because of the circuit, the inverting circuit input impedance is R_G. Once the circuit gain is fixed, the only way to increase R_G is to increase R_F. But, R_F is determined by a tradeoff between stability and bandwidth. The circuit gain and bandwidth requirements fix R_F, hence there is no room to further adjust R_F to raise the resistance of R_G. If the manufacturer's data sheet says that $R_F = 100\ \Omega$ when the closed-loop gain is two, then

R_G = 100 Ω or 50 Ω depending on the circuit configuration. This sets the circuit input impedance at 100 Ω. This analysis is not entirely accurate because R_B adds to the input impedance, but this addition is very small and dependent on IC parameters. CFA op amp circuits are usually limited to noninverting voltage applications, but they serve very well in inverting applications that are current-driven.

The CFA is limited to the bipolar process because that process offers the highest speed. The option of changing process to BIFET or CMOS to gain increased input impedance is not attractive today. Although this seems like a limiting factor, it is not because CFAs are often used in low impedance where the inputs are terminated in 50 Ω or 75 Ω. Also, most very high-speed applications require low impedances.

9.6 Equation Comparison

The pertinent VFA and CFA equations are repeated in Table 9–1. Notice that the ideal closed-loop gain equations for the inverting and noninverting circuits are identical. The ideal equations for the VFA depend on the op amp gain, a, being very large thus making Aβ large compared to one. The CFA needs two assumptions to be valid to obtain the ideal equations. First, the ideal equations for the CFA depend on the op amp transimpedance, Z, being very large thus making Aβ large compared to one. Second, R_B must be very small compared to $Z_F||Z_G$.

The ideal gain equations are identical, but the applications are very different because the VFA is best applied to lower frequency precision jobs while the CFA applications are in the very high frequency realm. The transimpedance in a CFA acts much like the gain does in a VFA. In each case, transimpedance or gain, it is the parameter that enables the use of feedback.

Table 9–1. Tabulation of Pertinent VFA and CFA Equations

CIRCUIT CONFIGURATION	CURRENT FEEDBACK AMPLIFIER	VOLTAGE FEEDBACK AMPLIFIER
NONINVERTING		
Forward or direct gain	$\frac{Z\left(1 + Z_F/Z_G\right)}{Z_F\left(1 + Z_B/Z_F \parallel Z_G\right)}$	a
Actual closed loop gain	$\frac{\frac{Z_F\left(1+\frac{Z_B}{Z_G}\right)}{Z_F\left(1+\frac{Z_B}{Z_F \parallel Z_G}\right)}}{1 + \frac{Z}{Z_F\left[1+\frac{\frac{Z_B}{1+Z_B}}{Z_F \parallel Z_G}\right]}}$	$\frac{a}{1 + \frac{aZ_G}{Z_F \parallel Z_G}}$
Closed loop gain	$1 + Z_F/Z_G$	$1 + Z_F/Z_G$
INVERTING		
Forward or direct gain	$\frac{Z}{Z_G\left(1 + Z_B/Z_F \parallel Z_G\right)}$	$aZ_F/\left(Z_F + Z_G\right)$
Ideal loop gain	$Z/Z_F\left(1 + Z_B/Z_F \parallel Z_G\right)$	$aZ_G/\left(Z_G + Z_F\right)$
Actual closed loop gain	$\frac{-Z_G\left(1 + \frac{Z_B}{Z_F \parallel Z_G}\right)}{1 + \frac{Z}{Z_F\left(1+\frac{Z_B}{Z_F \parallel Z_G}\right)}}$	$\frac{\frac{-aZ_F}{Z_F+Z_G}}{1 + \frac{aZ_G}{Z_F \parallel Z_G}}$
Closed loop gain	$-Z_F/Z_G$	$-Z_F/Z_G$

Chapter 10

Op Amp Noise Theory and Applications

Bruce Carter

10.1 Introduction

The purpose of op amp circuitry is the manipulation of the input signal in some fashion. Unfortunately in the real world, the input signal has unwanted noise superimposed on it.

Noise is not something most designers get excited about. In fact, they probably wish the whole topic would go away. It can, however, be a fascinating study by itself. A good understanding of the underlying principles can, in some cases, be used to reduce noise in the design.

10.2 Characterization

Noise is a purely random signal, the instantaneous value and/or phase of the waveform cannot be predicted at any time. Noise can either be generated internally in the op amp, from its associated passive components, or superimposed on the circuit by external sources. External noise is covered in Chapter 17, and is usually the dominant effect.

10.2.1 rms versus P-P Noise

Instantaneous noise voltage amplitudes are as likely to be positive as negative. When plotted, they form a random pattern centered on zero. Since noise sources have amplitudes that vary randomly with time, they can only be specified by a probability density function. The most common probability density function is Gaussian. In a Gaussian probability function, there is a mean value of amplitude, which is most likely to occur. The probability that a noise amplitude will be higher or lower than the mean falls off in a bell-shaped curve, which is symmetrical around the center (Figure 10–1).

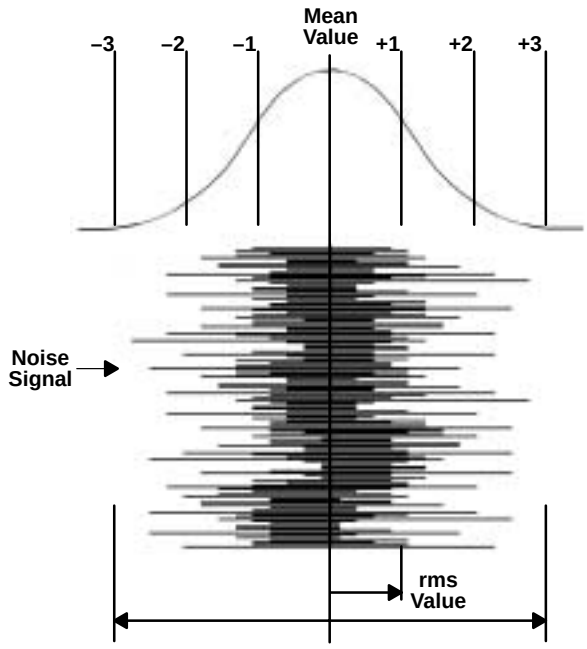


Figure 10–1. Gaussian Distribution of Noise Energy

σ is the standard deviation of the Gaussian distribution and the rms value of the noise voltage and current. The instantaneous noise amplitude is within $\pm 1\sigma$ 68% of the time. Theoretically, the instantaneous noise amplitude can have values approaching infinity. However, the probability falls off rapidly as amplitude increases. The instantaneous noise amplitude is within $\pm 3\sigma$ of the mean 99.7% of the time. If more or less assurance is desired, it is between $\pm 2\sigma$ 95.4% of the time and $\pm 3.4\sigma$ 99.94% of the time.

σ^2 is the average mean-square variation about the average value. This also means that the average mean-square variation about the average value, $\overline{i^2}$ or $\overline{e^2}$, is the same as the variance σ^2.

Thermal noise and shot noise (see below) have Gaussian probability density functions. The other forms of noise do not.

10.2.2 Noise Floor

When all input sources are turned off and the output is properly terminated, there is a level of noise called the *noise floor* that determines the smallest signal for which the circuit is useful. The objective for the designer is to place the signals that the circuit processes above the noise floor, but below the level where the signals will clip.

10.2.3 Signal-to-Noise Ratio

The noisiness of a signal is defined as:

$$\frac{S_{(f)}}{N_{(f)}} = \frac{\text{rms signal voltage}}{\text{rms noise voltage}} \quad (10\text{–}1)$$

In other words, it is a ratio of signal voltage to noise voltage (hence the name *signal-to-noise ratio*).

10.2.4 Multiple Noise Sources

When there are multiple noise sources in a circuit, the total root-mean-square (rms) noise signal that results is the square root of the sum of the average mean-square values of the individual sources:

$$E_{Totalrms} = \sqrt{e_{1rms}^2 + e_{2rms}^2 + \ldots e_{nrms}^2} \quad (10\text{–}2)$$

Put another way, this is the only “break” that the designer gets when dealing with noise. If there are two noise sources of equal amplitude in the circuit, the total noise is not doubled (increased by 6 dB). It only increases by 3 dB. Consider a very simple case, two noise sources with amplitudes of 2 V_{rms}:

$$E_{Totalrms} = \sqrt{2^2 + 2^2} = \sqrt{8} = 2.83\ V_{rms} \quad (10\text{–}3)$$

Therefore, when there are two equal sources of noise in a circuit, the noise is $20 \times \log\frac{2.83}{2} = 3.01$ dB higher than if there were only one source of noise — instead of double (6 dB) as would be intuitively expected.

This relationship means that the worst noise source in the system will tend to dominate the total noise. Consider a system in which one noise source is 10 V_{rms} and another is 1 V_{rms}:

$$E_{Totalrms} = \sqrt{10^2 + 1^2} = \sqrt{101} = 10.05\ V_{rms} \quad (10\text{–}4)$$

There is hardly any effect from the 1-V noise source at all!

10.2.5 Noise Units

Noise is normally specified as a spectral density in rms volts or amps per root Hertz, $V/\sqrt{Hz}$ or $A/\sqrt{Hz}$. These are not very “user-friendly” units. A frequency range is needed to relate these units to actual noise levels that will be observed.

For example:

- A TLE2027 op amp with a noise specification of 2.5 $nV/\sqrt{Hz}$ is used over an audio frequency range of 20 Hz to 20 kHz, with a gain of 40 dB. The output voltage is 0 dBV (1 V).
- To begin with, calculate the *root Hz* part: $\sqrt{20000 - 20} = 141.35$.
- Multiplying this by the noise spec: $2.5 \times 141.35 = 353.38$ nV, which is the equivalent input noise (EIN). The output noise equals the input noise multiplied by the gain, which is 100 (40 dB).

The signal-to-noise ratio can be now be calculated:

- $353.38\ nV \times 100 = 35.3\ \mu V$
- Signal-to-noise (dB) =

$$20 \times \log(1V \div 35.3\ \mu V) = 20 \times \log(28329) = 89\ dB \qquad (10–5)$$

The TLE2027 op amp is an excellent choice for this application. Remember, though, that passive components and external noise sources can degrade performance. There is also a slight increase in noise at low frequencies, due to the 1/f effect (see below).

10.3 Types of Noise

There are five types of noise in op amps and associated circuitry:

1) Shot noise
2) Thermal noise
3) Flicker noise
4) Burst noise
5) Avalanche noise

Some or all of these noises may be present in a design, presenting a noise spectrum unique to the system. It is not possible in most cases to separate the effects, but knowing general causes may help the designer optimize the design, minimizing noise in a particular bandwidth of interest. Proper design for low noise may involve a “balancing act” between these sources of noise and external noise sources.

10.3.1 Shot Noise

The name *shot noise* is short for Schottky noise. Sometimes it is referred to as *quantum noise*. It is caused by random fluctuations in the motion of charge carriers in a conductor. Put another way, current flow is not a continuous effect. Current flow is electrons, charged particles that move in accordance with an applied potential. When the electrons encounter a barrier, potential energy builds until they have enough energy to cross that barrier. When they have enough potential energy, it is abruptly transformed into kinetic energy as they cross the barrier. A good analogy is stress in an earthquake fault that is suddenly released as an earthquake.

As each electron randomly crosses a potential barrier, such as a pn junction in a semiconductor, energy is stored and released as the electron encounters and then shoots across the barrier. Each electron contributes a little *pop* as its stored energy is released when it crosses the barrier (Figure 10–2).

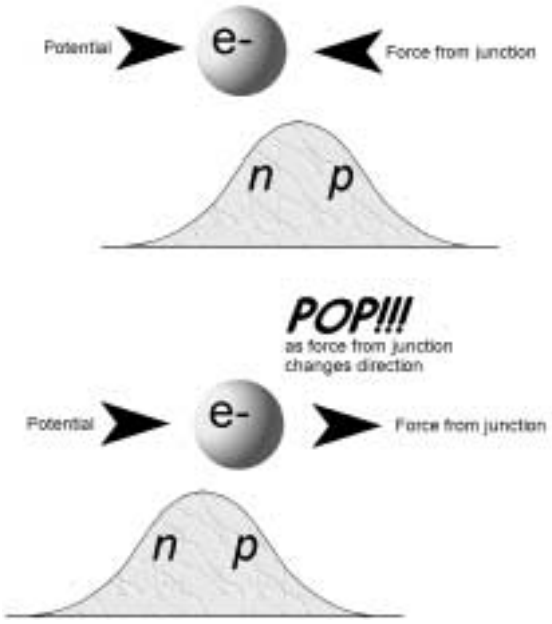


Figure 10–2. Shot Noise Generation

The aggregate effect of all of the electrons shooting across the barrier is the shot noise. Amplified shot noise has been described as sounding like lead shot hitting a concrete wall.

Some characteristics of shot noise:

- Shot noise is always associated with current flow. It stops when the current flow stops.
- Shot noise is independent of temperature.

- Shot noise is spectrally flat or has a uniform power density, meaning that when plotted versus frequency it has a constant value.
- Shot noise is present in any conductor — not just a semiconductor. Barriers in conductors can be as simple as imperfections or impurities in the metal. The level of shot noise, however, is very small due to the enormous numbers of electrons moving in the conductor, and the relative size of the potential barriers. Shot noise in semiconductors is much more pronounced.

The rms shot noise current is equal to:

$$I_{sh} = \sqrt{(2qI_{dc} + 4qI_0)B} \qquad (10–6)$$

Where:

q = Electron charge (1.6×10^{-19} coulombs)
I_{dc} = Average forward dc current in A
I_o = Reverse saturation current in A
B = Bandwidth in Hz

If the pn junction is forward biased, I_o is zero, and the second term disappears. Using Ohm's law and the dynamic resistance of a junction,

$$r_d = \frac{kT}{qI_{dc}} \qquad (10–7)$$

the rms shot noise voltage is equal to:

$$E_{sh} = kT\sqrt{\frac{2B}{qI_{dc}}} \qquad (10–8)$$

Where:

k = Boltzmann's constant (1.38×10^{-23} Joules/°K)
q = Electron charge (1.6×10^{-19} coulombs)
T = Temperature in °K
I_{dc} = Average dc current in A
B = Bandwidth in Hz

For example, a junction carries a current of 1 mA at room temperature. Its noise over the audio bandwidth is:

$$E_{sh} = 1.38 \times 10^{-23} \times 298 \sqrt{\frac{2(20000 - 20)}{(1.6 \times 10^{-19}) \times (1 \times 10^{-3})}} = 65 \text{ nV} = -144 \text{ dBV} \qquad (10–9)$$

Obviously, it is not much of a problem in this example.

Look closely at the formula for shot noise voltage. Notice that the shot noise voltage is inversely proportional to the current. Stated another way, shot noise voltage decreases

as average dc current increases, and increases as average dc current decreases. This can be an elegant way of determining if shot noise is a dominant effect in the op amp circuit being designed. If possible, decrease the average dc current by a factor of 100 and see if the overall noise increases by a factor of 10. In the example above:

$$E_{sh} = 1.38 \times 10^{-23} \times 298 \sqrt{\frac{2(20000 - 20)}{(1.6 \times 10^{-19}) \times (1 \times 10^{-5})}} = 650 \text{ nV} = -124 \text{ dBV} \qquad (10\text{–}10)$$

The shot noise voltage does increase by a factor of 10, or 20 dB.

10.3.2 Thermal Noise

Thermal noise is sometimes referred to as Johnson noise after its discoverer. It is generated by thermal agitation of electrons in a conductor. Simply put, as a conductor is heated, it will become noisy. Electrons are never at rest; they are always in motion. Heat disrupts the electrons' response to an applied potential. It adds a random component to their motion (Figure 10–3). Thermal noise only stops at absolute zero.

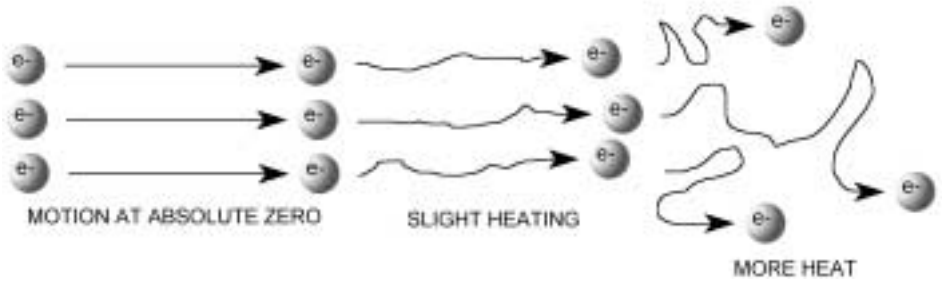


Figure 10–3. Thermal Noise

Like shot noise, thermal noise is spectrally flat or has a uniform power density (it is *white*), but thermal noise is independent of current flow.

At frequencies below 100 MHz, thermal noise can be calculated using Nyquist's relation:

$$E_{th} = \sqrt{4kTRB} \qquad (10\text{–}11)$$

or

$$I_{th} = \sqrt{\frac{4kTB}{R}} \qquad (10–12)$$

Where:

E_{th} =	Thermal noise voltage in Volts rms
I_{th} =	Thermal noise current in Amps rms
k =	Boltzmann's constant (1.38×10^{-23})
T =	Absolute temperature (Kelvin)
R =	Resistance in ohms
B =	Noise bandwidth in Hertz ($f_{max}–f_{min}$)

The noise from a resistor is proportional to its resistance and temperature. It is important not to operate resistors at elevated temperatures in high gain input stages. Lowering resistance values also reduces thermal noise.

For example:

The noise in a 100 kΩ resistor at 25°C (298°K) over the audio frequency range of 20 Hz to 20 kHz is:

$$\begin{aligned} E_{th} &= \sqrt{4kTRB} \\ &= \sqrt{4 \times (1.38 \times 10^{-23}) \times 298 \times 100,000 \times (20,000 - 20)} \\ &= 5.73\ \mu V \\ &= -104.8\ dBV \end{aligned} \qquad (10–13)$$

Decreasing the temperature would reduce the noise slightly, but scaling the resistor down to 1 kΩ (a factor of 100) would reduce the thermal noise by 20 dB. Similarly, increasing the resistor to 10 MΩ would increase the thermal noise to –84.8 dBV, a level that would affect a 16-bit audio circuit. The noise from multiple resistors adds according to the root-mean-square law in Paragraph 10.2.4. Beware of large resistors used as the input resistor of an op amp gain circuit, their thermal noise will be amplified by the gain in the circuit (Paragraph 10.4). Thermal noise in resistors is often a problem in portable equipment, where resistors have been scaled up to get power consumption down.

10.3.3 Flicker Noise

Flicker noise is also called *1/f noise*. Its origin is one of the oldest unsolved problems in physics. It is pervasive in nature and in many human endeavors. It is present in all active and many passive devices. It may be related to imperfections in crystalline structure of semiconductors, as better processing can reduce it.

Some characteristics of flicker noise:

- It increases as the frequency decreases, hence the name *1/f*
- It is associated with a dc current in electronic devices
- It has the same power content in each octave (or decade)

$$E_n = K_v \sqrt{\left(\ln \frac{f_{max}}{f_{min}}\right)} \qquad I_n = K_i \sqrt{\left(\ln \frac{f_{max}}{f_{min}}\right)} \tag{10–14}$$

Where:
K_e and K_i are proportionality constants (volts or amps) representing E_n and I_n at 1 Hz
f_{max} and f_{min} are the minimum and maximum frequencies in Hz

Flicker noise is found in carbon composition resistors, where it is often referred to as excess noise because it appears in addition to the thermal noise that is there. Other types of resistors also exhibit flicker noise to varying degrees, with wire wound showing the least. Since flicker noise is proportional to the dc current in the device, if the current is kept low enough, thermal noise will predominate and the type of resistor used will not change the noise in the circuit.

Reducing power consumption in an op amp circuit by scaling up resistors may reduce the 1/f noise, at the expense of increased thermal noise.

10.3.4 Burst Noise

Burst noise, also called popcorn noise, is related to imperfections in semiconductor material and heavy ion implants. It is characterized by discrete high-frequency pulses. The pulse rates may vary, but the amplitudes remain constant at several times the thermal noise amplitude. Burst noise makes a popping sound at rates below 100 Hz when played through a speaker — it sounds like popcorn popping, hence the name. Low burst noise is achieved by using clean device processing, and therefore is beyond the control of the designer. Modern processing techniques at Texas Instruments has all but eliminated its occurrence.

10.3.5 Avalanche Noise

Avalanche noise is created when a pn junction is operated in the reverse breakdown mode. Under the influence of a strong reverse electric field within the junction's depletion region, electrons have enough kinetic energy that, when they collide with the atoms of the crystal lattice, additional electron-hole pairs are formed (Figure 10–4). These collisions are purely random and produce random current pulses similar to shot noise, but much more intense.

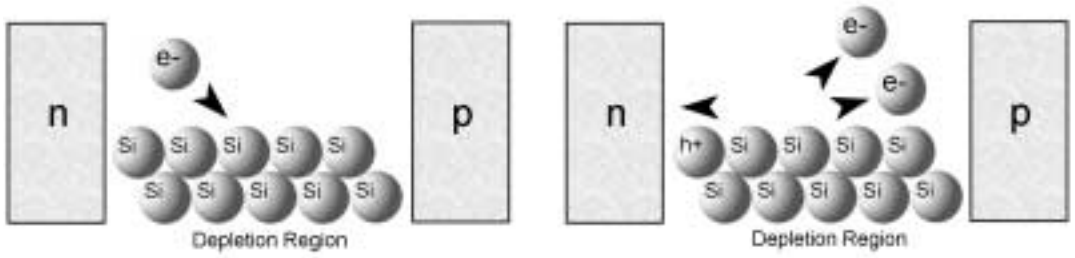


Figure 10–4. Avalanche Noise

When electrons and holes in the depletion region of a reversed-biased junction acquire enough energy to cause the avalanche effect, a random series of large noise spikes will be generated. The magnitude of the noise is difficult to predict due to its dependence on the materials.

Because the zener breakdown in a pn junction causes avalanche noise, it is an issue with op amp designs that include zener diodes. The best way of eliminating avalanche noise is to redesign a circuit to use no zener diodes.

10.4 Noise Colors

While the noise types are interesting, real op amp noise will appear as the summation of some or all of them. The various noise types themselves will be difficult to separate. Fortunately, there is an alternative way to describe noise, which is called *color*. The colors of noise come from rough analogies to light, and refer to the frequency content. Many colors are used to describe noise, some of them having a relationship to the real world, and some of them more attuned to the field of psycho-acoustics.

White noise is in the middle of a *spectrum* that runs from purple to blue to white to pink and red/brown. These colors correspond to powers of the frequency to which their spectrum is proportional, as shown in Table 10–1.

Table 10–1. Noise Colors

COLOR	FREQUENCY CONTENT
Purple	f^2
Blue	f
White	1
Pink	$\frac{1}{f}$
Red/Brown	$\frac{1}{f^2}$

There are an infinite number of variations between the colors. All inverse powers of frequency are possible, as are noises that are narrowband or appear only at one discrete frequency. Those, however, are primarily external sources of noise, so their presence is an important clue that the noise is external, not internal. There are no pure colors; at high frequencies, all of them begin to roll off and become pinkish. The op amp noise sources described above appear in the region between white noise and red/brown noise (Figure 10–5).

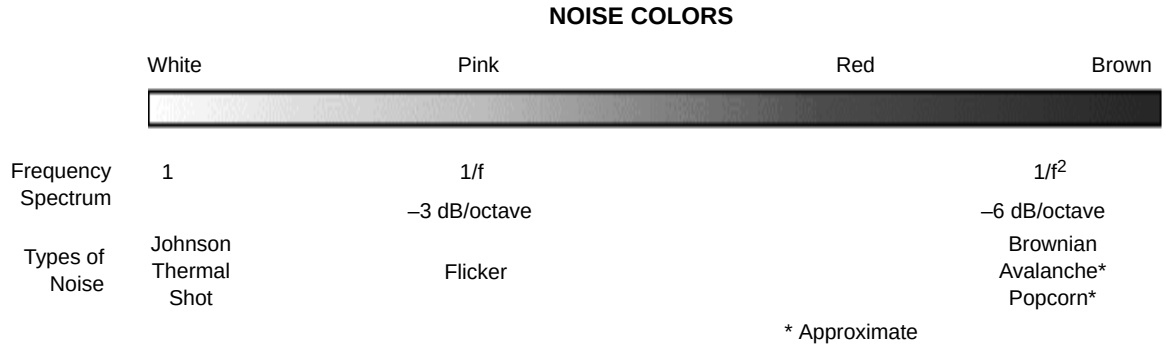


Figure 10–5. Noise Colors

10.4.1 White Noise

White noise is noise in which the frequency and power spectrum is constant and independent of frequency. The signal power for a constant bandwidth (centered at frequency f_o), does not change if f_o is varied. Its name comes from a similarity to white light, which has equal quantities of all colors.

When plotted versus frequency, white noise is a horizontal line of constant value.

Shot and thermal (Johnson) noise sources are approximately white, although there is no such thing as pure white noise. By definition, white noise would have infinite energy at infinite frequencies. White noise always becomes pinkish at high frequencies.

Steady rainfall or radio static on an unused channel approximate a white noise characteristic.

10.4.2 Pink Noise

Pink noise is noise with a 1/f frequency and power spectrum excluding dc. It has equal energy per octave (or decade for that matter). This means that the amplitude decreases logarithmically with frequency. Pink noise is pervasive in nature — many supposedly random events show a 1/f characteristic.

Flicker noise displays a 1/f characteristic, which also means that it rolls off at 3 dB/octave.

10.4.3 Red/Brown Noise

Red noise is not universally accepted as a noise type. Many sources omit it and go straight to brown, attributing red characteristics to brown. This has more to do with aesthetics than it does anything else (if brown noise is the low end of the spectrum, then pink noise should be named tan). So if pink noise is pink, then the low end of the spectrum should be red. Red noise is named for a connection with red light, which is on the low end of the visible light spectrum. But then this noise simulates Brownian motion, so perhaps it should be called Brown. Red/brown noise has a –6 dB/octave frequency response and a frequency spectrum of $1/f^2$ excluding dc.

Red/brown noise is found in nature. The acoustic characteristics of large bodies of water approximate red/brown noise frequency response.

Popcorn and avalanche noise approximate a red/brown characteristic, but they are more correctly defined as pink noise where the frequency characteristic has been shifted down as far as possible in frequency.

10.5 Op Amp Noise

This section describes the noise in op amps and associated circuits.

10.5.1 The Noise Corner Frequency and Total Noise

Op amp noise is never specified as shot, thermal, or flicker, or even white and pink. Noise for audio op amps is specified with a graph of equivalent input noise versus frequency. These graphs usually show two distinct regions:

- Lower frequencies where pink noise is the dominant effect
- Higher frequencies where white noise is the dominant effect

Actual measurements for the TLV2772 show that the noise has both white and pink characteristics (Figure 10–6). Therefore, the noise equations for each type of noise are not able to approximate the total noise out of the TLV2772 over the entire range shown on the graph. It is necessary to break the noise into two parts — the pink part and the white part — and then add those parts together to get the total op amp noise using the root-mean-square law of Paragraph 10.2.4.

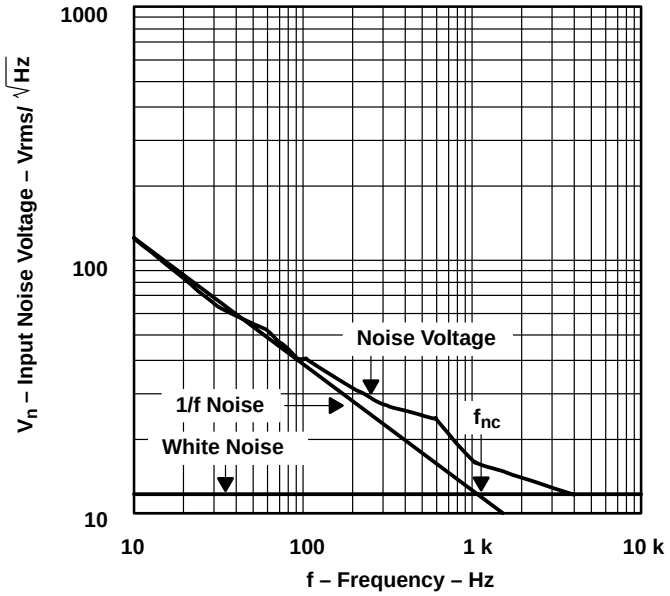


Figure 10–6. TLV2772 Op Amp Noise Characteristics

10.5.2 The Corner Frequency

The point in the frequency spectrum where 1/f noise and white noise are equal is referred to as the noise corner frequency, f_{nc}. Note on the graph in Figure 10–6 that the actual noise voltage is higher at f_{nc} due to the root-mean-square addition of noise sources as defined in Paragraph 10.2.4.

f_{nc} can be determined visually from the graph in Figure 10–6. It appears a little above 1 kHz. This was done by:

- Taking the white noise portion of the curve, and extrapolating it down to 10 Hz as a horizontal line.
- Taking the portion of the pink noise from 10 Hz to 100 Hz, and extrapolating it as a straight line.
- The point where the two intercept is $f_{nc,}$ the point where the white noise and pink noise are equal in amplitude. The total noise is then $\sqrt{2}$ x white noise specification (from Paragraph 10.2.4). This would be about 17 nV/ $\sqrt{Hz}$ for the TLV2772.

This is good enough for most applications. As can be seen from the actual noise plot in Figure 10–6, small fluctuations make precise calculation impossible. There is a precise method, however:

- Determine the 1/f noise at the lowest possible frequency.

- Square it.
- Subtract the white noise voltage squared (subtracting noise with root-mean-squares is just as valid as adding).
- Multiply by the frequency. This will give the noise contribution from the 1/f noise.
- Then divide by the white noise specification squared. The answer is f_{nc}.

For example:

The TLV2772 has a typical noise voltage of 130 nV/$\sqrt{\text{Hz}}$ at 10 Hz (from 5-V plot on data sheet).

The typical white noise specification for the TLV2772 is 12 nV/$\sqrt{\text{Hz}}$ (from data sheet)

$$1/f\ \text{noise}^2@10\text{Hz} = \left[\left(\frac{130\ \text{nV}}{\sqrt{\text{Hz}}}\right)^2 - \left(\frac{12\ \text{nV}}{\sqrt{\text{Hz}}}\right)^2\right] \times 10\ \text{Hz} = 167560(\text{nV})^2 \quad (10\text{–}15)$$

$$f_{nc} = \frac{1/f\ \text{noise}^2@10\text{Hz}}{\text{white noise}^2} = \frac{167560(\text{nV})^2}{\left(\frac{12\ \text{nV}}{\sqrt{\text{Hz}}}\right)^2} = 1164\ \text{Hz} \quad (10\text{–}16)$$

Once the corner frequency is known, the individual noise components can be added together as shown in Paragraph 10.2.2. Continuing the example above for a frequency range of 10 Hz to 10 kHz:

$$E_n = E_{\text{whitenoise}} \sqrt{f_{nc} \times 1n\frac{f_{max}}{f_{min}} + (f_{max} - f_{min})} \quad (10\text{–}17)$$

$$E_n = \frac{12\ \text{nV}}{\sqrt{\text{Hz}}} \sqrt{1164\ \text{Hz} \times 1n\frac{10^4}{10} + (10^4\ \text{Hz} - 10\ \text{Hz})} = 1.611\ \mu\text{V} = -116\ \text{dBV} \quad (10\text{–}18)$$

The example above presupposed that the bandwidth includes f_{nc}. If it does not, all of the contribution will be from either the 1/f noise or the white noise. Similarly, if the bandwidth is very large, and extends to three decades or so above f_{nc}, the contribution of the 1/f noise can be ignored.

10.5.3 Op Amp Circuit Noise Model

Texas Instruments measures the noise characteristics of a large sampling of devices. This information is compiled and used to determine the typical noise performance of the device. These noise specifications refer the input noise of the op amp. Some noise portions

can be represented better by a voltage source, and some by a current source. Input voltage noise is always represented by a voltage source in series with the noninverting input. Input current noise is always represented by current sources from both inputs to ground (Figure 10–7).

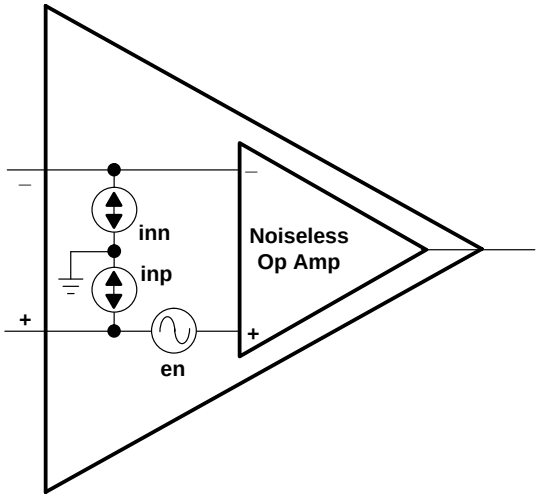


Figure 10–7. Op Amp Circuit Noise Model

In practice, op amp circuits are designed with low source impedance on the inverting and noninverting inputs. For low source impedances and CMOS JFET inputs, only the noise voltage is important; the current sources are insignificant in the calculations because they are swamped in the input impedances.

The equivalent circuit, therefore, reduces to that shown in Figure 10–8:

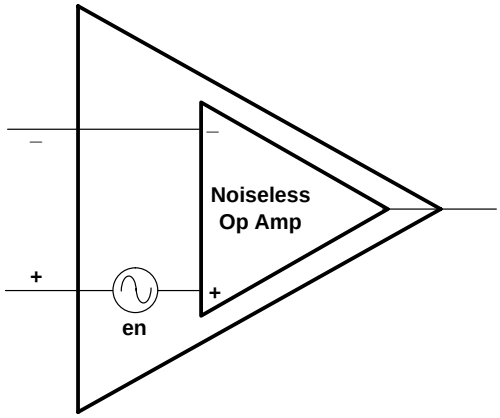


Figure 10–8. Equivalent Op Amp Circuit Noise Model

10.5.4 Inverting Op Amp Circuit Noise

If the previous circuit is operated in an inverting gain stage, the equivalent circuit becomes that shown in Figure 10–9:

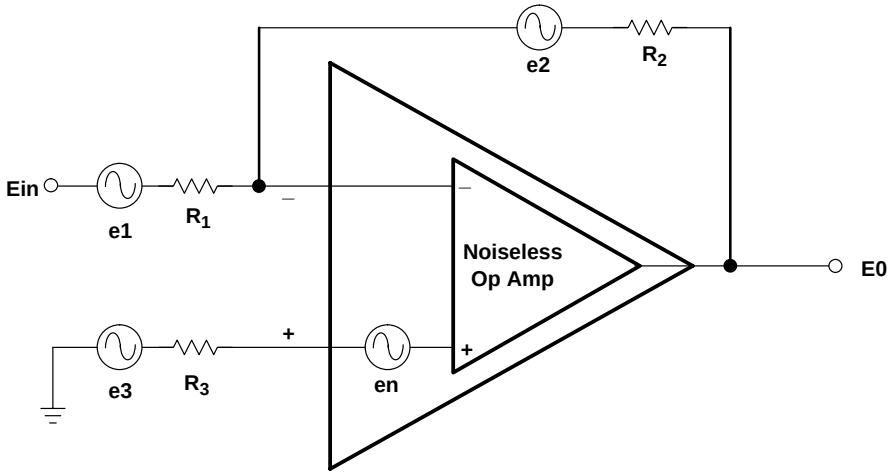


Figure 10–9. Inverting Op Amp Circuit Noise Model

The additional voltage sources e1 through e3 represent the thermal noise contribution from the resistors. As stated in Paragraph 10.3.2, the resistor noise can also be discounted if the values are low. Resistor noise will be omitted in the examples that follow. R3 is also not usually present, unless low common-mode performance is important. Deleting it and connecting the noninverting input directly to (virtual) ground makes the common mode response of the circuit worse, but may improve the noise performance of some circuits. There will be one less noise source to worry about. Therefore, the equivalent circuit becomes that shown in Figure 10–10:

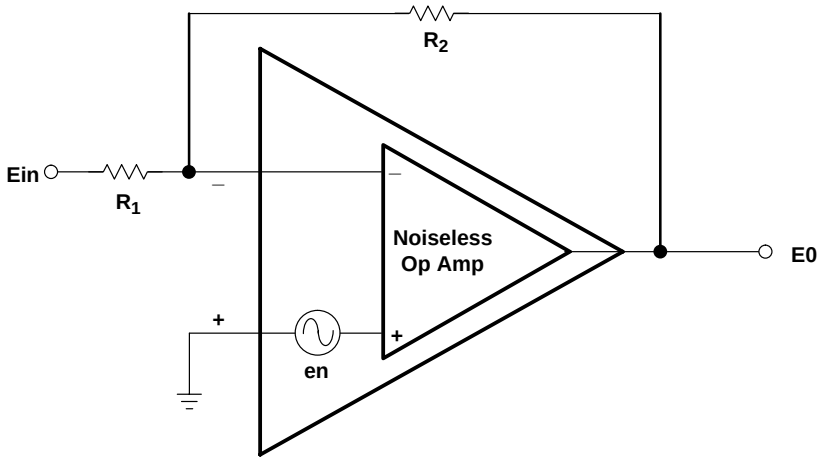


Figure 10–10. Inverting Equivalent Op Amp Circuit Noise Model

This simplifies the gain calculation:

$$E_0 = \sqrt{\left(E_{in}\frac{R_2}{R_1}\right)^2 + \left(e_n\left(1 + \frac{R_2}{R_1}\right)\right)^2} \qquad (10\text{–}19)$$

where e_n = the total noise over the bandwidth of interest.

10.5.5 Noninverting Op Amp Circuit Noise

Taking the simplified equivalent op amp circuit from Paragraph 10.5.2 as the base, the noise equivalent of a noninverting op amp circuit is shown in Figure 10–11:

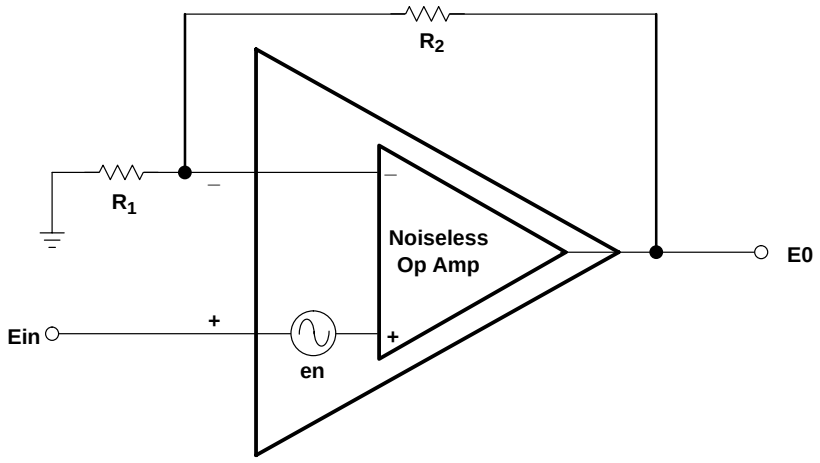


Figure 10–11. Noninverting Equivalent Op Amp Circuit Noise Model

The gain of this circuit is:

$$E_0 = \sqrt{\left(E_{in}\left(1 + \frac{R_2}{R_1}\right)\right)^2 + \left(e_n\left(1 + \frac{R_2}{R_1}\right)\right)^2} \qquad (10\text{–}20)$$

10.5.6 Differential Op Amp Circuit Noise

Taking the simplified equivalent op amp circuit from Paragraph 10.5.2 as the base, the noise equivalent of a differential op amp circuit is shown in Figure 10–12:

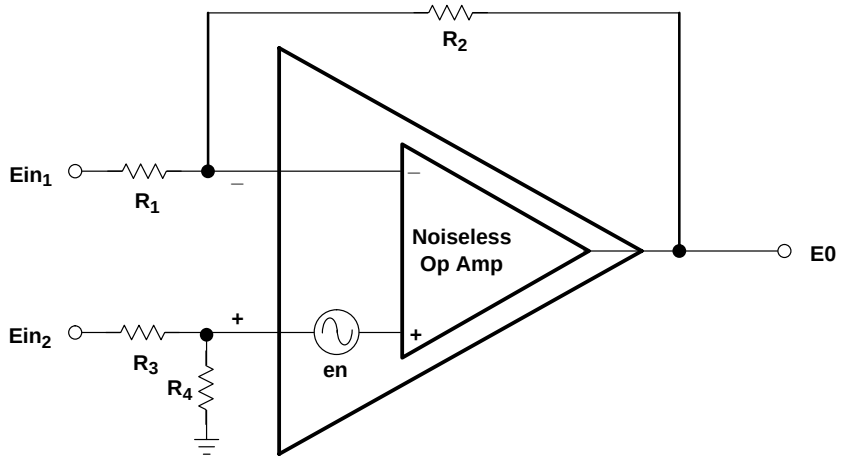


Figure 10–12. Differential Equivalent Op Amp Circuit Noise Model

Assuming that R1 = R3 and R2 = R4, the gain of this circuit is:

$$E_0 = \sqrt{\left((E_{in2} - E_{in1})\frac{R_2}{R_1}\right)^2 + \left(e_n\left(1 + \frac{R_2}{R_1}\right)\right)^2} \qquad (10\text{–}21)$$

10.5.7 Summary

The previous examples, though trivial, illustrate that noise always adds to the overall output of the op amp circuit. Reference 1 provides a much more in-depth derivation of op amp noise in circuits, including resistive effects.

10.6 Putting It All Together

This example is provided for analysis only — actual results depend on a number of other factors. Expanding on the techniques of Paragraph 10.2.5:

A low-noise op amp is needed over an audio frequency range of 20 Hz to 20 kHz, with a gain of 40 dB. The output voltage is 0 dBV (1V). The schematic is shown in Figure 10–13:

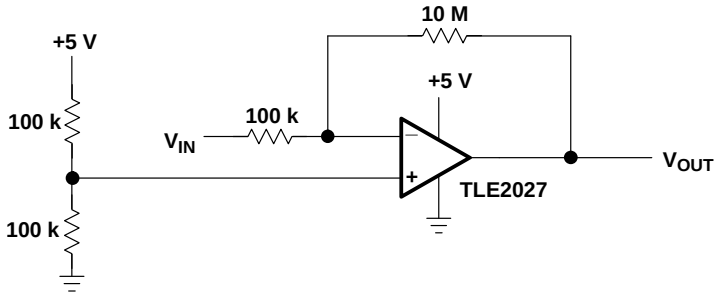


Figure 10–13. Split Supply Op Amp Circuit

It would be nice to use a TLE2027 with a noise figure of 2.5 $nV/\sqrt{Hz}$. The data sheet, however, reveals that this is a ±15-V part, and that noise figure is only specified at ±15 V. Furthermore, the specification for V_{OM+} and V_{OM-} (see Chapter 11) show that it can only swing to within approximately 2 V of its voltage rails. If they are +5 V and ground, the op amp is close to clipping with a 1-V output signal. This illustrates a common fallacy: the designer chooses an op amp based on one parameter only, without checking others that affect the circuit. An expert analog designer must develop an attention to details or be prepared to spend a lot of time in the lab with false starts and unexpected problems.

So, the only choice is to select a different op amp. The TLC2201 is an excellent choice. It is a low-noise op amp optimized for single supply operation. Figure 10–14 appears right on the first page of the data sheet, which should be extremely significant to the designer.

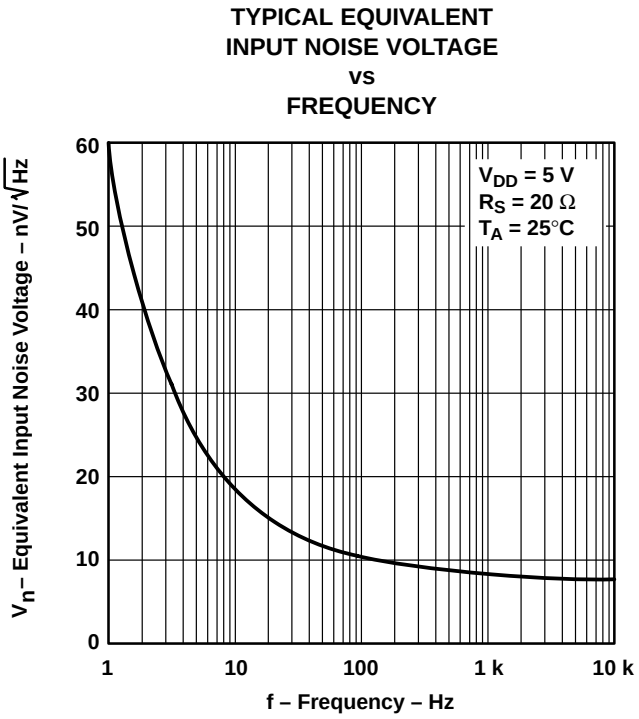


Figure 10–14. TLC2201 Op Amp Noise Performance

The first circuit change in this example is to change the TLE2027 to a TLC2201. Visually, the corner frequency f_{nc} appears to be somewhere around 20 Hz (from Paragraph 10.5.2), the lower frequency limit of the band we are interested in. This is good, it means for all practical purposes the 1/f noise can be discounted. It has 8 nV/$\sqrt{\text{Hz}}$ noise instead of 2.5 nV/$\sqrt{\text{Hz}}$, and from Paragraph 10.2.5:

- To begin with, calculate the *root Hz* part: $\sqrt{20000 - 20} = 141.35$.
- Multiplying this by the noise spec: $8 \times 141.35 = 1.131$ μV, which is the equivalent input noise (E_{IN}). The output noise equals the input noise multiplied by the gain, which is 100 (40 dB).

The signal-to-noise ratio can be now be calculated:

- $1.131\ \mu V \times 100 = 113.1\ \mu V$
- Signal-to-noise (dB) =

$$20 \times \log(1V \div 113.1\ \mu V) = 20 \times \log(8842) = 78.9\ \text{dB} \qquad (10\text{–}22)$$

Pretty good, but 10 dB less than would have been possible with a TLE2027. If this is not acceptable (lets say for 16-bit accuracy), one is forced to generate a ±15-V supply. Let's suppose for now that 78.9 dB signal-to-noise is acceptable, and build the circuit.

When it is assembled, it oscillates. What went wrong?

To begin with, it is important to look for potential sources of external noise. The way the schematic in Figure 10–14 is drawn provides a visual clue to the culprit: a long connection from the half-supply voltage reference to the high-impedance noninverting input. Added to that is a 50-kΩ source impedance, which does not effectively swamp external noise sources from entering the noninverting input. There is a big difference between simply providing a correct dc operating point, and providing one that has low impedances where they are needed. Most designers know the “fix”, which is to decouple the noninverting input as shown in Figure 10–15:

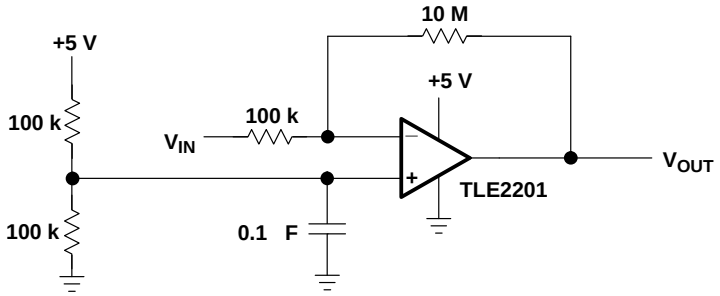


Figure 10–15. TLC2201 Op Amp Circuit

Better — it stopped oscillating. Probably a nearby noise source radiating into the noninverting input was providing enough noise to put the circuit into oscillation. The capacitor lowers the input impedance of the noninverting input and stops the oscillation. There is much more information on this topic in Chapter 17, including layout effects and component selection. For now, it is assumed that all of these have been taken into account.

The circuit is still slightly noisier than the 78.9 dB signal-to-noise ratio given above, especially at lower frequencies. This is where the real work of this example begins: that of eliminating component noise.

The circuit in Figure 10–15 has 4 resistors. Assuming that the capacitor is noiseless (not always a good assumption), that means four noise sources. For now, only the two resistors in the voltage divider that forms the voltage reference will be considered. The capacitor, however, has transformed the white noise from the resistors into pink (1/f) noise. From Paragraph 10.3.2 and 10.2.5, the noise from the resistors and the amplifier itself is:

$$E_{Totalrms} = \sqrt{5.73\ \mu V^2 + 5.73\ \mu V^2 + 113.1\ \mu V^2} = 113.1\ \mu V_{rms} \qquad (10–23)$$

Signal-to-noise (dB) =

$$20 \times \log(1V \div 113.1\ \mu V) = 20 \times \log(8842) = 78.9\ dB \qquad (10–24)$$

So far, so good. The amplifier noise is swamping the resistor noise, which will only add a very slight pinkish component at low frequencies. Remember, however, that this noise

voltage is multiplied by 101 through the circuit, but that was previously taken into account for the 78.9 dB signal-to-noise calculation above.

Reducing the value of the resistors to decrease their noise is an option. Changing the voltage divider resistors from 100 kΩ to 1 kΩ while leaving the 0.1 μV capacitor the same, changes the corner frequency from 32 Hz to 796 Hz, right in the middle of the audio band.

> **Note:**
>
> Resist the temptation to make the capacitor larger to move the pinkish effect below the lower limits of human hearing. The resulting circuit must charge the large capacitor up during power up, and down during power down. This may cause unexpected results.

If the noise from the half-supply generator is critical, the best possible solution is to use a low-noise, low-impedance half-supply source. Remember, however, that its noise will be multiplied by 101 in this application.

The effect of the 100-kΩ resistor on the inverting input is whitish, and will appear across the entire bandwidth of the circuit. Compared to the amplifier noise, it is still small, just like the noise from the noninverting resistors on the input. The noise contribution of resistors will be discounted.

Of much more concern, however, is the 10-MΩ resistor used as the feedback resistor. The noise associated with it appears as a voltage source at the inverting input of the op amp, and, therefore, is multiplied by a factor of 100 through the circuit. From Paragraph 10.3.2, the noise of a 10-MΩ resistor is –84.8 dBV, or 57.3 μV. Adding this and the 100-kΩ resistor noise to the amplifier noise:

$$E_{Totalrms} = \sqrt{5.73\ \mu V^2 + 113.1\ \mu V^2} = 126.8\ \mu V_{rms} = -77.9\ dBV \qquad (10–25)$$

Signal-to-noise (dB) =

$$20 \times \log(1V \div 126.8\ \mu V) = 20 \times \log(7887) = 77.9\ dB \qquad (10–26)$$

The noise contribution from the 10-MΩ resistor subtracts 1 dB from the signal-to-noise ratio. Changing the 10-MΩ resistor to 100 kΩ, and the input resistor from 100 kΩ to 1 kΩ preserves the overall gain of the circuit. The redesigned circuit is shown in Figure 10–16:

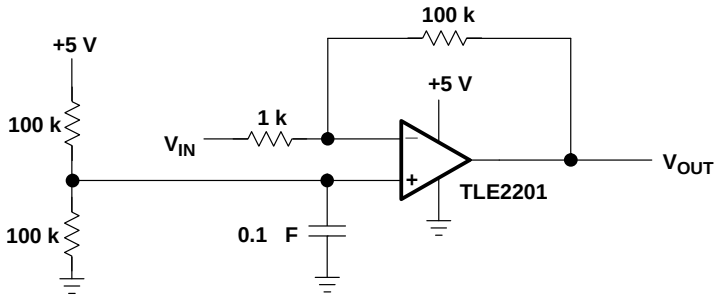


Figure 10–16. Improved TLC2201 Op Amp Circuit

For frequencies above 100 Hz, where the 1/f noise from the op amp and the reference resistors is negligible, the total noise of the circuit is:

$$E_{Totalrms} = \sqrt{0.57\ \mu V^2 + 5.73\ \mu V^2 + 113.1\ \mu V^2} = 113.2\ \mu V_{rms} = -78.9\ dBV \qquad (10–27)$$

Signal-to-noise (dB) =

$$20 \times \log(1V \div 113.2\ \mu V) = 20 \times \log(8830) = 78.9\ dB \qquad (10–28)$$

Proper selection of resistors, therefore, has yielded a signal-to-noise ratio close to the theoretical limit for the op amp itself. The power consumption of the circuit, however, has increased slightly, which may be unacceptable in a portable application. Remember, too, that this signal-to-noise ratio is only at an output level of 0 dBV, an input level of –40 dBV. If the input signal is reduced, the signal-to-noise ratio is reduced proportionally.

Music, in particular almost never sustains peak levels. The average amplitude may be down 20 dB to 40 dB from the peak values. This erodes a 79 dB signal-to-noise ratio to 39 dB in quiet passages. If someone “cranks up the volume” during the quiet passages, noise will become audible. This is done automatically with automatic volume controls. The only way a designer can combat this is to increase the voltage levels through the individual stages. If the preceding audio stages connecting to this example, for instance, could be scaled to provide 10 dB more gain, the TLC2201 would be handling an output level of 3.16 V instead of 1 V, which is well within its rail-to-rail limit of 0 V to 4.7 V. This would increase the signal-to-noise gain of this circuit to 88.9 dB — almost the same as would have been possible with a TLE2027 operated off of ±15V! But noise in the preceding stages would also increase. Combatting noise is a difficult problem, and there are always trade-offs involved.

10.7 References

(1) Texas Instruments Application Report, *Noise Analysis in Operational Amplifier Circuits*, SLVA043A, 1999

Chapter 11

Understanding Op Amp Parameters

Bruce Carter

11.1 Introduction

This chapter is about op amp data sheet parameters. The designer must have a clear understanding of what op amp parameters mean and their impact on circuit design. The chapter is arranged for speedy access to parameter information. Their definitions, typical abbreviations, and units appear in Section 11.2. Section 11.3 digs deeper into important parameters for the designer needing more in-depth information.

While these parameters are the ones most commonly used at Texas Instruments, the same parameter may go by different names and abbreviations at other manufacturers. Not every parameter listed here may appear in the data sheet for a given op amp. An op amp that is intended only for ac applications may omit dc offset information. The omission of information is not an attempt to “hide” anything. It is merely an attempt to highlight the parameters of most interest to the designer who is using the part the way it was intended. There is no such thing as an ideal op amp — or one that is universally applicable. The selection of any op amp must be based on an understanding of what particular parameters are most important to the application.

If a particular parameter cannot be found in the data sheet, a review of the application may well be in order and another part, whose data sheet contains the pertinent information, might be more suitable. Texas Instruments manufactures a broad line of op amps that can implement almost any application. The inexperienced designer could easily select an op amp that is totally wrong for the application. Trying to use an audio op amp with low total harmonic distortion in a high-speed video circuit, for example, will not work — no matter how superlative the audio performance might be.

Some parameters have a statistically normal distribution. The typical value published in the data sheet is the mean or average value of the distribution. The typical value listed is the 1σ value. This means that in 68% of the devices tested, the parameter is found to be $\pm$ the typical value or better. Texas Instruments currently uses 6σ to define minimum and maximum values. Usually, typical values are set when the part is characterized and never changes.

11.2 Operational Amplifier Parameter Glossary

There are usually three main sections of electrical tables in op amp data sheets. The *absolute maximum ratings* table and the *recommended operating conditions* table list constraints placed upon the circuit in which the part will be installed. *Electrical characteristics* tables detail device performance.

Absolute maximum ratings are those limits beyond which the life of individual devices may be impaired and are never to be exceeded in service or testing. Limits, by definition, are maximum ratings, so if double-ended limits are specified, the term will be defined as a range (e.g., operating temperature range).

Recommended operating conditions have a similarity to maximum ratings in that operation outside the stated limits could cause unsatisfactory performance. Recommended operating conditions, however, do not carry the implication of device damage if they are exceeded.

Electrical characteristics are measurable electrical properties of a device inherent in its design. They are used to predict the performance of the device as an element of an electrical circuit. The measurements that appear in the electrical characteristics tables are based on the device being operated within the recommended operating conditions.

Table 11–1 is a list of parameters and operating conditions that are commonly used in TI op amp data sheets. The glossary is arranged alphabetically by parameter name. An abbreviation cross-reference is provided after the glossary in Table 11–2 to help the designer find information when only an abbreviation is given. More detail is given about important parameters in Section 11.3.

Table 11–1. Op Amp Parameter GLossary

PARAMETER	ABBV	UNITS	DEFINITION	INFO
Bandwidth for 0.1 dB flatness		MHz	The range of frequencies within which the gain is ± 0.1 dB of the nominal value.	
Case temperature for 60 seconds		°C	Usually specified as an absolute maximum — It is meant to be used as guide for automated soldering processes.	
Common-mode input capacitance	C_{ic}	pF	Input capacitance a common-mode source would see to ground.	11.3.7.1
Common-mode input impedance	Z_{ic}	Ω	The parallel sum of the small-signal impedance between each input terminal and ground.	
Common-mode input voltage	V_{IC}	V	The average voltage at the input pins.	11.3.3
Common-mode rejection ratio	CMRR or k_{CMR}	dB	The ratio of differential voltage amplification to common-mode voltage amplification. Note: This is measured by determining the ratio of a change in input common-mode voltage to the resulting change in input offset voltage.	11.3.9

PARAMETER	ABBV	UNITS	DEFINITION	INFO
Continuous total dissipation		mW	Usually specified as an absolute maximum. It is the power that can be dissipated by the op amp package, including the load power. This parameter may be broken down by ambient temperature and package style in a table.	
Crosstalk	X_T	dBc	The ratio of the change in output voltage of a driven channel to the resulting change in output voltage from another channel that is not driven.	
Differential gain error	A_D	%	The change in ac gain with change in dc level. The ac signal is 40 IRE (0.28 V_{PK}) and the dc level change is ±100 IRE (±0.7 V). Typically tested at 3.58 MHz (NTSC) or 4.43 MHz (PAL) carrier frequencies.	
Differential input capacitance	C_{ic}	pF	(see common mode input capacitance)	11.3.7.1
Differential input resistance	r_{id}	Ω	The small-signal resistance between two ungrounded input terminals.	
Differential input voltage	V_{ID}	V	The voltage at the noninverting input with respect to the inverting input.	
Differential phase error	Φ_D	°	The change in ac phase with change in dc level. The ac signal is 40 IRE (0.28 V_{PK}) and the dc level change is ±100 IRE (±0.7 V). Typically tested at 3.58 MHz (NTSC) or 4.43 MHz (PAL) carrier frequencies.	
Differential voltage amplification	A_{VD}	dB	(see open loop voltage gain)	11.3.6
Fall time	t_f	ns	The time required for an output voltage step to change from 90% to 10% of its final value.	
Duration of short-circuit current			Amount of time that the output can be shorted to network ground — usually specified as an absolute maximum.	
Input common-mode voltage range	V_{ICR}	V	The range of common-mode input voltage that, if exceeded, may cause the operational amplifier to cease functioning properly. This is sometimes is taken as the voltage range over which the input offset voltage remains within a set limit.	11.3.3
Input current	I_I	mA	The amount of current that can be sourced or sinked by the op amp input — usually specified as an absolute maximum rating.	
Input noise current	I_n	$\frac{pA}{\sqrt{Hz}}$	The internal noise current reflected back to an ideal current source in parallel with the input pins.	11.3.13
Input noise voltage	V_n	$\frac{nV}{\sqrt{Hz}}$	The internal noise voltage reflected back to an ideal voltage source in parallel with the input pins.	11.3.13
Gain bandwidth product	GBW	MHz	The product of the open-loop voltage gain and the frequency at which it is measured.	11.3.13
Gain margin	A_m	dB	The reciprocal of the open-loop voltage gain at the frequency where the open-loop phase shift first reaches –180°.	
High-level output voltage	V_{OH}	V	The highest positive op amp output voltage for the bias conditions applied to the power pins.	11.3.5
Input bias current	I_{IB}	μA	The average of the currents into the two input terminals with the output at a specified level.	11.3.2

PARAMETER	ABBV	UNITS	DEFINITION	INFO
Input capacitance	c_i	pF	The capacitance between the input terminals with either input grounded.	11.3.7.1
Input offset current	I_{IO}	µA	The difference between the currents into the two input terminals with the output at the specified level.	11.3.2
Input offset voltage	$V_{IO,}$ V_{OS}	mV	The dc voltage that must be applied between the input terminals to cancel dc offsets within the op amp.	11.3.1
Input offset voltage long-term drift		$\frac{\mu V}{month}$	The ratio of the change in input offset voltage to the change time. It is the average value for the month.	11.3.1
Input resistance	r_i	MΩ	The dc resistance between the input terminals with either input grounded.	11.3.7.1
Input voltage range	V_I	V	The range of input voltages that may be applied to either the IN+ or IN– inputs	11.3.15
Large-signal voltage amplification	A_V	dB	(see open loop voltage gain)	
Lead temperature for 10 or 60 seconds		°C	Usually specified as an absolute maximum. It is meant to be used as guide for automated and hand soldering processes.	
Low-level output current	I_{OL}	mA	The current into an output with input conditions applied that according to the product parameter will establish a low level at the output.	
Low-level output voltage	V_{OL}	V	The smallest positive op amp output voltage for the bias conditions applied to the power pins.	11.3.5
Maximum peak output voltage swing	$V_{OM\pm}$	V	The maximum peak-to-peak output voltage that can be obtained without clipping when the op amp is operated from a bipolar supply.	11.3.5
Maximum peak-to-peak output voltage swing	$V_{O(PP)}$	V	The maximum peak-to-peak voltage that can be obtained without waveform clipping when the dc output voltage is zero.	
Maximum-output-swing bandwidth	B_{OM}	MHz	The range of frequencies within which the maximum output voltage swing is above a specified value or the maximum frequency of an amplifier in which the output amplitude is at the extents of it's linear range. Also called full power bandwidth.	11.3.15
Noise figure	NF	dB	The ratio of the total noise power at the output of an amplifier, referred to the input, to the noise power of the signal source.	
Open-loop transimpedance	Z_t	MΩ	In a transimpedance or current feedback amplifier, it is the frequency dependent ratio of change in output voltage to the frequency dependent change in current at the inverting input.	
Open-loop transresistance	R_t	MΩ	In a transimpedance or current feedback amplifier, it is the ratio of change in dc output voltage to the change in dc current at the inverting input.	
Open -loop voltage gain	A_{OL}	dB	The ratio of change in output voltage to the change in voltage across the input terminals. Usually the dc value and a graph showing the frequency dependence are shown in the data sheet.	
Operating temperature	T_A	°C	Temperature over which the op amp may be operated. Some of the other parameters may change with temperature, leading to degraded operation at temperature extremes.	

PARAMETER	ABBV	UNITS	DEFINITION	INFO
Output current	I_O	mA	The amount of current that is drawn from the op amp output. Usually specified as an absolute maximum rating — not for long term operation at the specified level.	
Output impedance	Z_o	Ω	The frequency dependent small-signal impedance that is placed in series with an ideal amplifier and the output terminal.	11.3.8
Output resistance	r_o	Ω	The dc resistance that is placed in series with an ideal amplifier and the output terminal.	
Overshoot factor	–	–	The ratio of the largest deviation of the output voltage from its final steady-state value to the absolute value of the step after a step change at the output.	
Phase margin	Φ_m	°	The absolute value of the open-loop phase shift at the frequency where the open-loop amplification first equals one.	11.3.15
Power supply rejection ratio	PSRR	dB	The absolute value of the ratio of the change in supply voltages to the change in input offset voltage. Typically both supply voltages are varied symmetrically. Unless otherwise noted, both supply voltages are varied symmetrically.	11.3.10
Rise time	t_r	nS	The time required for an output voltage step to change from 10% to 90% of its final value.	
Settling time	t_s	nS	With a step change at the input, the time required for the output voltage to settle within the specified error band of the final value. Also known as total response time, $t_{tot.}$	
Short-circuit output current	I_{OS}	mA	The maximum continuous output current available from the amplifier with the output shorted to ground, to either supply, or to a specified point. Sometimes a low value series resistor is specified.	
Slew rate	SR	V/µs	The rate of change in the output voltage with respect to time for a step change at the input.	11.3.12
Storage temperature	T_S	°C	Temperature over which the op amp may be stored for long periods of time without damage.	
Supply current	I_{CC}/I_{DD}	mA	The current into the V_{CC+}/V_{DD+} or $V_{CC–}/V_{DD–}$ terminal of the op amp while it is operating.	
Supply current (shutdown)	$I_{CC–}/I_{DD–}$ SHDN	mA	The current into the V_{CC+}/V_{DD+} or $V_{CC–}/V_{DD–}$ terminal of the amplifier while it is turned off.	
Supply rejection ratio	k_{SVR}	dB	(see power supply rejection ratio)	11.3.10
Supply voltage sensitivity	k_{SVS}, $\Delta V_{CC±}$, $\Delta V_{DD±}$, or ΔV_{IO}	dB	The absolute value of the ratio of the change in input offset voltage to the change in supply voltages.	11.3.10
Supply voltage	V_{CC}/V_{DD}	V	Bias voltage applied to the op amp power supply pin(s). Usually specified as a ± value, referenced to network ground.	
Temperature coefficient of input offset current	αI_{IO}	µA/°C	The ratio of the change in input offset current to the change in free-air temperature. This is an average value for the specified temperature range.	11.3.2

PARAMETER	ABBV	UNITS	DEFINITION	INFO
Temperature coefficient of input offset voltage	αV_{IO}	µV/°C	The ratio of the change in input offset voltage to the change in free-air temperature. This is an average value for the specified temperature range.	11.3.1
Total current into $V_{CC}+$/ V_{DD+}		mA	Maximum current that can be supplied to the positive power terminal of the op amp — usually specified as an absolute maximum.	
Total current out of V_{DD-}		mA	Maximum current that can be drawn from the negative power terminal of the op amp — usually specified as an absolute maximum.	
Total harmonic distortion	THD	dB	The ratio of the RMS voltage of the first nine harmonics of the fundamental signal to the total RMS voltage at the output.	
Total harmonic distortion plus noise	THD+N	dB	The ratio of the RMS noise voltage and RMS harmonic voltage of the fundamental signal to the total RMS voltage at the output.	11.3.14
Total power dissipation	P_D	mW	The total dc power supplied to the device less any power delivered from the device to a load. Note: At no load: $P_D = V_{CC+} \times I$ or $P_D = V_{DD+} \times I$	
Turn-on voltage (shutdown)	$V_{IH-SHDN}$	V	The voltage required on the shutdown pin to turn the device on.	
Turn-off voltage (shutdown)	$V_{IL-SHDN}$	V	The voltage required on the shutdown pin to turn the device off.	
Turn-on time (shutdown)	t_{EN}	µs	The time from when the turn-on voltage is applied to the shutdown pin to when the supply current has reached half of its final value.	
Turn-off time (shutdown)	t_{DIS}	µs	The time from when the turn-off voltage is applied to the shutdown pin to when the supply current has reached half of its final value.	
Unity gain bandwidth	B_1	MHz	The range of frequencies within which the open-loop voltage amplification is greater that unity.	11.3.15

Table 11–2. Cross-Reference of Op Amp Parameters

ABBV	PARAMETER
αI_{IO}	Temperature coefficient of input offset current
$\Delta V_{CC\pm}$/ $\Delta V_{DD\pm}$	Supply voltage sensitivity
αV_{IO}	Temperature coefficient of input offset voltage
ΔV_{IO}	Supply voltage sensitivity
A_D	Differential gain error
A_m	Gain margin
A_{OL}	Open loop voltage gain
A_V	Large-signal voltage amplification
A_{VD}	Differential voltage amplification

ABBV	PARAMETER
B_1	Unity gain bandwidth
B_{OM}	Maximum-output-swing bandwidth
c_i	Input capacitance
C_{ic}	Common-mode input capacitance
CMRR	Common-mode rejection ratio
GBW	Gain bandwidth product
$I_{CC–SHDN}$/ $I_{DD–SHDN}$	Supply current (shutdown)
I_{CC}/I_{DD}	Supply current
I_I	Input current
I_{IB}	Input bias current
I_{IO}	Input offset current
I_n	Input noise current
I_O	Output current
I_{OL}	Low-level output current
I_{OS}	Short-circuit output current
k_{CMR}	Common-mode rejection ratio
k_{SVR}	Supply rejection ratio
k_{SVS}	Supply voltage sensitivity
NF	Noise figure
P_D	Total power dissipation
PSRR	Power supply rejection ratio
r_i	Input resistance
R_{id}	Differential input resistance
R_o	Output resistance
R_t	Open-loop transresistance
SR	Slew rate
T_A	Operating temperature
t_{DIS}	Turn-off time (shutdown)
t_{EN}	Turn-on time (shutdown)
THD	Total harmonic distortion
t_f	Fall time
THD+N	Total harmonic distortion plus noise
t_r	Rise time
t_s	Settling time
T_S	Storage temperature
V_{CC}/V_{DD}	Supply voltage
V_I	Input voltage range
V_{IC}	Common-mode input voltage
V_{ICR}	Input common-mode voltage range

ABBV	PARAMETER
V_{ID}	Differential input voltage
V_{IHSHDN}	Turn-on voltage (shutdown)
$V_{IL–SHDN}$	Turn-off voltage (shutdown)
V_{IO}, V_{OS}	Input offset voltage
V_n	Input noise voltage
$V_{O(PP)}$	Maximum peak-to-peak output voltage swing
V_{OH}	High-level output voltage
V_{OL}	Low-level output voltage
$V_{OM±}$	Maximum peak output voltage swing
X_T	Crosstalk
Z_{ic}	Common-mode input impedance
Z_o	Output impedance
Z_t	Open-loop transimpedance
Φ_D	Differential phase error
Φ_m	Phase margin

11.3 Additional Parameter Information

Depending on the application, some op amp parameters are more important than others. This section contains additional information for parameters that impact a broad range of designs.

11.3.1 Input Offset Voltage

All op amps require a small voltage between their inverting and noninverting inputs to balance mismatches due to unavoidable process variations. The required voltage is known as the input offset voltage and is abbreviated V_{IO}. V_{IO} is normally modeled as a voltage source driving the noninverting input.

Figure 11–1 shows two typical methods for measuring input offset voltage — *DUT* stands for device under test. Test circuit (a) is simple, but since Vout is not at zero volts, it does not really meet the definition of the parameter. Test circuit (b) is referred to as a *servo loop*. The action of the loop is to maintain the output of the DUT at zero volts.

Bipolar input op amps typically offer better offset parameters than JFET or CMOS input op amps.

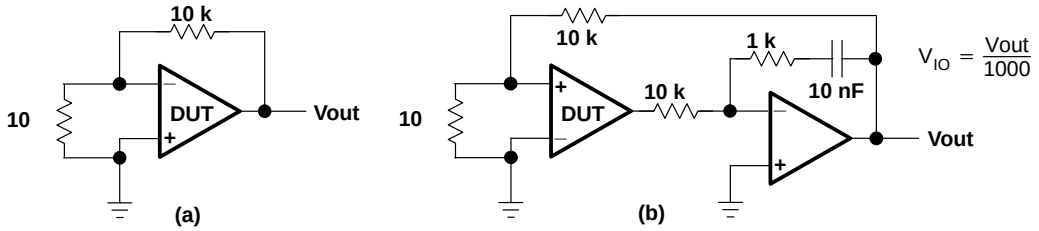


Figure 11–1. Test Circuits for Input Offset Voltage

TI data sheets show two other parameters related to V_{IO}; the average temperature coefficient of input offset voltage, and the input offset voltage long-term drift.

The average temperature coefficient of input offset voltage, αV_{IO}, specifies the expected input offset drift over temperature. Its units are $\mu V/^{\circ}C$. V_{IO} is measured at the temperature extremes of the part, and αV_{IO} is computed as $\Delta V_{IO}/\Delta^{\circ}C$.

Normal aging in semiconductors causes changes in the characteristics of devices. The input offset voltage long-term drift specifies how V_{IO} is expected to change with time. Its units are μV/month.

V_{IO} is normally attributed to the input differential pair in a voltage feedback amplifier. Different processes provide certain advantages. Bipolar input stages tend to have lower offset voltages than CMOS or JFET input stages.

Input offset voltage is of concern anytime that DC accuracy is required of the circuit. One way to null the offset is to use external null inputs on a single op amp package (Figure 11–2). A potentiometer is connected between the null inputs with the adjustable terminal connected to the negative supply through a series resistor. The input offset voltage is nulled by shorting the inputs and adjusting the potentiometer until the output is zero.

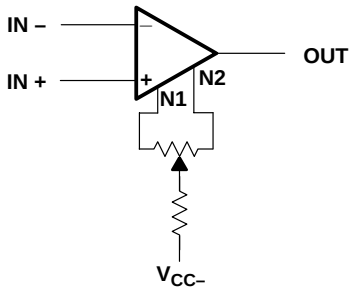


Figure 11–2. Offset Voltage Adjust

11.3.2 Input Current

The input circuitry of all op amps requires a certain amount of bias current for proper operation. The input bias current, I_{IB}, is computed as the average of the two inputs:

$$I_{IB} = \frac{\left(I_N + I_P\right)}{2} \qquad (11\text{–}1)$$

CMOS and JFET inputs offer much lower input current than standard bipolar inputs. Figure 11–3 shows a typical test circuit for measuring input bias currents.

The difference between the bias currents at the inverting and noninverting inputs is called the input offset current, $I_{IO} = I_N – I_P$. Offset current is typically an order of magnitude less than bias current.

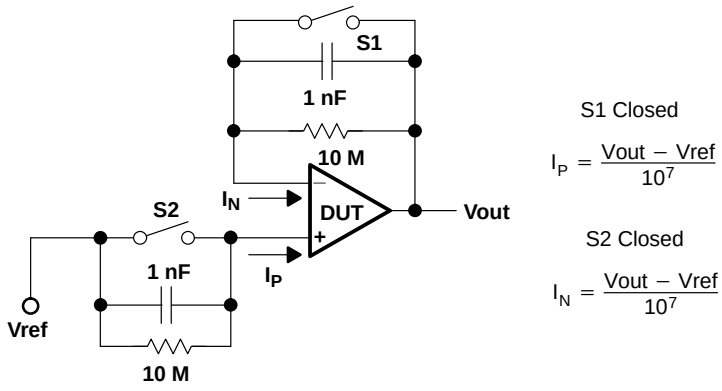


Figure 11–3. Test Circuit – I_{IB}

Input bias current is of concern when the source impedance is high. If the op amp has high input bias current, it will load the source and a lower than expected voltage is seen. The best solution is to use an op amp with either CMOS or JFET input. The source impedance can also be lowered by using a buffer stage to drive the op amp that has high input bias current.

In the case of bipolar inputs, offset current can be nullified by matching the impedance seen at the inputs. In the case of CMOS or JFET inputs, the offset current is usually not an issue and matching the impedance is not necessary.

The average temperature coefficient of input offset current, αI_{IO}, specifies the expected input offset drift over temperature. Its units are $\mu A/°C$. I_{IO} is measured at the temperature extremes of the part, and αI_{IO} is computed as $\Delta I_{IO}/\Delta °C$.

11.3.3 Input Common Mode Voltage Range

The input common voltage is defined as the average voltage at the inverting and noninverting input pins. If the common mode voltage gets too high or too low, the inputs will shut down and proper operation ceases. The common mode input voltage range, V_{ICR}, specifies the range over which normal operation is guaranteed.

Different input structures allow for different input common-mode voltage ranges:

The LM324 and LM358 use bipolar PNP inputs that have their collectors connected to the negative power rail. This allows the common-mode input voltage range to include the negative power rail.

The TL07X and TLE207X type BiFET op amps use P-channel JFET inputs with the sources tied to the positive power rail via a bipolar current source. This allows the common-mode input voltage range to include the positive power rail.

TI LinCMOS op amps use P-channel CMOS inputs with the substrate tied to the positive power rail. This allows the common-mode input voltage range to include the negative power rail.

Rail-to-rail input op amps use complementary N- and P-type devices in the differential inputs. When the common-mode input voltage nears either rail, at least one of the differential inputs is still active, and the common-mode input voltage range includes both power rails.

The trends toward lower, and single supply voltages make V_{ICR} of increasing concern.

Rail-to-rail input is required when a noninverting unity gain amplifier is used and the input signal ranges between both power rails. An example of this is the input of an analog-to digital-converter in a low-voltage, single-supply system.

High-side sensing circuits require operation at the positive input rail.

11.3.4 Differential Input Voltage Range

Differential input voltage range is normally specified as an absolute maximum. Exceeding the differential input voltage range can lead to breakdown and part failure.

Some devices have protection built into them, and the current into the input needs to be limited. Normally, differential input mode voltage limit is not a design issue.

11.3.5 Maximum Output Voltage Swing

The maximum output voltage, $V_{OM\pm}$, is defined as *the maximum positive or negative peak output voltage that can be obtained without wave form clipping, when quiescent DC output voltage is zero*. $V_{OM\pm}$ is limited by the output impedance of the amplifier, the saturation voltage of the output transistors, and the power supply voltages. This is shown pictorially in Figure 11–4.

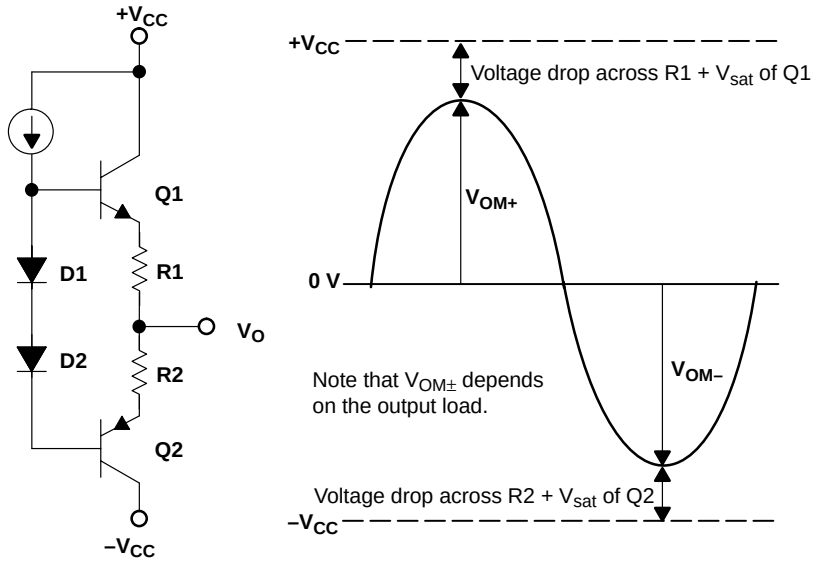


Figure 11–4. V_{OM}

This emitter follower structure cannot drive the output voltage to either rail. Rail-to-rail output op amps use a common emitter (bipolar) or common source (CMOS) output stage. With these structures, the output voltage swing is only limited by the saturation voltage (bipolar) or the on resistance (CMOS) of the output transistors, and the load being driven.

Because newer products are focused on single supply operation, more recent data sheets from Texas Instruments use the terminology V_{OH} and V_{OL} to specify the maximum and minimum output voltage.

Maximum and minimum output voltage is usually a design issue when dynamic range is lost if the op amp cannot drive to the rails. This is the case in single supply systems where the op amp is used to drive the input of an A to D converter, which is configured for full scale input voltage between ground and the positive rail.

11.3.6 Large Signal Differential Voltage Amplification

Large signal differential voltage amplification, A_{VD}, is similar to the open loop gain of the amplifier except open loop is usually measured without any load. This parameter is usually measured with an output load. Figure 11–11 shows a typical graph of A_{VD} vs. frequency.

A_{VD} is a design issue when precise gain is required. The gain equation of a noninverting amplifier:

$$\text{Gain} = \frac{1}{\beta} \times \frac{1}{1 + \frac{1}{A_{VD}\,\beta}} \qquad (11–2)$$

β is a feedback factor, determined by the feedback resistors. The term $\frac{1}{A_{VD}\,\beta}$ in the equation is an error term. As long as A_{VD} is large in comparison with $\frac{1}{\beta}$, it will not greatly affect the gain of the circuit.

11.3.7 Input Parasitic Elements

Both inputs have parasitic impedance associated with them. Figure 11–5 shows a model of the resistance and capacitance between each input terminal and ground and between the two terminals. There is also parasitic inductance, but the effects are negligible at low frequency.

Input impedance is a design issue when the source impedance is high. The input loads the source.

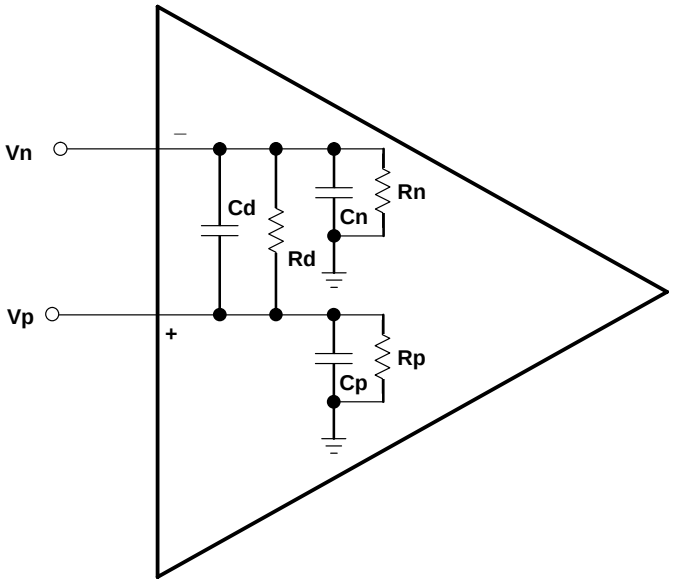


Figure 11–5. Input Parasitic Elements

11.3.7.1 Input Capacitance

Input capacitance, C_i, is measured between the input terminals with either input grounded. C_i is usually a few pF. In Figure 11–5, if $V_{p\ is}$ grounded, then $C_i = C_d \parallel C_n$.

Sometimes common-mode input capacitance, C_{ic} is specified. In Figure 11–5, if V_p is shorted to V_n, then $C_{ic} = C_p \parallel C_n$. C_{ic} is the input capacitance a common mode source would see referenced to ground.

11.3.7.2 Input Resistance

Input resistance, r_i is the resistance between the input terminals with either input grounded. In Figure 11–5, if V_p is grounded, then $r_i = R_d \parallel R_n$. r_i ranges from $10^7\ \Omega$ to $10^{12}\ \Omega$, depending on the type of input.

Sometimes common-mode input resistance, r_{ic}, is specified. In Figure 11–5, if V_p is shorted to V_n, then $r_{ic} = R_p \parallel R_n$. r_{ic} is the input resistance a common mode source would see referenced to ground.

11.3.8 Output Impedance

Different data sheets list the output impedance under two different conditions. Some data sheets list *closed-loop* output impedance while others list *open-loop* output impedance, both designated by Z_o.

Z_o is defined as the small signal impedance between the output terminal and ground. Data sheet values run from 50 Ω to 200 Ω.

Common emitter (bipolar) and common source (CMOS) output stages used in rail-to-rail output op amps have higher output impedance than emitter follower output stages.

Output impedance is a design issue when using rail-to-rail output op amps to drive heavy loads. If the load is mainly resistive, the output impedance will limit how close to the rails the output can go. If the load is capacitive, the extra phase shift will erode phase margin.

Figure 11–6 shows how output impedance affects the output signal assuming Z_o is mostly resistive.

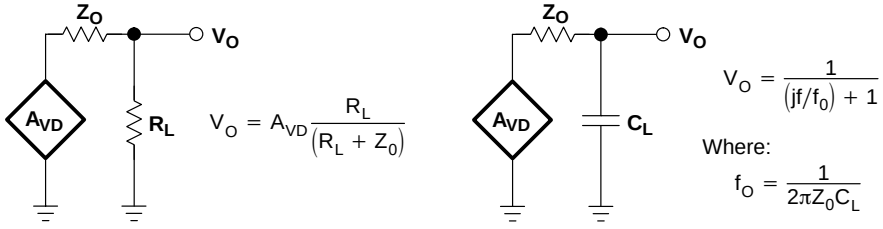


Figure 11–6. Effect of Output Impedance

Some new audio op amps are designed to drive the load of a speaker or headphone directly. They can be an economical method of obtaining very low output impedance.

11.3.9 Common-Mode Rejection Ratio

Common-mode rejection ratio, CMRR, is defined as the ratio of the differential voltage amplification to the common-mode voltage amplification, A_{DIF}/A_{COM}. Ideally this ratio would be infinite with common mode voltages being totally rejected.

The common-mode input voltage affects the bias point of the input differential pair. Because of the inherent mismatches in the input circuitry, changing the bias point changes the offset voltage, which, in turn, changes the output voltage. The real mechanism at work is $\Delta V_{OS}/\Delta V_{COM}$.

In a Texas Instruments data sheet, CMRR = $\Delta V_{COM}/\Delta V_{OS}$, which gives a positive number in dB. CMRR, as published in the data sheet, is a dc parameter. CMRR, when graphed vs. frequency, falls off as the frequency increases.

A common source of common-mode interference voltage is 50-Hz or 60-Hz ac noise. Care must be used to ensure that the CMRR of the op amp is not degraded by other circuit components. High values of resistance make the circuit vulnerable to common mode (and other) noise pick up. It is usually possible to scale resistors down and capacitors up to preserve circuit response.

11.3.10 Supply Voltage Rejection Ratio

Supply voltage rejection ratio, k_{SVR} (AKA power supply rejection ratio, PSRR), is the ratio of power supply voltage change to output voltage change.

The power voltage affects the bias point of the input differential pair. Because of the inherent mismatches in the input circuitry, changing the bias point changes the offset voltage, which, in turn, changes the output voltage.

For a dual supply op amp, $K_{SVR} = \frac{\Delta V_{CC\pm}}{\Delta V_{OS}}$ or $K_{SVR} = \frac{\Delta V_{DD\pm}}{\Delta V_{OS}}$. The term $\Delta V_{CC\pm}$ means that the plus and minus power supplies are changed symmetrically. For a single supply op amp, $K_{SVR} = \frac{\Delta V_{CC}}{\Delta V_{OS}}$ or $K_{SVR} = \frac{\Delta V_{DD}}{\Delta V_{OS}}$.

Also note that the mechanism that produces k_{SVR} is the same as for CMRR. Therefore k_{SVR} as published in the data sheet is a dc parameter like CMRR. When k_{SVR} is graphed vs. frequency, it falls off as the frequency increases.

Switching power supplies produce noise frequencies from 50 kHz to 500 kHz and higher. k_{SVR} is almost zero at these frequencies so that noise on the power supply results in noise on the output of the op amp. Proper bypassing techniques must be used (see Chapter 17) to control high-frequency noise on the power lines.

11.3.11 Supply Current

Supply current, I_{DD}, is the quiescent current draw of the op amp(s) with no load. In a Texas Instruments data sheet, this parameter is usually the total quiescent current draw for the whole package. There are exceptions, however, such as data sheets that cover single and multiple packaged op amps of the same type. In these cases, I_{DD} is the quiescent current draw for each amplifier.

In op amps, power consumption is traded for noise and speed.

11.3.12 Slew Rate at Unity Gain

Slew rate, SR, is the rate of change in the output voltage caused by a step input. Its units are V/μs or V/ms. Figure 11–7 shows slew rate graphically. The primary factor controlling slew rate in most amps is an internal compensation capacitor C_C, which is added to make the op amp unity gain stable. Referring to Figure 11–8, voltage change in the second stage is limited by the charging and discharging of the compensation capacitor C_C. The maximum rate of change is when either side of the differential pair is conducting $2I_E$. Essentially $SR = 2I_E/C_C$. Remember, however, that not all op amps have compensation capacitors. In op amps without internal compensation capacitors, the slew rate is determined by internal op amp parasitic capacitances. Noncompensated op amps have greater bandwidth and slew rate, but the designer must ensure the stability of the circuit by other means.

In op amps, power consumption is traded for noise and speed. In order to increase slew rate, the bias currents within the op amp are increased.

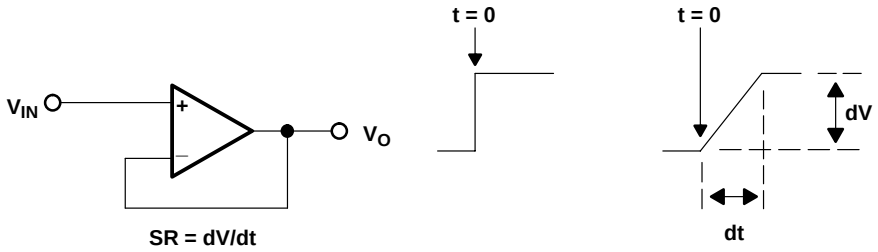


Figure 11–7.Figure 6. Slew Rate

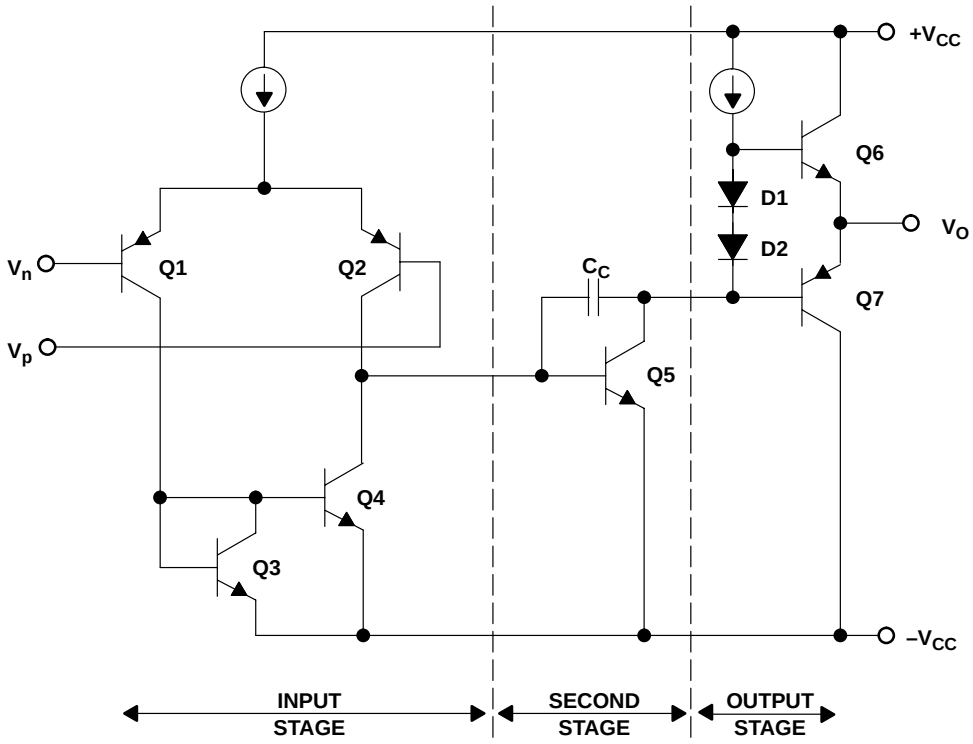


Figure 11–8.Figure 7. Simplified Op Amp Schematic

11.3.13 Equivalent Input Noise

Noise is covered in more detail in Chapter 10.

All op amps have parasitic internal noise sources. Noise is measured at the output of an op amp, and referenced back to the input. Therefore, it is called equivalent input noise.

Equivalent input noise parameters are usually specified as voltage, V_n, (or current, I_n) per root Hertz. For audio frequency op amps, a graph is usually included to show the noise over the audio band.

11.3.13.1 Spot Noise

The spectral density of noise in op amps has a pink and a white noise component. Pink noise is inversely proportional to frequency and is usually only significant at low frequencies. White noise is spectrally flat. Figure 11–9 shows a typical graph of op amp equivalent input noise.

Usually spot noise is specified at two frequencies. The first frequency is usually 10 Hz where the noise exhibits a 1/f spectral density. The second frequency is typically 1 kHz

where the noise is spectrally flat. The units used are normally $\frac{nV_{rms}}{\sqrt{Hz}}$ (or $\frac{pA_{rms}}{\sqrt{Hz}}$ for current noise). In Figure 11–9 the transition between 1/f and white is denoted as the corner frequency, f_{nc}.

11.3.13.2 Broadband Noise

A noise parameter like $V_{N(PP)}$, is the a peak to peak voltage over a specific frequency band, typically 0.1 Hz to 1 Hz, or 0.1 Hz to 10 Hz. The units of measurement are typically nV P–P.

Given the same structure within an op amp, increasing bias currents lowers noise (and increases SR, GBW, and power dissipation).

Also the resistance seen at the input to an op amp adds noise. Balancing the input resistance on the noninverting input to that seen at the inverting input, while helping with offsets due to input bias current, adds noise to the circuit.

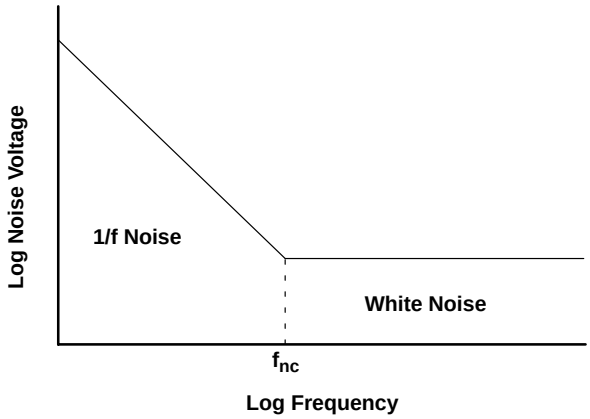


Figure 11–9. Typical Op amp Input Noise Spectrum

11.3.14 Total Harmonic Distortion Plus Noise

Total harmonic distortion plus noise, THD + N, compares the frequency content of the output signal to the frequency content of the input. Ideally, if the input signal is a pure sine wave, the output signal is a pure sine wave. Due to nonlinearity and noise sources within the op amp, the output is never pure.

THD + N is the ratio of all other frequency components to the fundamental and is usually specified as a percentage:

$$\text{THD} + \text{N} = \left[\frac{(\Sigma \text{ Harmonic voltages} + \text{Noise Voltages})}{\text{Fundamental}}\right] \times 100\% \qquad (11\text{–}3)$$

Figure 11–10 shows a hypothetical graph where THD + N = 1%. The fundamental is the same frequency as the input signal. Nonlinear behavior of the op amp results in harmonics

of the fundamental being produced in the output. The noise in the output is mainly due to the input noise of the op amp. All the harmonics and noise added together make up 1% of the fundamental.

Two major reasons for distortion in an op amp are the limit on output voltage swing and slew rate. Typically an op amp must be operated at or below its recommended operating conditions to realize low THD.

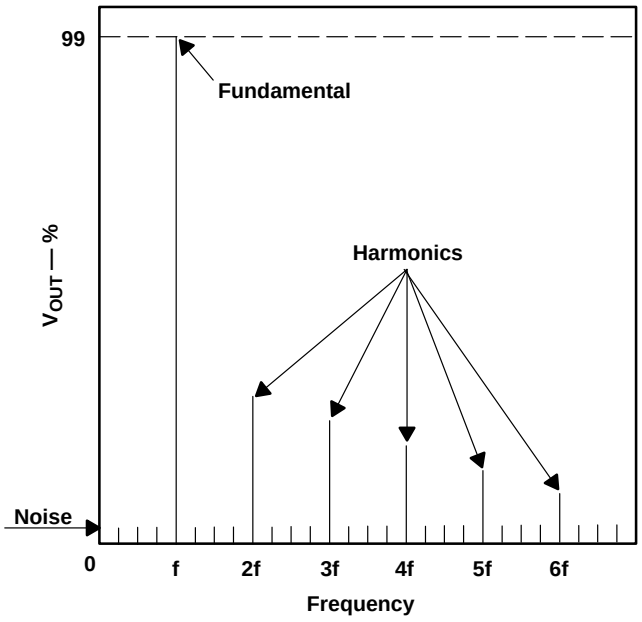


Figure 11–10. Output Spectrum with THD + N = 1%

11.3.15 Unity Gain Bandwidth and Phase Margin

There are five parameters relating to the frequency characteristics of the op amp that are likely to be encountered in Texas Instruments data sheets. These are unity-gain bandwidth (B_1), gain bandwidth product (GBW), phase margin at unity gain (ϕ_m), gain margin (A_m), and maximum output-swing bandwidth (B_{OM}).

Unity-gain bandwidth (B_1) and gain bandwidth product (GBW) are very similar. B_1 specifies the frequency at which A_{VD} of the op amp is 1:

$$B_1 = (f)A_{VD} = 1 \qquad (11–4)$$

GBW specifies the gain-bandwidth product of the op amp in an open loop configuration and the output loaded:

$$GBW = A_{VD} \times f \tag{11–5}$$

GBW is constant for voltage-feedback amplifiers. It does not have much meaning for current-feedback amplifiers because there is not a linear relationship between gain and bandwidth.

Phase margin at unity gain (ϕ_m) is the difference between the amount of phase shift a signal experiences through the op amp at unity gain and 180°:

$$\phi_m = 180^\circ - \phi @ B1 \tag{11–6}$$

Gain margin is the difference between unity gain and the gain at 180° phase shift:

$$\text{Gain margin} = 1 - \text{Gain} @ 180^\circ \text{ phase shift} \tag{11–7}$$

Maximum output-swing bandwidth (B_{OM}) specifies the bandwidth over which the output is above a specified value:

$$B_{OM} = f_{MAX,} \text{ while } V_O > V_{MIN} \tag{11–8}$$

The limiting factor for B_{OM} is slew rate. As the frequency gets higher and higher the output becomes slew rate limited and can not respond quickly enough to maintain the specified output voltage swing.

In order to make the op amp stable, capacitor, C_C, is purposely fabricated on chip in the second stage (Figure 11–8). This type of frequency compensation is termed dominant pole compensation. The idea is to cause the open-loop gain of the op amp to roll off to unity before the output phase shifts by 180°. Remember that Figure 11–8 is very simplified, and there are other frequency shaping elements within a real op amp.

Figure 11–11 shows a typical gain vs. frequency plot for an internally compensated op amp as normally presented in a Texas Instruments data sheet.

As noted earlier, A_{VD} falls off with frequency. A_{VD} (and thus B_1 or GBW) is a design issue when precise gain is required of a specific frequency band.

Phase margin (ϕ_m) and gain margin (A_m) are different ways of specifying the stability of the circuit. Since rail-to-rail output op amps have higher output impedance, a significant phase shift is seen when driving capacitive loads. This extra phase shift erodes the phase margin, and for this reason most CMOS op amps with rail-to-rail outputs have limited ability to drive capacitive loads.

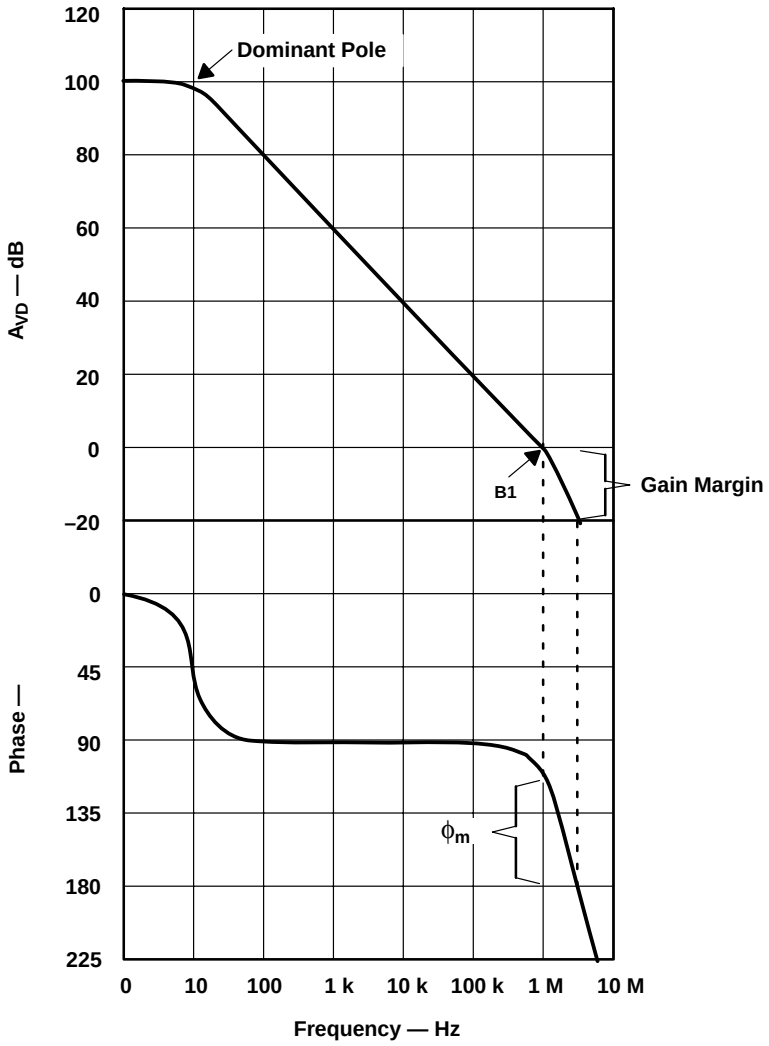


Figure 11–11. Voltage Amplification and Phase Shift vs. Frequency

11.3.16 Settling Time

It takes a finite time for a signal to propagate through the internal circuitry of an op amp. Therefore, it takes a period of time for the output to react to a step change in the input. In addition, the output normally overshoots the target value, experiences damped oscillation, and settles to a final value. Settling time, t_S, is the time required for the output voltage to settle to within a specified percentage of the final value given a step input.

Figure 11–12 shows this graphically:

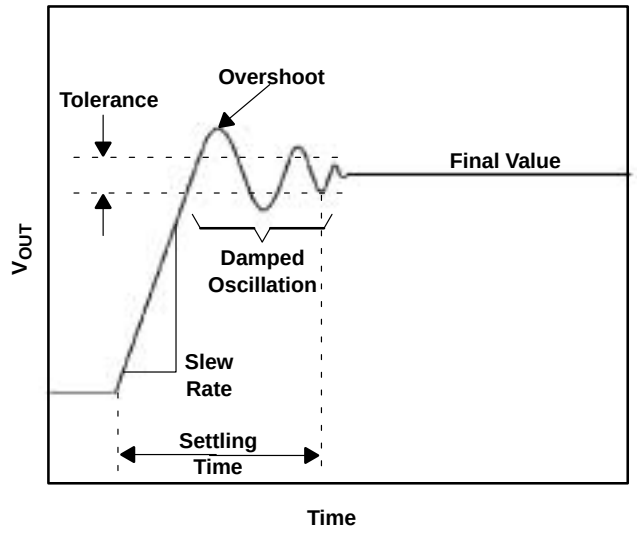


Figure 11–12. Settling Time

Settling time is a design issue in data acquisition circuits when signals are changing rapidly. An example is when using an op amp following a multiplexer to buffer the input to an A to D converter. Step changes can occur at the input to the op amp when the multiplexer changes channels. The output of the op amp must settle to within a certain tolerance before the A to D converter samples the signal.

Chapter 12

Instrumentation: Sensors to A/D Converters

Ron Mancini

12.1 Introduction

The typical transducer measurement system block diagram is shown in Figure 12–1. The transducer is the electronic system's interface with the real world, and it issues data about a variable. The transducer converts the data into an electrical signal adequate for processing by the circuitry that follows the transducer. Bias and excitation circuitry does the care and feeding of the transducer, thus this circuitry provides offset voltages, bias currents, excitation signals, external components, and protection that is required for the transducer to operate properly. The output of the transducer is an electrical signal representing the measured variable.

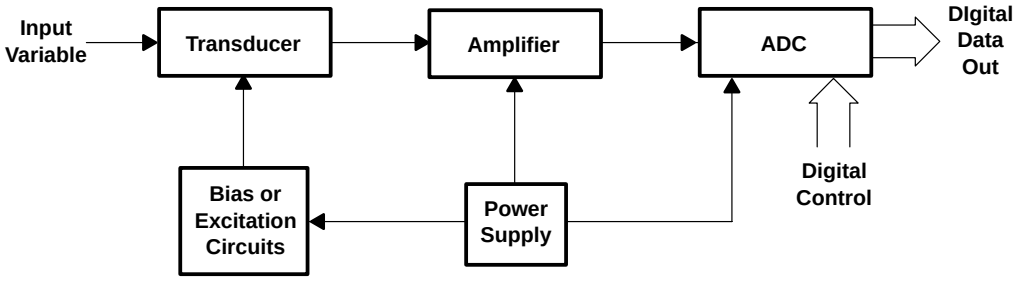


Figure 12–1. Block Diagram of a Transducer Measurement System

The variables that must be measured are determined by the customer's application, and the *measured variable* normally dictates the transducer selection. If the measured variable is temperature, then some sort of temperature sensing transducer must be employed, and the range of temperatures to be measured or the accuracy of the measurement is the primary factor influencing temperature transducer selection. Notice that the electrical output of the transducer is not a major concern at this point in the transducer selection. The transducer's electrical output is always a consideration, although picking the right transducer for the job is the primary goal. The correct transducer for the job can have an Ω/°C change, μV/°C change, or mV/°C change. All transducers have offset voltages or currents, and they can be referenced to ground, either power supply rail, or some other voltage. The selection of the transducer is out of circuit designer's hands; thus, the circuit designer must accept what the application demands.

The ADC selection is based on several system criteria such as resolution, conversion speed, power requirements, physical size, processor compatibility, and interface structure. The ADC must have enough bits to obtain the resolution required by the accuracy specification. The formula for calculating the resolution of an ADC is given in Equation 12–1 where n is the number of significant bits contained in the ADC.

$$\text{RESOLUTION} = 2^n \tag{12–1}$$

Some confusion exists about the word *bits* because the same word is used for binary bits and significant bits. Binary bits are ones and zeros used to calculate binary numbers; for example, a converter with 8 different digital states has 8 significant bits and $2^n = 256$ binary bits (see Figure 12–2). The voltage value of a single bit, called a least significant bit (LSB), is calculated in Equation 12–2.

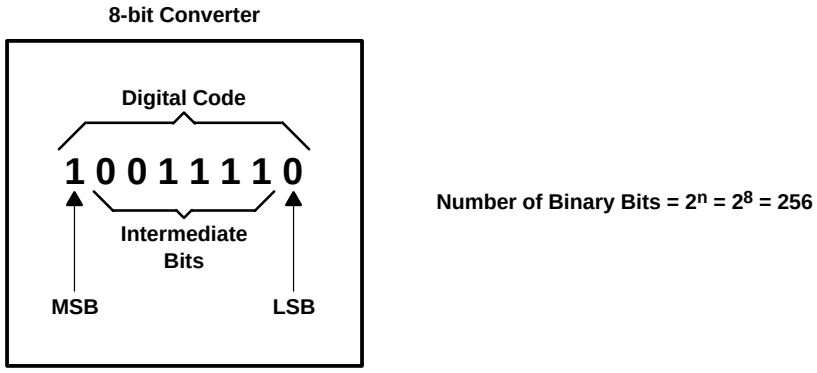


Figure 12–2. Significant Bits versus Binary Bits

$$\text{LSB} = \frac{\text{FSV}}{2^n} \tag{12–2}$$

FSV is the full-scale voltage of the converter in volts; hence, a 12-bit converter FSV = 10 volts has an LSB equal to $10/2^{12} = 2.441406$ mV/bit. In an ADC, an LSB is the maximum voltage change required to for a one-bit output change, and in an ADC an LSB is defined as $\text{LSB} = \text{FSV}/(2^n - 1)$.

Conversion speed is not critical in temperature measurement applications because temperature changes occur at slow rates. Directional control in a rocket traveling at Mach 2 happens much faster than temperature changes, so conversion speed is an important factor in rocket applications. The speed of an ADC is generally thought of as the conversion time plus the time required between conversions, and conversion speed dictates the converter structure. When conversion speed is a primary specification a flash converter is used, and flash converters require low impedance driving circuits. This is an example of a converter imposing a specification requirement, low output impedance, on the driving amplifier.

The system definition specifies the voltages available for design and the maximum available current drain. Some systems have multiple voltages available, and others are limited to a single voltage. The available voltage impacts the converter selection. Size, processor compatibility, and interface structure are three more factors that must be considered when selecting the ADC. The available package that the converter comes in determines what footprint or size that the converter takes on the printed circuit board. Some applications preclude large power-hungry ADCs, so these applications are limited to recursive or $\Sigma\Delta$-type ADCs. The converter must be compatible with the processor to preclude the addition of glue logic; thus, the processor dictates the ADC's structure. This defines the interface structure, sometimes the ADC structure, and the ADC timing.

Notice that the amplifier designer has not been consulted during this decision process, but in actuality, the systems engineers do talk to the amplifier designers if only to pacify them. The selection of the ADC is out of circuit designer's hands; thus, the circuit designer must accept what the application demands.

There are many different transducer/ADC combinations, and each combination has a different requirement. Although they may be natural enemies, any transducer may be coupled with any ADC, thus the amplifier must make the coupling appear to be seamless. There is no reason to expect that the selected transducer's output voltage span matches the selected ADC's input voltage span, so an amplifier stage must match the transducer output voltage span to the ADC input voltage span. The amplifier stage amplifies the transducer output voltage span and shifts its dc level until the transducer output voltage span matches the ADC input voltage span. When the spans are matched, the transducer/ADC combination achieve the ultimate accuracy; any other condition sacrifices accuracy and/or dynamic range.

The transducer output voltage span seldom equals the ADC input voltage span. Transducer data is lost and/or ADC dynamic range is not fully utilized when the spans are unequal, start at different dc voltages, or both. In Figure 12–3 (A), the spans are equal (3 V), but they are offset by 1 V. This situation requires level shifting to move the sensor output voltage up by one volt so the spans match. In Figure 12–3 (B), the spans are unequal (2 V and 4 V), but no offset voltage exists. This situation requires amplification of the sensor output to match the spans. When the spans are unequal (2 V versus 3 V) and offset (1 V), as is the case in Figure 12–3 (C), level shifting and amplification are required to match the spans.

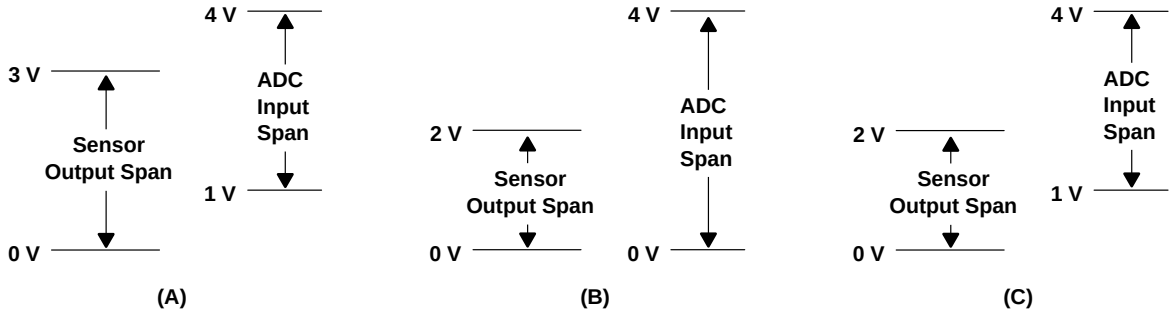


Figure 12–3. Example of Spans That Require Correction

The output span of the transducer must be matched to the input span of the ADC to achieve optimum performance. When the spans are mismatched either the transducer output voltage does not fit into the ADC input span thus losing sensor data, or the transducer output voltage does not fill the ADC input span thus losing ADC accuracy. The latter situation requires an increase in ADC dynamic range (increased cost) because a higher bit converter must be used to achieve the same resolution. The best analog circuit available for matching the spans is the op amp because it level shifts and amplifies the input voltage to make the spans equal. The op amp is so versatile that it shifts the signal's dc level and amplifies the input signal simultaneously.

A similar but different problem exists in the digital to analog converter (DAC) to actuator interface. The DAC output voltage or current span must match the actuator input voltage span to achieve maximum performance. The procedure for matching the DAC output span to the actuator input span can be quite different from the procedure for matching the transducer output span to the ADC input span. Transducer outputs are usually low-level signals, thus care must be taken to preserve their signal to noise ratio. Actuator input signals may require significant power, thus robust op amps are required to drive some actuators.

The system specifications eventually determine the transducer, ADC, and analog circuit specifications. System specifications are seen as absolute specifications; they must be met for the design to function in a satisfactory manner. Component specifications are divided into several categories; absolute maximum ratings (AMR), guaranteed minimum/maximum specifications (V_{MAX} or V_{MIN}), typical specifications (V), and guaranteed but not tested specifications (GNT).

If any of the device parameters are taken beyond the AMR, the device can be destroyed (expect destruction). The manufacturer guard bands the AMR to guarantee safety and quality, and you should guard band the AMRs, too. Typical specifications are the most appealing, but throw them out because they are meaningless in most cases. In the vast majority of cases the typical specifications are not related to meaningful data; rather, they are marketing dreams. Never design with typical specifications unless you are in the habit of designing with meaningless data or have a good reason for believing that typical is close to reality. A violation of this rule is a specification like output voltage swing that de-

pends heavily on test conditions such as the value of the load resistor. When the load resistance is much higher than that specified for the test condition, the output voltage swing is closer to the typical rather than the guaranteed specification. The laws of physics guarantee this truth, but the amount of extra voltage swing that is achieved is hard to calculate.

Guaranteed min/max specifications define the limits of a parameter. The parameter will always exceed the minimum value, and it will never exceed the maximum value. The guaranteed min/max specifications are your design specifications. Guaranteed but not tested (GNT) specifications are usually applied to parameters that are very expensive to test. The manufacturer either tests some other parameter related to the specification or they sample test each lot to insure compliance. GNT specifications are design specifications. There is a fifth specification called guaranteed by design (GBD), and if they are not critical specifications, GBD is a useful design specification.

All specifications have conditions associated with testing. They specify ambient temperature, supply voltage, test signals, test loads, and other conditions, and they define how the measurements are made. Inspect the test conditions carefully; an op amp that specifies a 5-V output swing with a 50-Ω load is much more capable of driving a load than an op amp that specifies a 5-V output swing with a 10-kΩ load. Beware, you assume the risk of a parameter being out of specification when you use devices at conditions other than the test conditions.

An error budget is a logical and orderly method of tabulating errors, and it helps the designer keep track of the errors by error sources. Meeting the system specifications translates into minimizing errors, choosing components with acceptable errors, and canceling or eliminating errors when possible. The error budget is first applied to the ADC and transducer because they are the components that the designer has the least control over. When these two error budgets are combined and subtracted from the system error allowance the result is the error allowance for the amplifier and peripheral circuits. The designer must choose components and design wisely to stay within the error allowance, or the system specifications are not met.

Sometimes the systems specifications can’t be met, and this fact is greeted with moans, name calling, and finger pointing. The error budgets are the design engineer’s only defense against subjective accusations. The error budgets document the design trail, and they show where changes need to be made to do the best job possible. It is very hard to maintain an error budget because errors must be converted to equivalent units (volts, bits, or amps), and the impact of the error term may not be calculable at that point in the design. When this situation occurs, tabulate the error terms in a table and calculate their effect later in the design process.

This chapter teaches the designer how to characterize the transducer and ADC, how to determine amplifier and design specifications through the use of error budgets, and how to complete circuit design. The equations developed in this chapter are not nearly as important as the design philosophy.

12.2 Transducer Types

This is not a treatise on transducers, but an appreciation for the many different types of transducers gives a feel for the extent and complexity of the transducer characterization problem (following section). The variety of electrical output that transducers offer loosely groups transducers. Various types of transducer outputs are resistive, optical, ac-excited, junction voltage, and magnetic, and each of these outputs must be converted to an electrical signal that can be amplified to fit the input span of an ADC. There are excellent references[1] that deal with transducer characterization, but the transducer manufacturer should be your first source of reference material.

The transducer manufacturer publishes data similar to that contained in an IC data sheet, and they take the same liberties with typical specifications that the IC manufacturers do. Sort through the data to determine the meaning of the various specifications, and then pay special attention to the test conditions that prevailed when the data was gathered. Look for application notes that show transducer excitation, bias, or interface circuits. Search several manufacturers for similar information because nobody manages to cover every aspect of a design

Some transducers such as strain gages, thermistors, RTDs, and potentiometers sense a change in resistance (ΔR). The ΔR sensing devices are used in at least three circuit configurations; the voltage divider, current excited, and Wheatstone bridge circuits shown in Figures 12–4, 12–5, and 12–6. The transducer resistance is R_T, and the change in transducer resistance caused by a change in the measured variable is ΔR.

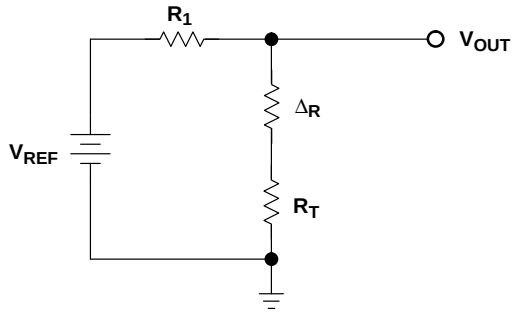


Figure 12–4. Voltage Divider Circuit for a Resistive Transducer

The voltage divider circuit uses a stable reference voltage to convert the transducer resistance into voltage, and its output voltage is given in Equation 12–3.

$$V_{OUT} = V_{REF} \frac{\Delta R + R_T}{\Delta R + R_T + R_1} \qquad (12–3)$$

If R_1 is comparable in value with R_T, the circuit has very low sensitivity because the circuit must measure a small change in resistance in the presence of a large resistance. When

the bias resistor, R_1, is selected as a large value, V_{REF} and R_1 act as a current source, and the transducer resistance can be neglected in the calculations thus yielding Equation 12–4. When $R_1 >> (R_T + \Delta R)$ Equation 12–3 reduces to Equation 12–4.

$$V_{OUT} = \frac{V_{REF}}{R_1}\left(\Delta R + R_T\right) \qquad (12–4)$$

Equation 12–5 is the equivalent of Equation 12–4, and it is obtained by exciting the transducer with a bias current as shown in Figure 12–5. The bias current can be made very accurate by employing op amps in a current source configuration as shown in Figure 12–6, thus the approximation $R_1 >> (R_T + \Delta R)$ need not enter the calculations.

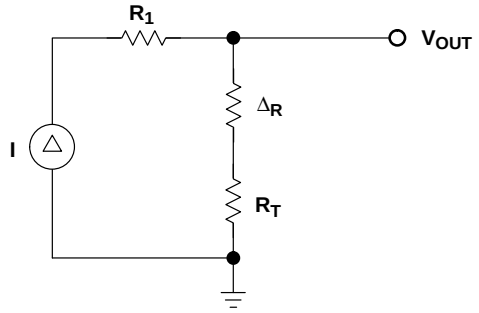


Figure 12–5. Current Source Excitation for a Resistive Transducer

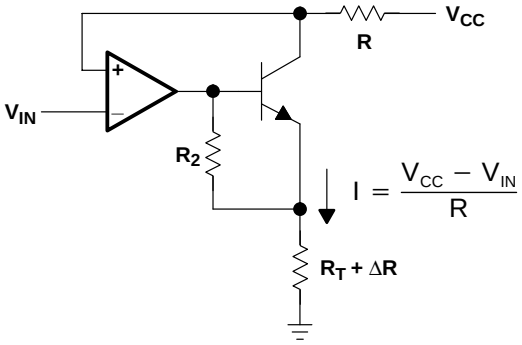


Figure 12–6. Precision Current Source

$$V_{OUT} = I\left(\Delta R + R_T\right) \qquad (12–5)$$

The Wheatstone bridge shown in Figure 12–7 is a precision device used to measure small changes in resistance. One leg of the bridge is made up of a voltage divider consisting of equal stable resistors (R_1 and R_2) and the reference voltage. When R_X and ΔR equal zero, R_{TX} is selected equal to R_T. As the transducer resistance changes ΔR assumes

some value, and R_X is switched until the bridge output voltage nulls to zero. At this point the value of ΔR is read from the R_X dial. Bridge circuits are used to convert resistive transducer values to dial readings, but there are methods of using transducers in bridge circuits that yield a voltage change proportional to the resistance change. The bridge circuit has a high output impedance, thus op amps configured in an instrumentation configuration (both inputs are equal high resistances) must be used to amplify the output voltage from bridge circuits.

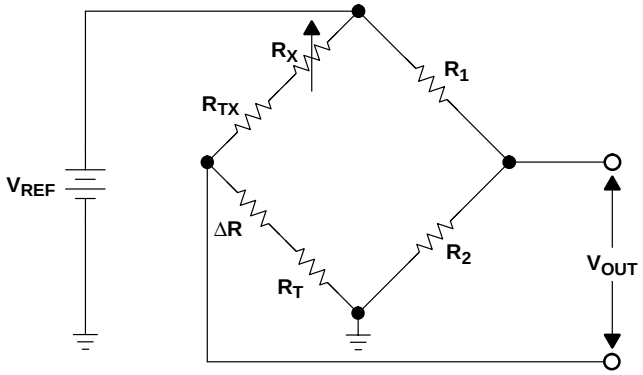


Figure 12–7. Wheatstone Bridge Circuit

The three most popular optical transducers are the photoconductive cell, the photodiode and the photovoltaic cell. The photoconductive cell acts like a light sensitive resistor, thus one of the circuits shown in Figures 12–4, 12–5, or 12–7 that convert resistance changes to voltage is used in photoconductive cell applications. The photodiode is a very fast diode with a small output current, and the circuit shown in Figure 12–8 is used to convert current to voltage. The photodiode is reversed biased with a constant voltage, so the photodiode terminating voltage stays constant thus maintaining linearity. The photodiode amplifier output voltage equation is Equation 12–6.

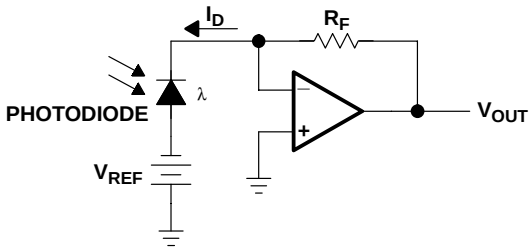


Figure 12–8. Photodiode Amplifier

$$V_{OUT} = I_D R_F \quad (12–6)$$

The phototransistor has a junction that is light sensitive, and the junction has a transparent cover so that it can sense ambient light. The collector-base junction of the transistor is reverse biased, and normal transistor action takes place with the ambient light induced base current taking place of the normal base current (see Figure 12–9).

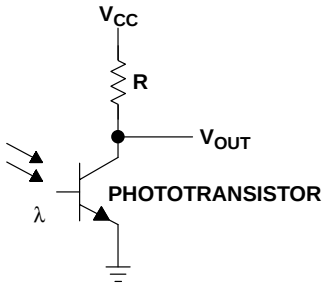


Figure 12–9. Phototransistor Amplifier

The photovoltaic or solar cell circuit is shown in Figure 12–10. The circuit zero-biases the cell for minimum leakage current, and the cell's output current is a linear function of the area exposed to light. When the photovoltaic cell is properly masked and evenly flooded with light, it operates as a linear distance transducer (see Figure 12–10 and Equation 12–6).

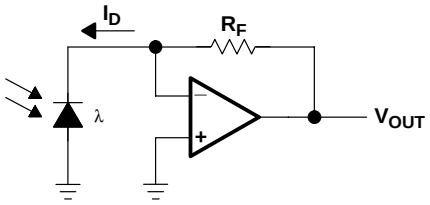


Figure 12–10. Photovoltaic Cell Amplifier

AC-excited transducers are usually used to make motion and/or distance sensors. In one type of ac-excited transducer, a stationary winding is excited with an ac current, and another winding is moved past the stationary winding inducing a voltage in the second winding. In a well-designed transducer, the induced voltage is proportional to distance, hence the output voltage is proportional to distance. Another ac-excited transducer uses two plates; one plate is excited with an ac current, and the other plate is ground. An object coming near the excited plate changes the capacitance between the plates, and the result is an output voltage change.

Resolvers and synchros are position transducers that indicate position as a function of the phase angle between the exciting signal and the output signal. Resolvers and synchros normally are multiple-winding devices excited from two or more sources. They indicate position very accurately, but their special circuitry requirements, cost, and weight limit them to a few applications such as airfoil control surfaces and gyros.

AC-excited transducers require a rectifier circuit to make the output voltage unipolar prior to integration. Coarse transducers use a diode or diode bridge to rectify the output voltage, but diodes are not adequate for precision applications because their forward voltage drop is temperature sensitive and poorly regulated. The diode problems are overcome through the use of feedback in the active full wave rectifier circuit shown in Figure 12–11. An integrating capacitor, C, is added to the circuit so the output voltage is a dc voltage proportional to the average voltage value of the input voltage.

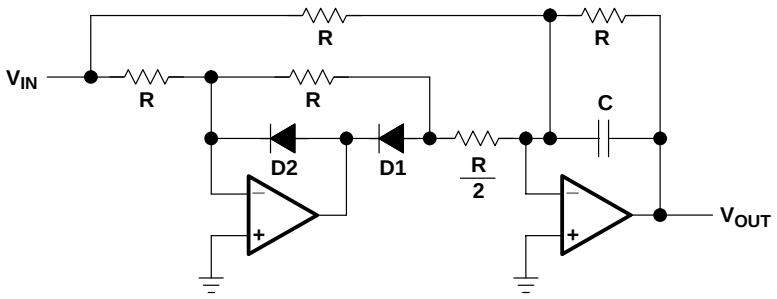


Figure 12–11. Active Full-Wave Rectifier and Filter

Semiconductor or wire junctions (thermocouples) are often used as temperature transducers because there is a linear relationship between temperature and output voltage over a restricted temperature range. Thermocouples have small voltages varying from μV/°C to mV/°C, and they normally are configured with thermistors and zeroing resistors in the output circuit. Thermocouples have small output voltages and high output resistance, thus a special op amp called an instrumentation amplifier is required for thermocouple amplification. An instrumentation amplifier has very high and equal input impedances, thus they don't load the input signal source.

Semiconductor junctions have a nominal temperature coefficient (TC) of –2 mV/°C. The TC is linear, but it varies from diode to diode because of manufacturing techniques, semiconductor materials, and bias currents. In a well-controlled application where thermal mass is insignificant, semiconductor junctions make excellent temperature transducers. The junction effect is so stable and linear that commercial temperature transducers have become available in a single IC.

Magnetic fields can be sensed by the Hall effect, and special semiconductors called Hall-effect sensors have been developed to sense magnetic fields. Current is passed through

the semiconductor in a direction perpendicular to the magnetic field. A pair of voltage pick off leads is placed perpendicular to the direction of current flow, and the output voltage is proportional to the magnetic field strength. The manufacturing process for Hall effect transducers is a standard semiconductor manufacturing process, so Hall effect transducers are offered for sale as transistors or ICs.

12.3 Design Procedure

A step-by-step design procedure that results in the proper op amp selection and circuit design is given below. This design procedure works best when the op amp has almost ideal performance, thus the ideal op amp equations are applicable. When nonideal op amps are used, parameters like input current affect the design, and they must be accounted for in the design process. The latest generation of rail-to-rail op amps makes the ideal op amp assumption more valid than it ever was.

No design procedure can anticipate all possible situations, and depending on the op amp selected, procedure modifications may have to be made to account for op amp bias current, input offset voltage, or other parameters. This design procedure assumes that system requirements have determined the transducer and ADC selection and that changing these selections adversely impacts the project.

1) Review the system specifications to obtain specifications for noise, power, current drain, frequency response, accuracy, and other variables that might affect the design.

2) Characterize the reference voltage including initial tolerances and drift.

3) Characterize the transducer to determine its salient parameters including output voltage swing, output impedance, dc offset voltage, output voltage drift, and power requirements. These parameters determine the op amp's required input voltage range (V_{IN1} to V_{IN2}), and input impedance requirements. The offset voltage and voltage drift are tabulated as errors. At this point it is assumed that the selected op amp's input voltage span is greater than the transducer's output voltage excursion. Design peripheral circuits if required.

4) Scrutinize the ADC's specification sheet to determine it's required input voltage range because this range eventually sets the op amp's output voltage swing requirement (V_{OUT1} to V_{OUT2}). Determine the ADC's input resistance, input capacitance, resolution, accuracy, full-scale range, and allowable input circuit charge time. Calculate the LSB value.

5) Create an error budget (in bits) for the transducer and ADC. Use the transducer/ADC error budget to determine the value and range of the critical op amp parameters. Select an op amp, and justify the selection by creating an error budget for the op amp circuit.

6) Scan the transducer and ADC specifications, and make a set of analog interface amplifier (AIA) specifications

7) Complete the AIA circuit design.

8) Build the circuit, and test it.

12.4 Review of the System Specifications

The power supply has only one voltage available, and that voltage is 5V ± 5% = 5 V ± 250 mV. The power supply is connected with the negative terminal at ground and the positive terminal at V_{CC}. This is not a portable application, thus the allowed current drain, 50 mA, is adequate for the job. No noise specifications are given, but the proposed power, ground, and signal traces are being done on high-quality circuit board material with planes and good size copper. A system of this quality should experience no more than 50 mV of noise on the logic power lines and 10 mV of noise on the analog power lines.

This is a temperature measuring system that requires updates every 10 seconds. Clearly, ADC conversion speed or input charging rate is not cause for consideration. The low conversion speed translates into lower logic speed, and slow logic means less noise generated. The temperature transducer is located at the end of a three-foot long cable, so expect some noise picked up by the cable to be introduced into the circuit. Fortunately, the long time between ADC conversions enable extensive filtering to reduce the cable noise.

The system accuracy required is 11 bits. The application measures several parameters so it is multiplexed, and a TLV2544 12 bits resolution ADC has been selected. The temperature transducer is a diode, and the temperature span to be measured is –25°C to 100°C. The ambient temperature of the electronics package is held between 15°C to 35°C.

12.5 Reference Voltage Characterization

A reference voltage is required to bias the transducer and act as a reference voltage for the analog interface amplifier (AIA). Selecting a reference with a total accuracy better than the accuracy specification (11 bits) does not guarantee meeting the system accuracy specification because other error sources exist in the design. Resistor tolerances, amplifier tolerances, and transducer tolerances all contribute to the inaccuracy, and the reference can’t diminish these errors. The quandary here is a choice between an expensive reference and expensive accurate components, or an adjustment to null out initial errors. This quandary boils down to which is the lesser of two evils; expensive components or the expense of an adjustment.

System engineering has decided that they want the adjustment, so the reference does not have to have 11-bit accuracy. A TL431A voltage reference is chosen for the design. The output voltage specification at 25°C and 10-mA bias current is 2495 mV ± 25 mV. This

reference has a temperature drift of 25 mV over 70°C, and this translates to 7.14 mV drift over a 20°C temperature range. There is another drift caused by the cathode voltage change, and this drift is 2.7 mV/V. The supply voltage regulation is 0.5 V, but much of this tolerance is consumed by the initial tolerance and wiring scheme, so the less than 0.1 V is due to regulator drift. The total drift is 7.14 mV + 0.27 mV = 7.41 mV. This yields a total drift of 0.3% maximum. The amplifier usually uses a fraction of the reference voltage, so the final AIA will not drift the full 0.3%.

12.6 Transducer Characterization

The temperature transducer is a special silicon diode that is characterized for temperature measurement work. When this diode is forward biased at 2.0 mA ± 0.1 mA its forward voltage drop is 0.55 V ± 50 mV, and its temperature coefficient is –2 mV/°C. The wide acceptable variation in bias current makes this an easy device to work with. The circuit for the bias calculations is shown in Figure 12–12.

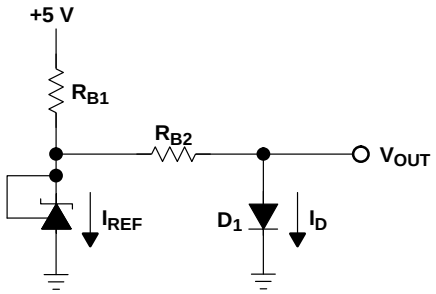


Figure 12–12. Reference and Transducer Bias Circuit

The current through R_{B1} is calculated in Equation 12–7. Remember, the reference must be biased at 10 mA, and the transducer must be biased at 2 mA.

$$I = I_{REF} + I_D = 10 + 2 = 12 \text{ mA} \quad (12\text{–}7)$$

The value of R_{B1} is calculated in Equation 12–8, and the value of R_{B2} is calculated in Equation 12–9.

$$R_{B1} = \frac{V_{+5} - V_{ref}}{I} = \frac{5 - 2.495}{12} = 208\ \Omega \quad (12\text{–}8)$$

$$R_{B2} = \frac{V_{ref}}{I_D} = \frac{2.495}{2} = 1247\ \Omega \quad (12\text{–}9)$$

Both resistors are selected from the list of 1% decade values, thus R_{B1} = 210 Ω, 1%, and R_{B2} = 1240 Ω, 1%. The resistor values have been established, so it is time to calculate

the worst case excursions of I_D (Equations 12–10 and 12–11). The resistors are assumed to have a 2% tolerance in these calculations. The extra 1% allows for temperature changes, vibration, and life. Three percent tolerances would have been used if the electronics' ambient temperature range were larger.

$$I_{D(MIN)} = \frac{V_{REF(MIN)}}{R_{B2(MAX)}} = \frac{2.47 - 0.025 - 0.007}{1.02\ (1.24)} = 1.93\ mA \quad (12–10)$$

$$I_{D(MAX)} = \frac{V_{REF(MAX)}}{R_{B2(MIN)}} = \frac{2.52 - 0.025 - 0.007}{0.98\ (1.24)} = 2.10\ mA \quad (12–11)$$

The bias current extremes do not exceed the transducer bias current requirements, so the transducer will meet the specifications advertised. The converter is 12 bits and the full-scale voltage is assumed to be 5 V, so the value of an LSB is calculated in Equation 12–12. The nominal transducer output voltage is 550 mV at an ambient temperature of 25°C. At –25°C, the transducer output voltage is 550 mV+ (–2 mV/°C)(–50°C) = 650 mV. At 125°C, the transducer output voltage is 550 mV + (–2 mV°C)(75°C) = 400 mV. This data is tabulated in Table 12–1.

$$LSB = \frac{FSV}{2^N} = \frac{5}{2^{12}} = 1.22\ mV \quad (12–12)$$

Table 12–1. Transducer Output Voltage

TRANSDUCER TEMPERATURE	TRANSDUCER OUTPUT VOLTAGE	ANALOG INTERFACE AMPLIFIER INPUT VOLTAGE
–25°C	650 mV	V_{IN1} = 650 mV
25°C	550 mV	550 mV
100°C	400 mV	V_{IN2} = 400 mV

The steady state (V_{TOS}) offset voltage is ± 50 mV, thus transducer output voltage (V_{TOV}) ranges from 350 mV to 700 mV. The offset voltage is stripped out by the adjustments in the AIA, so it is not of any concern here. V_{TOS} spans 100 mV, thus it is a 100 mV/1.22 mV/bit = 82 bit error unless it is adjusted out.

The output impedance of the transducer is equivalent to the resistance of a forward biased diode (Equation 12–13).

$$R_D = \frac{26}{I} = \frac{26}{2} = 13\ \Omega \quad (12–13)$$

At this stage of the design there are two parameters that influence the accuracy of the measurement, and they are the temperature coefficient of the transducer and the output

impedance of the transducer. The temperature transducer has been biased correctly, thus its temperature coefficient should be the advertised value of –2 mV/°C. The output impedance of the transducer forms a voltage divider with the input resistance of the AIA, but this error can't be calculated until the AIA is selected. The final transducer error contribution is that portion of the V_{TOS} that can't be adjusted out, and this error is determined during the AIA design.

12.7 ADC Characterization

This particular ADC was selected because it has a multiplexer and it enables different modes of operation. The temperature measurement is done in the single-shot mode because this mode allows the user to set the charge time at the input to the converter. During charging, the ADC's input resistance is low, but after the ADC input is charged the input resistance rises to 20 kΩ. This high input resistance does not load the AIA output circuit, thus the AIA achieves full rail-to-rail output voltage swing.

The internal reference is used in this application, and the reference sets the input voltage span required to obtain full accuracy for the ADC. Using the internal reference, the input voltage span is 0 V to 4 V. The offset voltage (V_{ADCOS}) is ± 150 mV, and the voltage drift is 40 PPM/°C. The voltage drift over the full temperature range is 40 PPM/°C(20°C) = 800 PPM. There are 244 PPM/LSB in a 12-bit converter, so the drift voltage error is 800/244 ≈ 4 bits error.

The ADC output is full scale (all bits 1) when the input voltage is 4 V, and it is zero (all bits 0) when the input voltage is 0 V. This data is tabulated in Table 12–2. Because the full scale output voltage has changed to 4 volts the LSB is calculated to be $4/(2^{12})$ = 976.6 μV/bit.

Table 12–2. ADC Input Voltage

ADC INPUT VOLTAGE	DIGITAL OUTPUT	ANALOG INTERFACE AMPLIFIER OUTPUT VOLTAGE
0 V	000000000000	V_{OUT1} = 0 V
4 V	111111111111	V_{OUT2} = 4 V

12.8 Op Amp Selection

It is time to select the op amp, and the easiest way to do this is to list the known specifications or requirements, list a candidate op amp's specifications, and calculated the projected error that the candidate op amp yields.

Table 12–3. Op Amp Selection

DESIGN SPECIFICATION	ESTIMATED VALUE	CANDIDATE OP AMP: TLV247X
R_{IN}	10^6 (13) Ω	10^{12} Ω
V_{TOV}	350 mV to 700 mV	–0.2 V to 5.2 V
R_{OUT}		1.8 Ω
V_{INADC}	0 V to 4 V	0.15 V to 4.85 V
V_{OS}	——	2.2 mV
I_B	——	100 pA
V_N	——	$\frac{28\ nV}{\sqrt{Hz}}$
I_N	——	$\frac{0.39\ pA}{\sqrt{Hz}}$
Analog noise	——	10 mV
k_{SVR}	——	63 dB

There should be almost no error from R_{IN} because the transducer output impedance is very low. The high side of the op amp's output voltage swing (4.85 V) is much higher than the ADC input voltage (4 V). The low side of the op amp's output voltage swing (0.185 V) is less than the ADC input voltage swing (0 V). The ADC input circuit is 20 kΩ and that doesn't load the op amp output stage, so the op amp output voltage swing is very close to the ADC input voltage range. R_{OUT} should present no problems acting as a voltage divider with the ADC input resistance. V_{OS} and I_{IB} create offset voltages that add to the reference offset voltage, and they have to be adjusted out as a group. The system noise overshadows the op amp noise, thus the op amp noise is accepted unless later calculation prove otherwise.

12.9 Amplifier Circuit Design

Enough information exists for the AIA to be designed. The TLV247X op amp is selected because it meets all the system requirements. The first step in the design is to determine the AIA input and output voltages, and this has already been done. These voltages are taken from Tables 12–1 and 12–2, and repeated here as Table 12–4.

Table 12–4. AIA Input and Output Voltages

INPUT VOLTAGE	OUTPUT VOLTAGES	
V_{IN1} = 650 mV	V_{OUT1} = 0 V	1st pair of data points
V_{IN2} = 400 mV	V_{OUT2} = 4 V	2nd pair of data points

The equation of an op amp is the equation of a straight line as given in Equation 12–14.

$$Y = mX + b \tag{12–14}$$

Two pairs of data points shown in Table 12–4 are substituted in Equation 12–14 making Equations 12–15 and 12–16.

$$4 = 0.4m + b \tag{12–15}$$

$$0 = 0.65m + b \tag{12–16}$$

Equation 12–15 is solved and substituted into Equation 12–16 to obtain Equation 12–17.

$$4 = 0.4\left(\frac{-b}{0.65}\right) + b \tag{12–17}$$

Solving Equation 12–17 yields b = 10.4, and solving Equation 12–15 yields m = –16. Substituting these values back into Equation 12–14 yields Equation 12–18, and Equation 12–18 (the final equation for the AIA) is put in electronic terminology.

$$V_{OUT} = -16V_{IN} + 10.4 \tag{12–18}$$

The circuit that yields the transfer function developed in Equation 12–18 is shown in Figure 12–13.

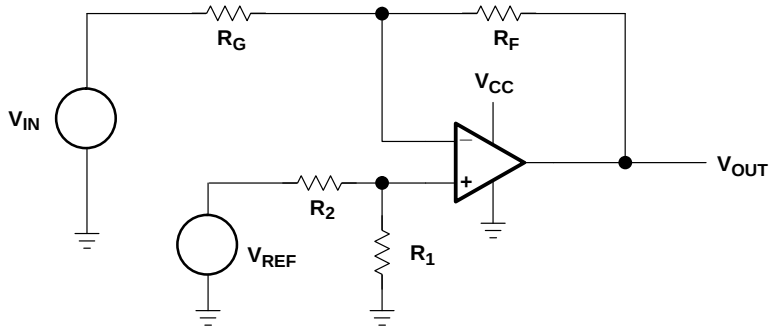


Figure 12–13. AIA Circuit

The equations for the AIA circuit are given below.

$$V_{OUT} = -V_{IN}\left(\frac{R_F}{R_G}\right) + V_{REF}\left(\frac{R_1}{R_1 + R_2}\right)\left(\frac{R_F + R_G}{R_G}\right) \tag{12–19}$$

$$|m| = \frac{R_F}{R_G} \tag{12–20}$$

$$b = V_{REF}\left(\frac{R_1}{R_1 + R_2}\right)\left(\frac{R_F + R_G}{R_G}\right) \tag{12–21}$$

Equation 12–18 gives the value for m as 16, and using Equation 12–20 yields $R_F = 16R_G$. Select R_F = 383 kΩ and R_G = 23.7 kΩ because they are standard 1% resistor values, and

this yields m = 16.16. The resistors R1 and R2 are calculated with the aid of Equations 12–22 and 12–23.

$$\frac{R_1}{R_1 + R_2} = \frac{b}{V_{REF}}\left(\frac{R_G}{R_F + R_G}\right) = \frac{10.4}{2.495}\left(\frac{23.7}{23.7 + 383}\right) = 0.2424 \qquad (12–22)$$

$$R_1 = \frac{0.2424}{0.7576} R_2 = 0.32 R_2 \qquad (12–23)$$

The parallel combination of R_1 and R_2 should equal the parallel combination of R_F and R_G so that the input voltage offset caused by the op amp input current is cancelled. Select R_2 = 105 kΩ and R_1 = 33.2 kΩ because they are standard 1% values, and then b = 10.3. The value of the parallel combination of R_1, R_2 ($R_1||R_2$ = 25.22 kΩ) almost matches the value of the parallel combination of R_F, R_G ($R_F||R_G$ = 22.3kΩ), and this is an adequate match for input current cancellation. The downsides of selecting large resistor values for R_F are current noise amplification, increased resistor noise, smaller bandwidth because of stray capacitance, and increased offset voltage due to input current. Bandwidth clearly is not a factor in this design. The op amp input current is 100 pA, so it won't cause much offset with a 383-kΩ feedback resistor (38.3 μV). The noise current and voltage are calculated later when the error budget is made.

The gain, m, and the intercept, b, are not accurate because the exact resistor values were not available in the 1%-resistor selection chart. This is a normal situation, and in less demanding designs the small error either does not matter or is corrected someplace else in the signal chain. That error is critical in this design, so it must eliminated. There are several nondrift type errors that have accumulated up to this point, and now is the time to correct all the nondrift errors with the addition of adjustments. Two adjustments are used; one adjustment controls the gain, m, and the other controls the intercept, b. The value of the adjustable resistor must be large enough to deliver an adequate adjustment range, but any value larger than that decreases the adjustment resolution.

The data that determines the adjustment range required is tabulated in Table 12–5. Drift and gain errors are calculated in volts, but drift errors are calculated in bits because they are not eliminated by adjustments. Remember, a LSB for this system is 4/4096 = 976.6 μV/bit.

Table 12–5. Offset and Gain Error Budget

ERROR PARAMETER	INTERCEPT	GAIN	DRIFT
V_{REF}	±25 mV		
V_{REF} drift			7.41 mV ≈ 8 LSB
Transducer offset		± 50 mV	
Transducer R_{OUT}			13 Ω ≈ 0 LSB
ADC reference		± 150 mV	1 LSB
Total unadjusted ADC error			2 LSB
Gain error		1.6 LSB	
ADC drift			4 LSB
V_{OS} op amp	2.2 mV		
I_B op amp	100 pA		
V_N op amp			$\frac{28\ nV}{\sqrt{Hz}}$ ≈ 1 LSB
I_N op amp			$\frac{139\ pA}{\sqrt{Hz}}$ ≈ 0 LSB
V_{NPS} PS noise			10 mV ≈ 2 LSB
R_{OUT} op amp			1.8 Ω ≈ 0 LSB
$V_{OUT\ LOW}$ op amp			70 mV ≈ 72 LSB
Total error			18 LSB

The adjustment for the intercept, b, depends on R_1, R_2, and V_{REF}. This adjustment has to account for the reference offset, the op amp input voltage offset, the op amp input current, and the resistor tolerances. The offset voltage inherent in the reference is given as ± 25 mV. The op amp input offset voltage is 2.2 mV; usually op amp offset voltage calculations include multiplying this offset by the closed loop gain, but this isn't done because the offset voltage is adjusted out in the input circuit. The op amp input current is converted to a common-mode voltage by the parallel combination of the reference resistors, so it is neglected in this calculation.

The worst case reference input voltage for the op amp, $V_{REF(MIN)}$, is calculated in Equation 12–24, where the resistor tolerances are assumed to be 3%, and the reference voltage error is 50 mV.

$$V_{REF(MIN)} = \left(V_{REF} - 50\ mV\right)\frac{0.97R_1}{0.97R_1 + 1.03R_2}$$
$$= (2.495 - 0.05)\frac{0.97(33.2)}{0.97(33.2) + 1.03(105)} = 0.566\ V \quad (12–24)$$

The nominal reference voltage at the op amp input is 0.6 V, so the reference voltage has to have about 40 mV adjustment around the nominal, or a total adjustment range of 80 mV. The nominal current through the voltage divider is $I_{DIVIDER}$ = (2.495/(105 + 33.2) kΩ = 0.018 mA. A 4444-kΩ resistor drops 80 mV, thus the adjustable resistor (a potentiometer) must be greater than 4444 kΩ. Select the adjustable resistor, R_{1A}, equal to 5 kΩ because this is an available potentiometer value, and the offset adjustment is ± 45 mV. Half of the potentiometer value is subtracted from R_1 to yield R_{1B}, and this subtraction centers the adjustment about the nominal value of 0.6V. R_{1B} = 33.2 kΩ – 2.5 kΩ = 30.7 kΩ. Select R_{1B} as 30.9 kΩ.

The adjustment for the gain employs R_F and R_G to insure that the gain can always be set at the value required to insure that the transducer output swing fills the ADC input range. The gain equation (Equation 12–18) is algebraically manipulated, worst case values are substituted for m and b, and it is presented as Equation 12–25.

$$G = \frac{V_{OUT} - 10.4}{-V_{IN}} = \frac{3.85 - 10.4}{-0.35} = 18.71 \quad (12–25)$$

Equation 12–26 is Equation 12–27 with 3% resistor tolerances.

$$0.97R_F = 18.71\left(1.03R_G\right) \quad (12–26)$$

Doing the arithmetic in Equation 12–26 yields R_F = 19.86 R_G. Thus, on the high side the gain must go from 16 to 19.86, or it must increase by 3.86. Assuming that the low side gain variation is equal, and rounding off to 4 sets the gain variation from 12 to 20. When R_G = 23.7 kΩ R_F varies from 284.4 kΩ to 474 kΩ. R_F is divided into a potentiometer R_{FA} = 200 kΩ and R_{FB} = 280 kΩ, thus the nominal gain can be varied from 11.8 to 20.2.

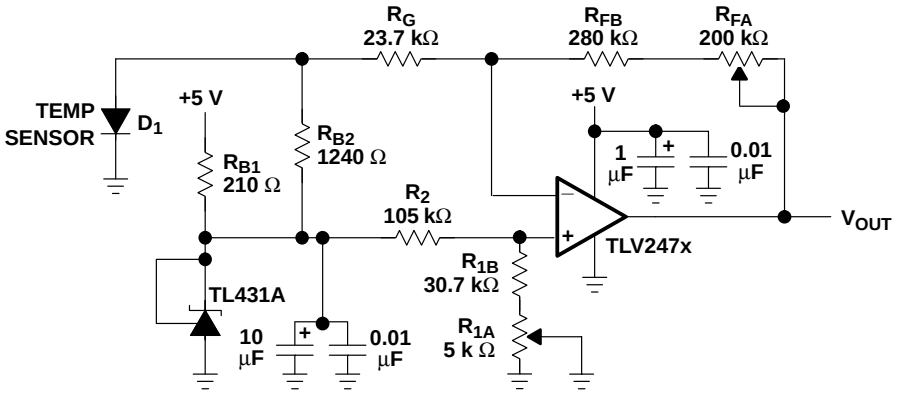


Figure 12–14. Final Analog Interface Circuit

There is no easy method of setting two interacting adjustments because when the gain is changed the offset voltage changes. They quickest method of adjustment is to connect the transducer to the circuit, adjust the offset, and then adjust the gain. It takes several series of adjustments to get to the point where the both parameters are set correctly.

The impedance and noise errors are calculated prior to completing the error budget. The op amp input impedance works against the transducer output impedance to act like a voltage divider. The value of the voltage divider is calculated in Equation 12–27, and as Equation 12–27 indicates, the output resistance of the transducer is negligible compared to the input resistance of the op amp. This is not always the case!

$$V_D = V_T \frac{R_{IN}}{r_c + R_{IN}} = V_T \frac{10^6}{13 + 10^6} \approx V_T \qquad (12\text{–}27)$$

The ADC input impedance works against the op amp output impedance to act like a voltage divider. The value of the voltage divider is calculated in Equation 12–28. The voltage divider action introduces about a 0.009% error into the system, and this is within 13-bit accuracy, so it can be neglected.

$$V_{IN} = V_{OUT} \frac{20(10^3)}{1.8 + 20(10^3)} = \frac{20}{20.0018} = 0.99991\ V_{OUT} \qquad (12\text{–}28)$$

The noise specification is given in nV/(Hz$^{0.5}$), and this must be converted to volts. There are involved formulas for the conversion, but the simplest thing to do is assume the noise is wide band. If the numbers add up to a significant error, detail calculations have to be made. The voltage noise is multiplied by the closed loop gain, thus $V_{NWB} = V_N\ (G_{MAX})$ = 28 nV(20) = 560 nV = 0.56 μV. The current noise is multiplied by the parallel combination of R_F and R_G, thus I_{NWB} $I_N(R_F||R_G)$ = 139 pA(22.5 kΩ) = 3.137 nV. The system noise is 10 mV, and this noise comes in through the inputs and the power supply. The power sup-

ply contribution is reduced by the power supply rejection ratio, and it is 10 mV/63 dB = 10 mV/1412 = 7.08 nV. This calculation assumes that high-frequency noise is not a problem, but if this is not true, CMRR must be reduced per the data sheet CMRR versus frequency curves.

Some of the system noise propagates through the inputs and is rejected by the common-mode rejection of the op amp. The op amp is not configured as a differential amplifier, so a portion of the closed loop gain will multiply some of the system noise.

The ac gain of the AIA is given in Equation 12–29.

$$V_{OUT} = \alpha V_{SN} \left(\frac{R_1}{R_1 + R_2}\right)\left(\frac{R_F}{R_F + R_G}\right) - \alpha V_{SN} \frac{R_F}{R_G} \qquad (12\text{–}29)$$

$$= \alpha V_{SN} (4.12 - 16) = 11.8\, \alpha V_{SN}$$

All of the system noise does not get in on the inputs, rather most of the system noise is found on the power supply. The fraction of the system noise that gets into the ground system and onto the op amp inputs is very small. This fraction, α, is normally about 0.01 because the power supplies are heavily decoupled to localize the noise. Considering this, the system noise is 1.18 mV, or less than 2 LSBs.

The op amp output voltage range does not include 0 volts, and the ADC output voltage low value is 0 volts, so this introduces another error. The guaranteed op amp low voltage is 185 mV at a load current of 2.5 mA. The output current in this design is 185 mV/20 kΩ = 9.25 μA. This output current approximates a no load condition, hence the nominal low voltage typical specification of 70 mV is used. This leads to a 72 LSB error, by far the biggest error.

Referring to Table 12–5, notice that the total error is 90 LSB. Losing 90 LSBs out of 4096 total LSBs is approximately 11.97 bits accurate, so the 11-bit specification is met. The final circuit is shown in Figure 12–14.

Notice that large decoupling capacitors have been added to the power supply and reference voltage. The decoupling capacitors localize IC noise, prevent interaction between circuits, and help keep noise from propagating. Two decoupling capacitors are used, a large electrolytic for medium and low frequencies, and a ceramic for high frequencies. Although this portion of the design is low frequency, the op amp has a good frequency response, and the decoupling capacitors prevent local oscillations through the power lines. If cable noise is a problem, an integrating capacitor can be put in parallel with R_F to form a low-pass filter.

12.10 Test

The final circuit is ready to build and test. The testing must include every possible combination of transducer input and ADC output to determine that the AIA functions in all manufacturing situations. The span of the adjustments, op amp output voltage range, and ADC input range must be checked for conformance to the design criteria. After the design has been tested for the specification limits it should be tested for user abuse. What happens when the power supply is ramped up, turned on instantly, or something between these two limits? What happens when the inputs are subjected to over voltage, or when the polarity is reversed? These are a few ideas to guide your testing.

12.11 Summary

The systems engineers select the transducer and ADC, and their selection criterion is foreordained by the application requirements. The AIA design engineer must accept the selected transducer and ADC, and it is the AIA designer's job to make these parts play together with adequate accuracy. The AIA design often includes the design of peripheral circuits like transducer excitation circuits, and references.

The design procedure starts with an analysis of the transducer and ADC. The analysis is followed by a characterization of the transducer and reference. At this point enough information is available to make an error budget and select candidate op amps. The op amp is selected in the next step in the procedure, and the circuit design follows.

The output voltage span of the transducer and corresponding input voltage span of the ADC are coupled as two pairs of data points that form the equation of a straight line. The data point pairs are substituted into simultaneous equations, and the equations are solved to determine the slope and intercept of a straight line (an op amp solution). The op amp circuit configuration is selected based on the sign of the slope and the intercept. Finally, the passive components used in the op amp circuit are calculated with the aid of the op amp circuit design equations.

The final circuit must be tested for conformance to the system specifications, but the prudent engineer tests beyond these specifications to determine the AIA's true limits.

12.12 References

1 Wobschall, Darold, *Circuit Design for Electronic Instrumentation*, McGraw-Hill Book Company, 1979

Chapter 13

Wireless Communication: Signal Conditioning for IF Sampling

Perry Miller and Richard Cesari

13.1 Introduction

High-speed operational amplifiers (op amps) are used extensively in wireless communication systems. These amplifiers typically operate at intermediate frequencies (IF) ≤ 500 MHz and most frequently operate below 25 MHz. Applications for high-speed op amps include filtering circuits in radio receivers, IF amplifiers, mixer circuits, and bandpass amplifiers.

13.2 Wireless Systems

This chapter focuses on the requirements for the op amp and a number of techniques used in wireless communication systems to interface high-speed op amps to analog-to-digital converters (ADCs) and digital-to-analog converters (DACs). This section provides several examples of different op amp usage.

Figure 13–1 shows an example of a dual-IF receiver. In this application, several stages with different IF frequencies are used to get the desired performance. The receiver converts the received radio frequency (RF) input from the antenna to a baseband signal. This type of system requires the ability to receive and operate over a wide range of signal strength. The inherent system noise level determines the lower operating limit and is a critical factor in the overall performance of the receiver. The receiver performance is measured in terms of receiver sensitivity, which is defined as the ratio between the power of the wanted baseband signal at the output of the ADC and the total power of all unwanted signals (include random noise, aliasing, distortion, and phase noise contributed by the local oscillator) introduced by the different circuit elements in the receiver. A low-sensitivity receiver can cause signal saturation in the ADC input.

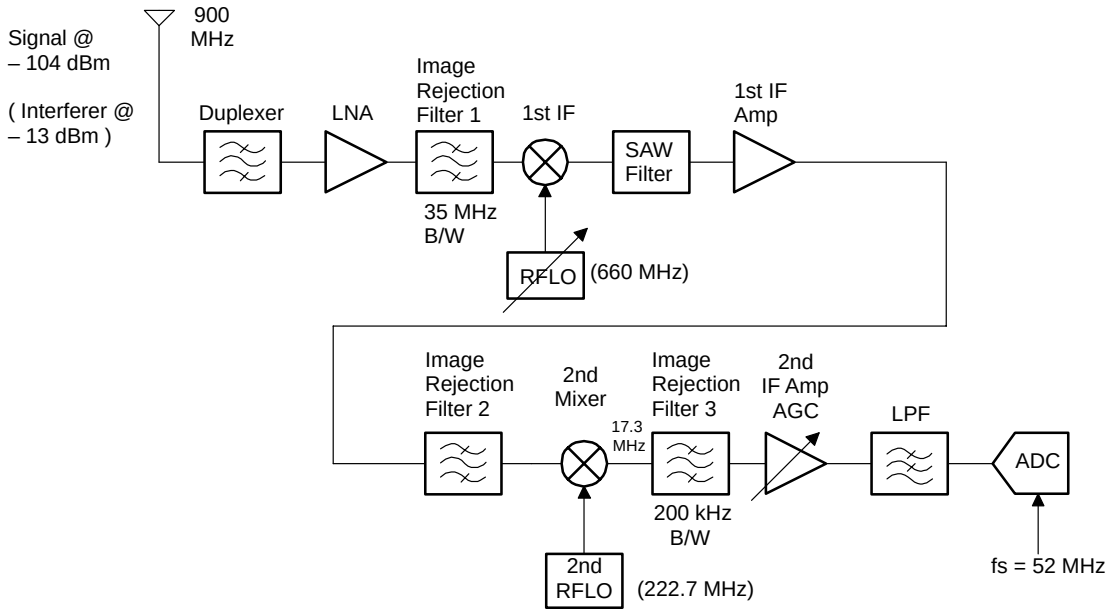


Figure 13–1. A Typical GSM Cellular Base Station Receiver Block Diagram

The receiver contains two mixer stages reminiscent of a classic superhetrodyne receiver with good selectivity. The process of hetrodyning involves the translation of one frequency to another by the use of a mixer and local oscillator (LO) offset at the proper frequency to convert the RF signal to the desired IF. The LO signal is at a much higher level than the RF signal. In Figure 13–1, the 900 MHz RF signal is picked up by the antenna and amplified by a low-noise amplifier (LNA). After being sufficiently amplified by the LNA to overcome the noise level, the RF signal passes through a bandpass filter (BPF) used to provide image rejection and sufficient selectivity prior to the first stage mixing.

High selectivity prevents adjacent channel energy from getting into the input of the ADC and decreasing the receiver dynamic range. A strong signal in an adjacent channel causes intermodulation products in the receiver that can result in loss of the received signal. The band-pass filter is implemented with a surface acoustic wave (SAW) filter. The SAW filter provides very sharp edges to the passband, with minimum ripple and phase distortion.

The first stage mixer down-converts the band-limited RF signal with the LO signal, producing a number of new frequencies in the spectrum, including the sum frequency component, the difference frequency component, and spurious responses. The first stage IF filter provides sufficient filtering after the mixing circuit. It selects the difference frequency component while rejecting the sum frequency component and undesirable spurious responses. Passing the difference frequency component on to the next stage of the receiver makes it much easier to provide the gain and filtering needed for proper receiver function-

ality. Image rejection also places a constraint on the choice of the IF (10 MHz – 20 MHz). The spurious responses are the result of power supply harmonics and intermodulation products created during the mixing of the RF signal and the LO signal. If not substantially suppressed, spurious responses often corrupt the IF signal and cause it to be accepted as a valid IF signal by the IF amplifier.

The first stage IF amplifier minimizes the effects of the first stage filter loss on the noise figure and amplifies the signal to a suitable level for the second stage mixer. The output from the second stage mixer is applied to the second stage IF amplifier, automatic gain control (AGC) amplifier, and the subsequent low-pass filter, producing a 1-V full-scale input to the ADC. The ADC samples and digitizes this baseband analog input. The AGC amplifier ensures that if the received signal amplitude goes up rapidly, the ADC is not saturated. At the other extreme, if there is a fast power ramp down, the AGC prevents the signal quality from passing below an acceptable level. High-speed current-feedback operational amplifiers (CFA) are typically used to filter and amplify the IF signals because this type of operational amplifier has good slew rate, wide bandwidth, large dynamic range, and a low noise figure.

In this type of receiver, the ADC is a key component requiring sampling rates ≥ 40 MSPS with 12 bits to 14 bits of resolution and is usually a pipeline architecture device. The output of the ADC is highly dependent on the ADC's sampling frequency, nonlinearities in the ADC and the analog input signal, and the converter maximum frequency.

Table 13–1 tabulates the contribution of each stage depicted in Figure 13–1 to the system level budget for a typical GSM receiver. GSM is the global system for mobile communications. It is one of the most popular digital cellular formats in the world.

Table 13–1. GSM Receiver Block System Budget

ELEMENT	NOISE FIGURE (dB)	GAIN (dB)	ANF†
Duplexer	1	–1	1
LNA	1.6	+18	0.51616
Image rejection Filter #1		–2	0.00517
1st stage mixer	9.87	–7	0.21853
Noise Filter		–2	0.07363
1st stage amplifier	7.92	+49	1.7957
Image rejection filter #2		–2	4.31E–06
2nd stage mixer	10.8	–7	5.68E–05
2nd stage image filter		–2	0.0009
AGC	11	+50	2.16E–09
Anti-alias filter		–2	3.83E–10
ADC	7.63	–2	3.83E–10
Total			3.61 **(5.7 dB)**

† ANF = Adjusted noise figure (linear).

Figure 13–2 shows a more flexible implementation for the receiver using a digital signal processor (DSP). Using a DSP allows a single receiver to access several different wireless systems through changes in software configuration.

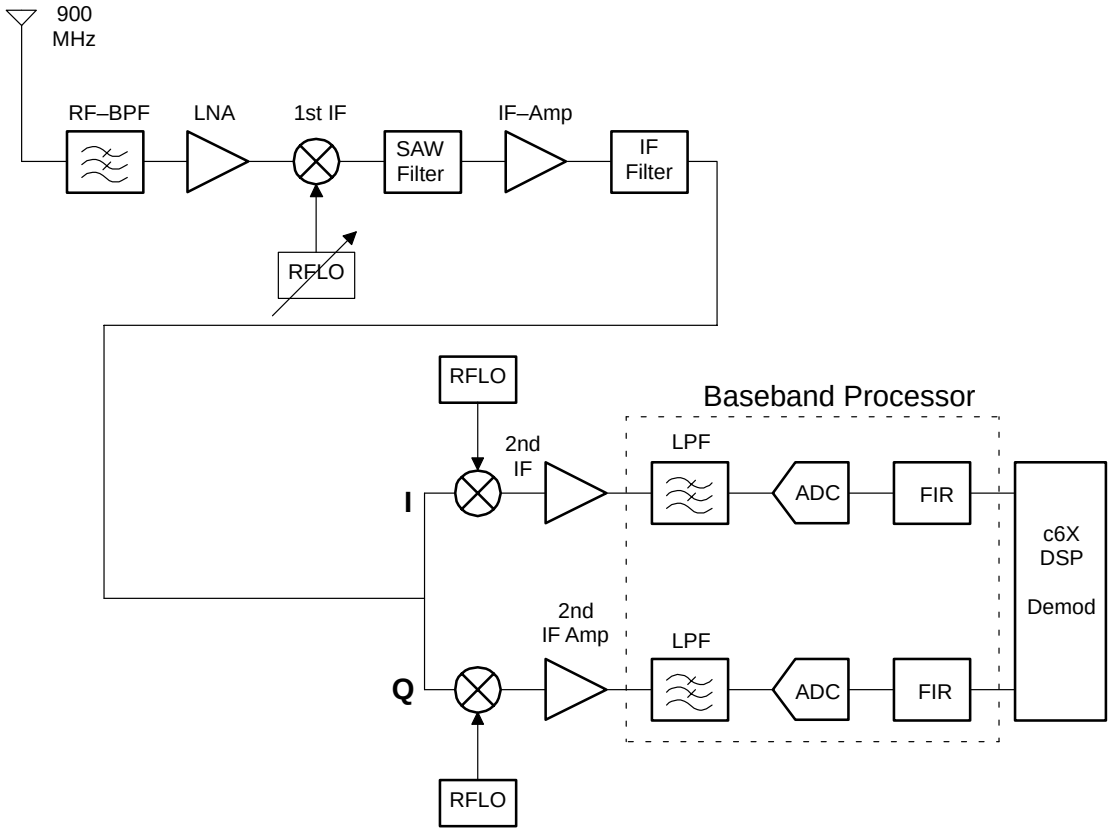


Figure 13–2. An Implementation of a Software-Configurable Dual-IF Receiver

Figure 13–3 shows a basic W-CDMA transmit chain. A voiceband CODEC (coder–decoder), op amps, and a DSP are used to digitize and band-limit the audio signal. The digitized signal is then compressed to the appropriate data rate either in hardware or by a software program implemented on the DSP. Redundancy (error correction), encryption, and the appropriate form of modulation (QPSK for W–CDMA or GMSK for GSM) are added to the compressed digitized signal. This signal goes via an interpolating filter to the communication DAC as shown in Figure 13–4. Eight times interpolation[1],[2] is shown in Figure 13–4, but other multiple-of-2 interpolations are possible and quite often used. Assuming that the modulated bit stream is a 3.84 MSPS W–CDMA signal, for 8x interpolation, the sampling clock frequency would need to be 30.72 MHz. The DAC converts the modulated bit stream to analog, and the conversion is usually performed by a pair of DACs: one for I channel and one for Q channel — see Figure 13–3. The reconstruction filter, at the output

of the DAC, is usually a high-order Bessel or elliptic filter used to low-pass filter the analog output from the DAC.

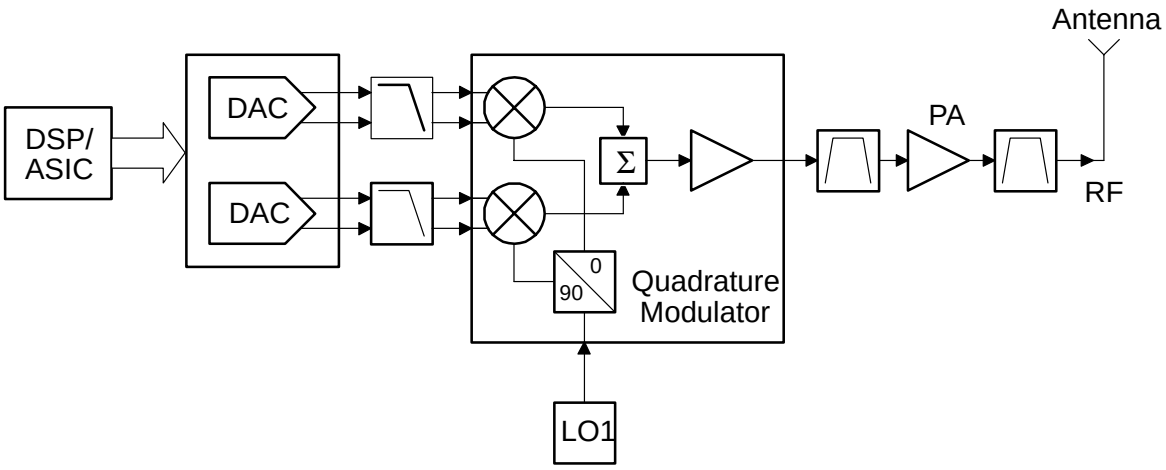


Figure 13–3. Basic W-CDMA Cellular Base Station Transmitter Block

The modulator block converts the baseband I and Q signal to the appropriate carrier frequency, typically 864 MHz. The up-converted 864 MHz signal is amplified to a suitable level by the power amplifier (PA) and sent out via the antenna over the air or to a nearby wireless base station.

The RF power amplifier is a large-signal device with power gain and efficiency on the order of 50% for GSM and about 30% for code division multiple access (CDMA), an access method in which multiple users are permitted simultaneously on the same frequency.

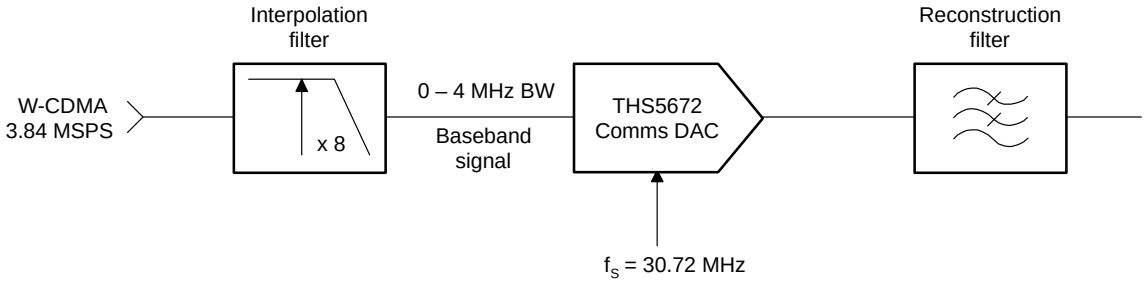


Figure 13–4. Communication DAC with Interpolation and Reconstruction Filters

13.3 Selection of ADCs/DACs

In communication applications, the dc nonlinearity specifications that describe the converter's static performance are less important than the dynamic performance of the ADC. The receiver (overall system) specifications depend very much on the ADC dynamic performance parameters: effective number of bits (ENOB), SFDR (spurious free dynamic range), THD (total harmonic distortion), and SNR (signal-to-noise ratio). Good dynamic performance and fast sampling rate are required for accurate conversion of the baseband analog signal at RF or IF frequencies. The SFDR specification describes the converter's in-band harmonic characterization and it represents the converter's dynamic range. SFDR is slew rate and converter input frequency dependent.

The output from an ADC is highly dependent on the converter sampling frequency and the maximum frequency of the analog input signal. A low-pass or band-pass anti-aliasing filter placed immediately before the ADC band-limits the analog input. Band-limiting ensures that the original input signal can be reconstructed exactly from the ADC's output samples when a sampling frequency (f_S) of twice the information bandwidth of the analog input signal is used (Nyquist sampling). Undesirable signals, above f_S/2, of a sufficient level, can create spectrum overlap and add distortion to the desired baseband signal. This must not be allowed to dominate the distortion caused by ADC nonlinearities. Sampling at the Nyquist rate places stringent requirements on the anti-aliasing filter — usually a steep transition 10^{th} or higher order filter is needed.

Oversampling techniques (sampling rate greater than the Nyquist rate) can be employed to drastically reduced the steepness of the anti-aliasing filter rolloff and simplify the filter design. However, whenever oversampling is used, a faster ADC is required to digitize the input signal. Very fast ADCs can be costly and they consume a fair amount of power (≥ 1000 mW). In a system application, such as a wireless base station, where large numbers of ADCs are used, the individual device power consumption must be kept to the bare minimum (≤ 400 mW). High-resolution ADCs, with slower sampling rates, offer potential cost savings, lower power consumption, and good performance, and are often used in some applications. In this case, undersampling or bandpass sampling techniques (the analog signal digitization by the ADC exceeds half the sampling frequency (f_S) of the ADC, but the signal information bandwidth is ≤ f_S/2) are employed.

Operating the ADC in a bandpass sampling application requires knowledge of the converter's dynamic performance for frequencies above f_S/2. In general, as the input signal frequency to the converter increases, ENOB, SNR, SFDR, and harmonic performance degrades.

The fact that the analog input to the ADC cannot be represented exactly with a limited number of discrete amplitude levels introduces quantization error into the output digital samples. This error is given by the rms quantization error voltage $e^2_{qns} = \frac{1}{12} q_s^2$.

The mean squared quantization noise power is is $P_{qn} = \frac{q_s^2}{12R}$ where q_s is the quantization step size and R is the ADC input resistance, typically 600 Ω to 1000 Ω.

Communication ADCs similar to the THS1052 and THS1265 typically have a full scale range (FSR) of 1 Vp–p to 2 Vp–p. Generally, wireless systems are based on a 50-Ω input/output termination, therefore, the ADC input is made to look like 50 Ω. Based on this assumption, the quantization noise power for a 12-bit, 65 MSPS ADC (THS1265) is –73.04 dBm.

For a noise-limited receiver, the receiver noise power can be computed as the thermal noise power in the given receiver bandwidth plus the receiver noise figure NF[3].

For 200 kHz BW (GSM channel), temperature 25°C, and 4 dB to 6 dB NF, the receiver noise power is –115 dBm. Therefore, to boost the receiver noise to the quantization noise power level requires a gain of 42 dB.

In Figure 13–1, the GSM–900 signal is at –104 dBm (GSM–900 spec for smallest possible signal at which the raw bit-error rate must meet or exceed 1%) and, therefore, the signal-to-noise ratio (SNR) at baseband or at the converter and due to the thermal noise component is given by $SNR_{thermal} = \frac{E_b}{N_0}$ = –104 dBm +115 dBm = 9 dB.

In order for the raw BER to be 1% in a GSM system, testing and standard curves [4] indicate, that a baseband SNR (derived from the sum of both thermal noise and ADC noise) of 9 dB is needed for this performance node.

The process gain G_p is defined as:

$$G_p = \frac{f_s}{BW} = \frac{52 \times 10^6}{200 \times 10^3} = 2.6 \times 10^2 = 24.15 \text{ dB} \qquad (13–1)$$

where GSM channel BW = 200 kHz and f_s = 52 MHz (the ADC sampling frequency).

The converter noise at baseband should be much better than the radio noise (= thermal noise + process gain). Furthermore, the thermal noise alone brings the system only to the reference bit-error rate (BER).

Therefore the converter noise (at baseband) = SNR_{adc} + process gain G_p.

The ADC SNR_{adc} should be 20 dB to 40 dB above the $SNR_{thermal}$ of the thermal noise component (+9 dBm). In this example, ADC SNR_{adc} is selected to be 37 dB better than the $SNR_{thermal}$ of the thermal noise component (9 dB).

In other words, if the converter SNR_{adc} is desired to be 37 dB better than the thermal noise component ($SNR_{thermal}$), the baseband converter is chosen to be 9 + 37 dB = 46 dB.

The total noise (N_{sum}) = thermal noise (N_t) + converter noise (N_{conv}).

Therefore, the noise-to-signal ratio is:

$$\frac{N_{sum}}{S_{GSM}} = \frac{N_t}{S_{GSM}} + \frac{N_{conv}}{S_{GSM}} = \frac{1}{10^{0.9}} + \frac{1}{10^{4.6}} \quad (13\text{–}2)$$

The signal-to-noise ratio $\frac{S_{GSM}}{N_{sum}}$ = 7.942 (linear), which is 8.999 dB.

This shows that the converter noise degenerates the baseband SNR due to thermal noise alone, by only 0.0001 dB (9.000 dBm – 8.999 dB) when the signal is at reference sensitivity level.

At f_s = 52 MSPS, the converter SNR_{adc} required to fit the GSM–900 signal is (46 – 24.15) dB = 22 dB.

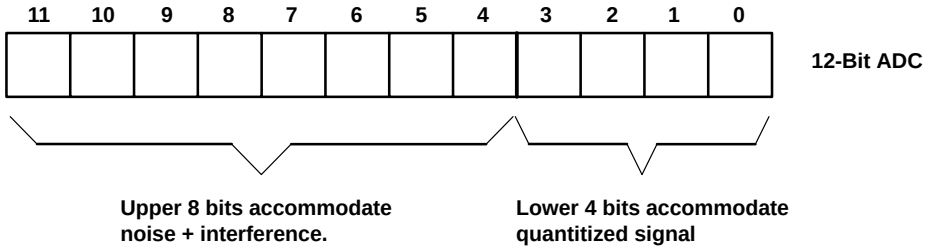
The effective number of bits (ENOB) required for fitting the GSM–900 signal is

$$ENOB = \frac{SNR - 1.76}{6.02} \quad (13\text{–}3)$$

and thus 4 bits are needed for the GSM–900 signal.

Assuming that filter #3 attenuates the interferer by 50 dB, the interferer drops from –13 dBm to –53 dBm, or 40 dB above the GSM signal. The requirements for the number of bits needed to accommodate the interferer is $\frac{40\ dB}{6\ dB/bit}$ = 6.3 bits.

Approximately 6 bits are needed to accommodate the interferer, plus 2 bits of head room for constructive interference, for a total of 8 bits. The ADC requirements are 4 bits for the GSM signal plus 8 bits for the interferer for a total of 12 bits:



It follows from the above analysis that the GSM signal is 8 bits down, or 8 bits x 6 dB/bit = 48 dB down from the 1 V FSR of the ADC. The full-scale input power to an ADC having a 50-Ω termination can be calculated as:

$$\frac{V^2}{R} = \frac{(1)^2}{50} = 2 \times 10^{-2} = 13\ dBm\ (full\ scale\ or\ FSR) \quad (13\text{–}4)$$

Therefore, the 4 bits for the GSM signal gives 13 dBm – (8 bits x 6 dB/bit) = –35 dBm, the smallest possible ADC input signal power with the interferer that will meet the GSM speci-

fication. Without the interferer present, the signal level is chosen to be about 20 dB below the ADC full scale, or –8 dBm, to accommodate constructive interference and to accommodate any large ADC input signal that may arise from short-term errant gain due to gain settling in the AGC. Thus, with the smallest possible signal specified by the GSM–900 spec, the signal is amplified from –104 dBm to –8dBm without the interferer, and to –35 dBm with the interferer.

For a practical receiver, as shown in Figure 13–1, an AGC is necessary to assure that the LSB represents a uniform noise input while the peak power does not exceed the ADC's FSR.

The receiver block shown in Figure 13–1 uses a high-speed, fairly wide bandwidth (100 MHz to 550 MHz) communication ADC to convert the baseband signal to a high-speed parallel bit stream for processing in a DSP. For an ADC to accurately produce a digital version of the baseband analog input, the device must have very good resolution and dynamic performance.

The signal path shown in Figure 13–1 needs 95 dB of gain in order to bring the –104 dBm GSM signal up to –9 dBm (equivalent to about 0.112 Vp–p across 50 Ω), with allowances for losses in the filters, and mixers. Usually this gain is split evenly between the RF and baseband, but baseband gain is less expensive and consumes less power. The LNA provides 18 dB, which, after filtering and cable losses, yields about 16 dB, while the mixer provides –7 dB of conversion gain/loss.

Modern communication DACs are, effectively, an array of matched current sources optimized for frequency domain performance. To handle both strong and weak signals, communication DACs require large dynamic range. The dynamic specifications of most interest are SFDR, SNR, THD, IMD (two-tone intermodulation distortion), ACPR (adjacent channel power rejection), and settling time.

Besides these, there are a number of dc parameters, such as integral nonlinearity (INL) and differential nonlinearity (DNL), that are considered important because of their influence on the SFDR parameter. DNL errors occur only at certain points in the converter's transfer function. INL and DNL errors appear as spurious components in the output spectrum and can degrade the signal-to-noise ratio of the DAC.

Typical SFDR figures for 12-bit DACs and 14-bit DACs, with a 5-MHz single tone input at 50 MSPS, range from 75 dB to 80 dB. In order to prevent adjacent communication channels from interfering with each other, the DAC must exhibit a good SFDR specification. Communication DACs normally have differential outputs, and the current-mode architecture is used to give the DAC a higher update rate.

13.4 Factors Influencing the Choice of Op Amps

IF amplifiers and filters can be built from discrete components, though most modern applications use integrated circuits. High-speed wideband op amps are employed as buffer amplifiers in the LO circuit, at the front end of ADCs, at the output of the DAC, in the external voltage reference circuits for ADCs and DACs, and in the AGC amplifier and anti-aliasing stage. Op amps operating at IF frequencies, such as the AGC amplifier in Figure 13–1, must attain a large gain control range. How well the amplifier handles large and small signals is a measure of its dynamic range. The current-feedback op amp can be used everywhere except for the anti-aliasing filter and in the reconstruction filter stage. The op amp must have a level gain response from almost dc to at least 500 MHz, after which a gentle rolloff is acceptable. Also, the phase response is important to avoid dispersing the signal — this requires a linear phase response.

Several factors influence the choice of the current-feedback op amp (CFA) and voltage-feedback amplifier (VFA) for use in wireless communication systems:

- The ADC/DAC resolution
- ADC/DAC dynamic specification
- Operating frequencies
- Type of signal
- Supply voltages and
- Cost

In both the receiver and transmit circuits, shown in Figures 13–1 and 13–3, the SFDR and IMD are the key ADC/DAC parameters that have the most influence on op amp selection. A minimum requirement is that the op amp's SFDR or THD parameter, measured at the frequency of operation, should be 5 dB to 10 dB better than the converter's SFDR. For a perfect 12-bit ADC, the SFDR is 72 dB, thus the op amp in front of the ADC should exhibit a SFDR (or THD) of 77 dB to 82 dB.

When an op amp is used as a buffer amplifier, it must faithfully reproduce the input to a very high degree of accuracy. This requires that the amplifier be designed and optimized for settling time. Fast settling time is mandatory when driving the analog input of an ADC because the op amp output must settle to within 1 LSB of its final value (within a time period set by the sampling rate) before the ADC can accurately digitize the analog input. The amplifier settling time determines the maximum data transfer rate for a given accuracy. For example, to settle within 1 LSB of full scale range implies that the settling accuracy of the ADC is ± ½ LSB. Hence, a 12-bit ADC will require the op amp to settle to $\frac{1}{2 \times 2^{12}} = 1.22 \times 10^{-4}$ of final value, or 0.0122% of final value. An LSB = 244 μV for a 12-bit ADC with 1 V full-scale range. Values for the settling time and other important op amp parameters as they relate to the receiver and transmit blocks are listed in Table 13–2.

The op amp dynamic parameters in Table 13–2 represent the range of values to achieve low noise, good SFDR, high slew rate, good bandwidth, etc.

Table 13–2. High-Speed Op Amp Requirements

PARAMETER	VALUE
Noise voltage	2.7 nV/$\sqrt{\text{Hz}}$ to 8 nV/$\sqrt{\text{Hz}}$
Noise current	1 pA/$\sqrt{\text{Hz}}$ to 30 pA/$\sqrt{\text{Hz}}$
THD	70 dBc to 95 dBc
Slew rate	260 V/μV to 3500 V/μV
Small signal bandwidth	200 MHz to 600 MHz
Large signal bandwidth	≥ 100 MHz
Common-mode input voltage	3 V
Supply voltage	± 5 V
Settling time	8 ns to 20 ns
Output current	40 mA to 100 mA
Output impedance	≤ 20 Ω
PSRR	– 60 dB
CMRR	–70 dB
Input offset voltage	10 mV (typical)

Op amps operating from ± 5-V supplies typically have 6 V to 8 V of common-mode range. Single-supply op amps often handle much smaller voltage ranges, and in some communication applications, could exhibit limited linear operation over a wide signal swing. With the exception of rail-to-rail op amps, most op amps can swing to within 1 V to 1.5 V of the positive rail. Typically signal-to-noise ratio, slew rate, and bandwidth suffer for devices operating from low supply voltages.

When selecting current-feedback op amps, the gain-bandwidth plots are essential. They are needed because with current feedback ordinary loop-gain-proportional bandwidth relationships do not hold.

13.5 Anti-Aliasing Filters

Spurious effects in the receiver channel (Figure 13–1) appear as high frequency noise in the baseband signal present at the ADC. The spurious signals (> f_S/ 2) must be blocked from getting to the ADC (sampling at Nyquist rate, f_S) where they will cause aliasing errors in the ADC output.

A suitable anti-aliasing low-pass analog filter placed immediately before the ADC can block all frequency components capable of causing aliasing from reaching the ADC. The

anti-aliasing filter cutoff frequency ($f_{c)}$ is set to the highest baseband signal frequency of interest (f_{max}) so that $f_c = f_{max}$. Sampling theorem requires that the ADC minimum Nyquist rate sampling frequency $fs = 2f_{max}$. This ensures that the original base band or IF signal can be reconstructed exactly from the ADC's digital outputs. It is important to know that only an anti-aliasing filter having a *brickwall*-type response could fully satisfy the exacting requirements imposed by the sampling theorem. The rolloff of real filters increases more gradually from cutoff to the stop band, and therefore, in practice, the ADC sampling frequency is usually slightly higher than $2f_{max}$.

The anti-aliasing filter must reduce the out-of-band aliasing producing signals to less than 1 LSB of the ADC resolution, without introducing additional distortion of the baseband or IF signal in-band components and without predominating distortion due to the ADC nonlinearities. The spectrum overlap (aliasing) requirements are determined by:

- Highest frequency of interest
- Sampling rate
- ADC resolution

The highest signal frequency of interest sets the filter cutoff frequency. For example, suppose the input signal is to be sampled to 12-bit accuracy with a sampling frequency of 52 MHz. If the IF signal is 17 MHz an 18 MHz filter –3dB cutoff frequency could be chosen. All frequencies above the Nyquist frequency should be attenuated to $\leq$ ½ LSB, but generally only frequencies above the ADC's limit of resolution will be a problem; i.e., f_{alias} = (52 – 17) MHz = 35 MHz. The frequency rolloff is 18 MHz to 35 MHz (about 1 octave) and the required attenuation is 72 dB (12-bit ADC). A very high-order filter is required to accomplish this task. Practical ant-aliasing filters are limited to fifth-order or sixth-order type because of amplifier bandwidth, phase margin, layout parasitics, supply voltage, and component tolerances. Keep in mind that as the rolloff sharpens, the passband ripple and phase distortion increase.

For communication applications, linear phase characteristic and gain accuracy (low passband ripple) are important. And normally, Chebychev or elliptic (Cauer) filter types are used for the anti-aliasing filter.

For good transient response or to preserve a high degree of phase coherence in complex signals, the filter must be of linear-phase type (Bessel-type filter).

The THS4011 or THS4021 voltage-feedback op amp is a good choice for implementing anti-aliasing filter in this example.

The quality of the capacitors and resistors used to implement the design is critical for performance anti-aliasing filter.

13.6 Communication D/A Converter Reconstruction Filter

Modern communication DACs are, effectively, an array of matched current sources optimized for frequency domain performance. The most important dynamic specifications are SFDR, SNR, THD, IMD, ACPR and settling time. The dc parameters INL and DNL are considered important because of their influence on the SFDR parameter. Typical SFDR figures for 12-bit to 14-bit DACs, with a 5-MHz single-tone input at 50 MSPS, ranges from 75 dB to 80 dB. In order to prevent adjacent communication channels from interfering with each other, the DAC must exhibit a good SFDR specification.

Communication DACs normally have differential outputs and current-mode architecture is used to give the DAC a higher update rate.

Figure 13–4 shows an interpolating filter block before the DAC and the reconstruction analog filter at the output of the DAC. The interpolating filter is a digital filter whose system clock frequency is an integer multiple of the filter input data stream and is employed to reduce the DAC's in-band aliased images. This eases the job of the reconstruction filter at the output of the DAC. The filter is used to smooth the input data — the output waveform frequency is the same as the input to the interpolating filter. Figure 13–5 shows the interpolation filter output. The system clock frequency of the DAC and the interpolating filter are running at the same rate; therefore, the frequency spectrum of the DAC output signal, repeated at integer multiples of the sampling rate, becomes increasingly separated as the sampling rate is increased. The further apart the repeated DAC output spectrum, then the less steep the attenuation characteristic of the anti-aliasing needs to be. Consequently, a simpler anti-aliasing filter with a less steep rolloff from pass band to stop band can be used without any increase in distortion due to aliasing. In Figure 13–4, the system clock is 30.772 MHz (3.84 MSPS x 8). Figure 13–6 shows the attenuation characteristics needed for the reconstruction anti-aliasing filter.

The aliasing frequency is f_{alias} = (28.7 – 2) = 26.7 MHz, and the required attenuation is 84 dB (14-bit DAC). A third-order anti-aliasing low-pass elliptic or Bessel filter could be used to meet the attenuation requirements. Either type of high-order filter gives a relatively flat response up to just below ½ the sampling frequency, followed by a sharp cutoff. But this arrangement provides no correction for the sinc function (Sinx)/x falloff in amplitude naturally produced by the sample and hold function in the DAC.

The tradeoff in building a simpler reconstruction filter is that a faster DAC is required to convert the input digital data stream to analog signal.

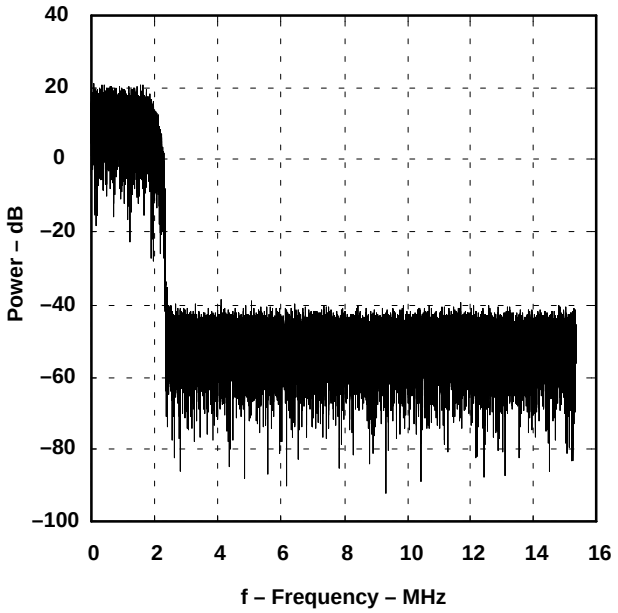


Figure 13–5. QPSK Power Spectral Density Without Raised Cosine Filter — W-CDMA

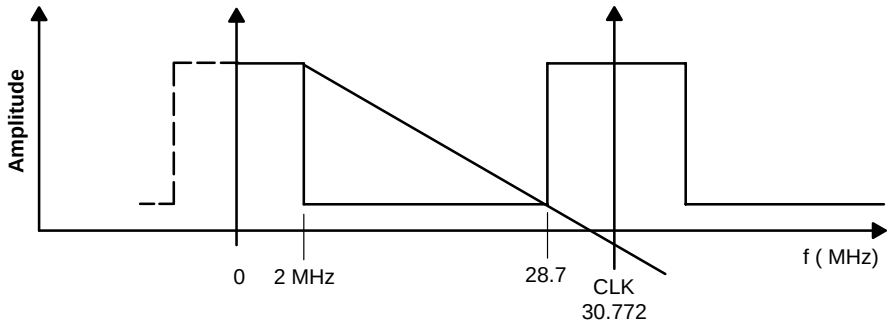


Figure 13–6. Reconstruction Filter Characteristics

Figure 13–7 shows a first-order reconstruction (low-pass) filter consisting of a high-speed differential amplifier configured for unity gain. The DAC outputs are terminated into 50 Ω. In Figure 13–7, the value for the filter capacitor is given by the expression:

$$C_{fb} = \frac{3.1827 \times 10^{-4}}{\frac{f_{clk}}{2}} \tag{13–5}$$

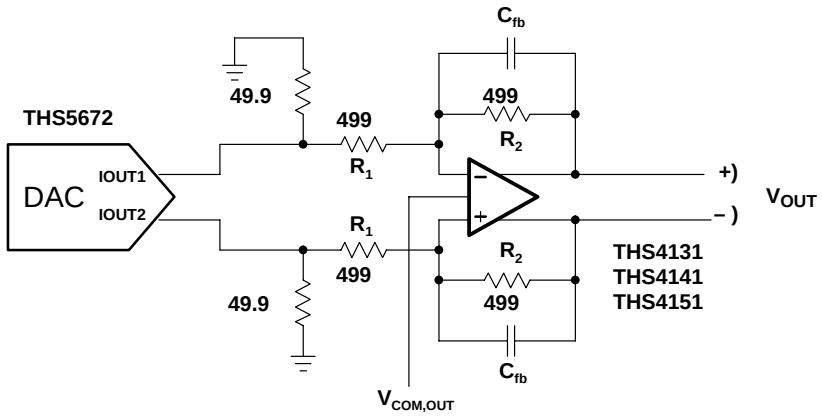


Figure 13–7. A Single-Pole Reconstruction Filter

13.7 External V_{ref} Circuits for ADCs/DACs

Figure 13–8 shows an op amp voltage follower circuit that is often used to interface the external precision voltage reference supplying the ADC/DAC external reference voltage (see for example, Miller and Moore, [5], [6],1999, 2000 for a more detailed discussion on voltage reference circuits used in ADC an DAC systems). V_{in} is the output from a precision voltage reference, such as the Thaler Corp. VRE3050. The low-pass filter (formed by C_1R_1) filters noise from the reference and op amp buffer. The –3 dB corner frequency of the filter is $1/2\pi C_1R_1$and the transfer function for this circuit can be written as

$$\frac{V_o}{V_{in}} = \frac{\left(1 + sC_2R_2\right)}{\left(s^2 + \frac{sC_2R_2}{C_1C_2R_1R_2} + \frac{1}{C_1C_2R_1R_2}\right) C_1C_2R_1R_2} \tag{13–6}$$

which has a zero at $s = C_2R_2$.

With the approximation $C_2R_2 = 2C_1R_1$, the denominator polynomial is solved for complex poles p_1 and p_2 of the response, which results in:

$$P_1 = -\frac{1}{2C_1R_1} + j\frac{1}{2C_1R_1} \quad (13\text{–}7)$$

$$P_2 = -\frac{1}{2C_1R_1} - j\frac{1}{2C_1R_1} \quad (13\text{–}8)$$

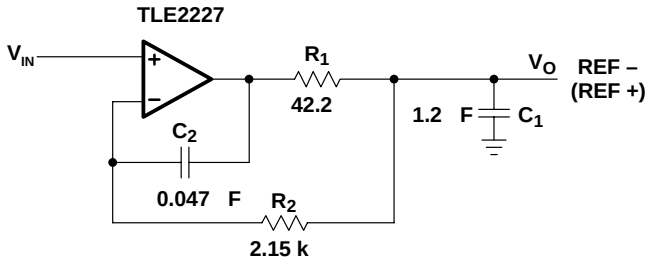


Figure 13–8. Voltage Reference Filter Circuit

The zero in the numerator of the transfer function improves the relative stability of the circuit. Resistor R_2 should be kept fairly low, since a small amount of bias current flows through it and causes dc error and noise. Resistor R_1 value ranges from 10 Ω to 50 Ω. Resistor R_1 is in the feedback loop, so any small leakage current that is due to capacitor C_1 flows through R1and the voltage dropped across R_1 is divided by the loop gain. For all practical purposes, the voltage across C_2 is 0 V and hence give rise to negligible leakage current.

A design example for a 3-kHz bandwidth filter is illustrated:

Choose C1 = 1.2 μF and R1 = 42.2 Ω.

Having determined the value for C1 and R1, the capacitor C2 value is estimated to be approximately 4% to 5% of C1 value (C2 = 0.047 μF) and resistor R2 is calculated using the approximation $C_2R_2 = 2C_1R_1$ (R2 = 2.15 kΩ).

The calculated –3-dB bandwidth for the circuit is 3.1 kHz and this value agrees with the circuit's frequency response plot shown in Figure 13–9. This circuit topology is good for driving large capacitive loads.

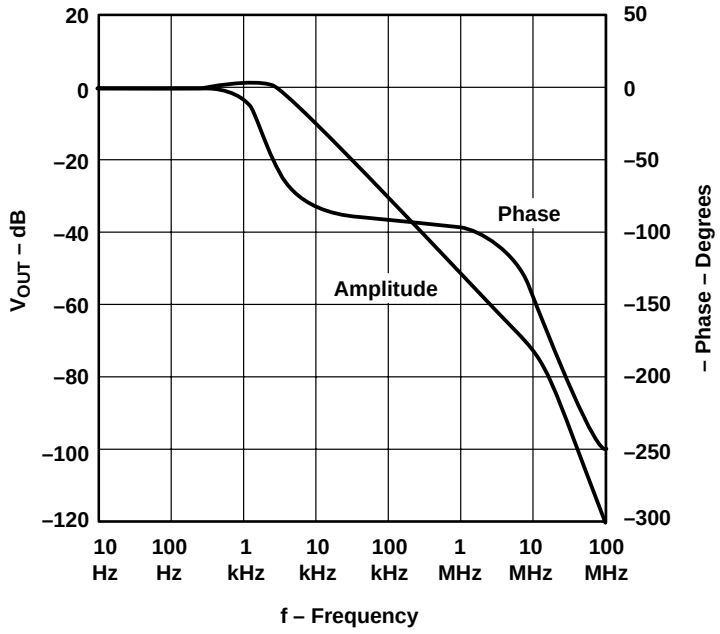


Figure 13–9. Voltage Follower Frequency Response Plot

Figure 13–10 shows an external reference circuit that provides a wide adjustment range of the ADC full-scale range. Resistors R281 and R282 play two roles in this circuit:

- Form part of the low-pass filter used to rolloff noise
- Isolate the ADC's reference input load capacitance from the buffer op amp output

Potentiometer R247 sets the external common-mode voltage (V_{ocm}) for the differential amplifier in Figure 13–11.

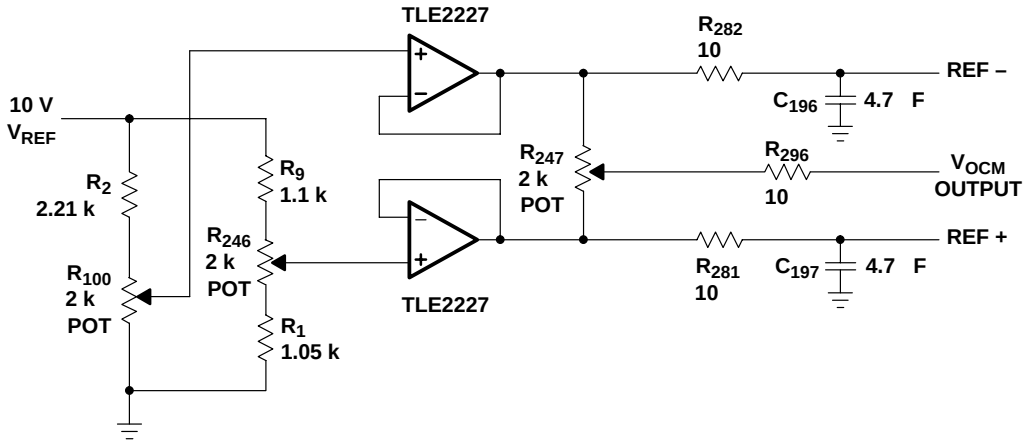


Figure 13–10. External Voltage Reference Circuit for ADC/DAC.

13.8 High-Speed Analog Input Drive Circuits

Communication ADCs, for the most part, have differential inputs and require differential input signals to properly drive the device. Drive circuits are implemented with either RF transformers or high-speed differential amplifiers with large bandwidth, fast settling time, low output impedance, good output drive capabilities, and a slew rate of the order of 1500 V/μS. The differential amplifier is usually configured for a gain of 1 or 2 and is used primarily for buffering and converting the single-ended incoming analog signal to differential outputs. Unwanted common-mode signals, such as hum, noise, dc, and harmonic voltages are generally attenuated or cancelled out. Gain is restricted to wanted differential signals, which are often 1 V to 2 V.

The analog input drive circuit, as shown in Figure 13–11, employs a complementary bipolar (BiCom) THS4141 device. BiCom offers fast speed, linear operation over a wide frequency range, and wide power-supply voltage range, but draws slightly more current than a BiCMOS device. The circuit closed-loop response is shown in Figure 13–12, where the –3-dB bandwidth is 120 MHz measured at the output of the amplifier. The analog input V_{in} is ac-coupled to the THS4141 and the dc voltage V_{ocm} is the applied input common-mode voltage. The combination R47– C57 and R26 – C34 are selected to meet the desired frequency rolloff. If the input signal frequency is above 5 MHz, higher-order low-pass filtering techniques (third-order or greater) are employed to reduce the op amp's inherent second harmonic distortion component.

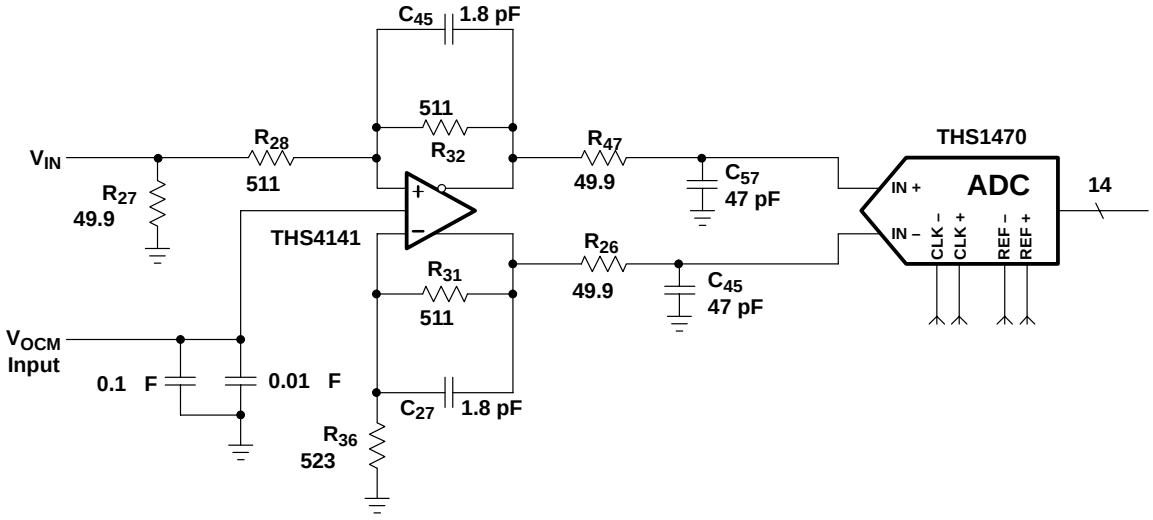


Figure 13–11. Single-Ended to Differential Output Drive Circuit

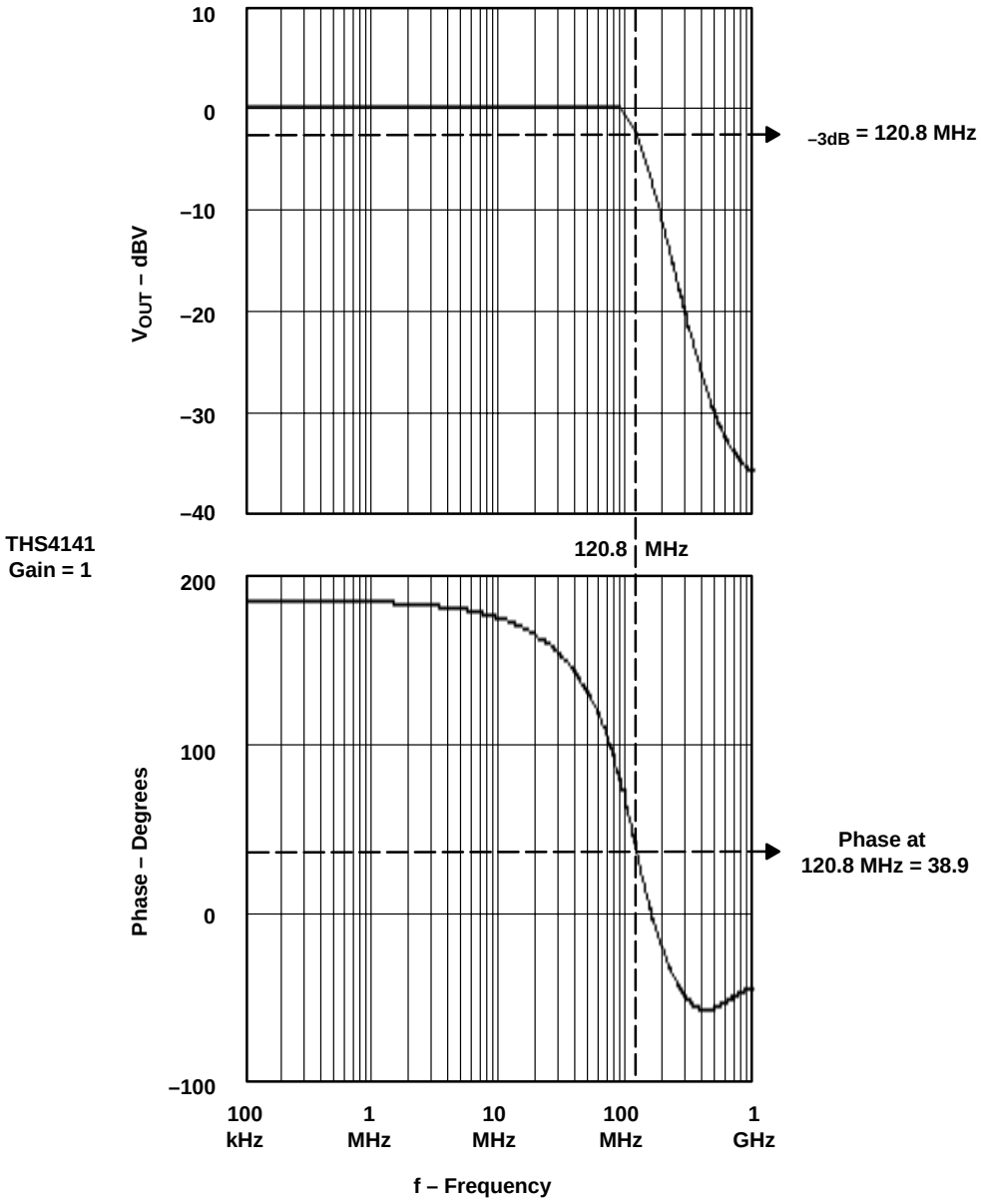


Figure 13–12. Differential Amplifier Closed-Loop Response

Figure 13–13 shows a design example of an ac-coupled, single-ended analog input drive circuit. This circuit uses a THS3201 current-feedback op amp and operates up to 975 MHz. The amplifier is configured as a noninverting amplifier with a gain of 2, where gain is 1 + R4/R3. In a current-feedback amplifier, the feedback resistor sets the amplifier

bandwidth and the frequency response shape as well as defining the gain of the circuit. R4 and R3 cannot be arbitrary values.

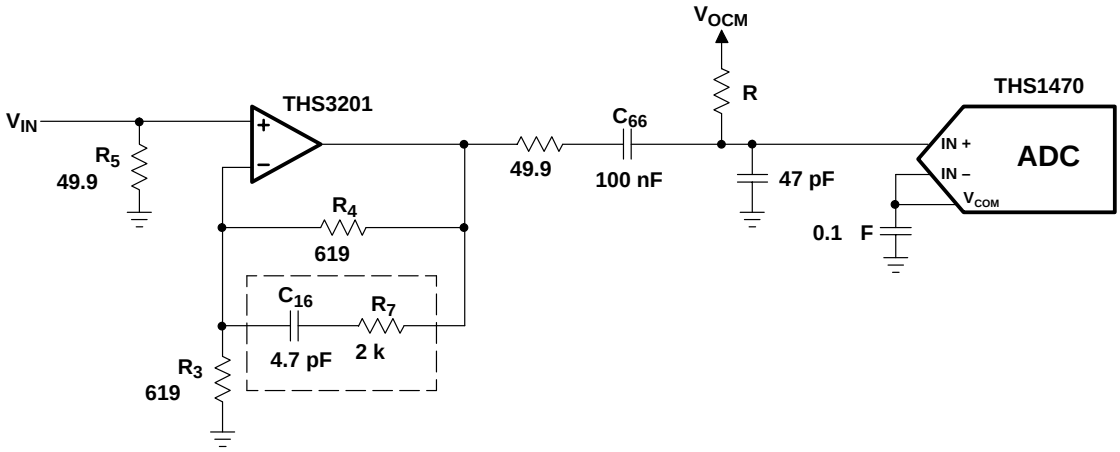


Figure 13–13. ADC Single-Ended Input Drive Circuit

The frequency response plots for a gain of 2, with R3 = R4 = 619 Ω, are shown in Figures 13–13 and 13–14. R4 affects the amplifier bandwidth and frequency response peaking. R3 has no effect on bandwidth and frequency peaking; it only affects the gain. The –3-dB bandwidth is 520 MHz. Increasing the resistance of R4 decreases the bandwidth. Conversely, lowering R4 resistance increases the bandwidth at the expense of increased peaking in the ac response.

The phase plot in Figure 13–14 exhibits a fairly linear phase shift (a flat group delay response) and hence the amplifier output should show excellent signal reproduction.

Unlike a voltage-feedback amplifier, the power supply voltage affects the bandwidth of a current-feedback amplifier. For example, lowering the supply voltage of the THS3201 from ± 5 V to ± 2.5 V reduces the bandwidth from 925 MHz to 350 MHz.

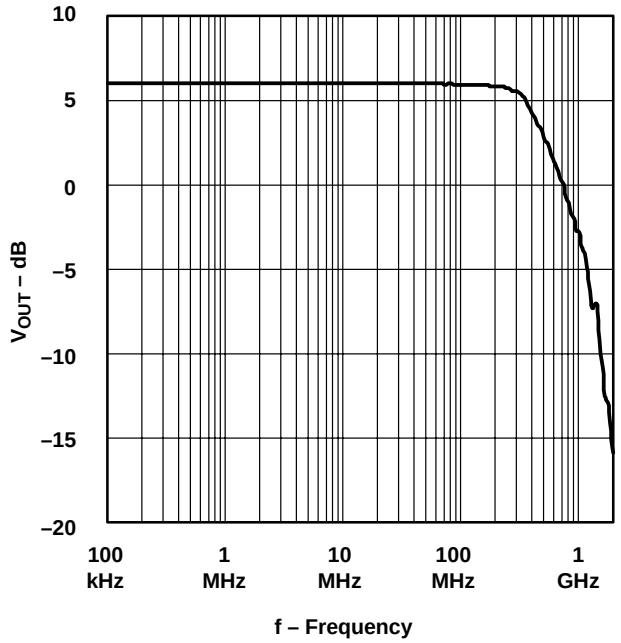


Figure 13–14. Gain vs Frequency Plot for THS3201

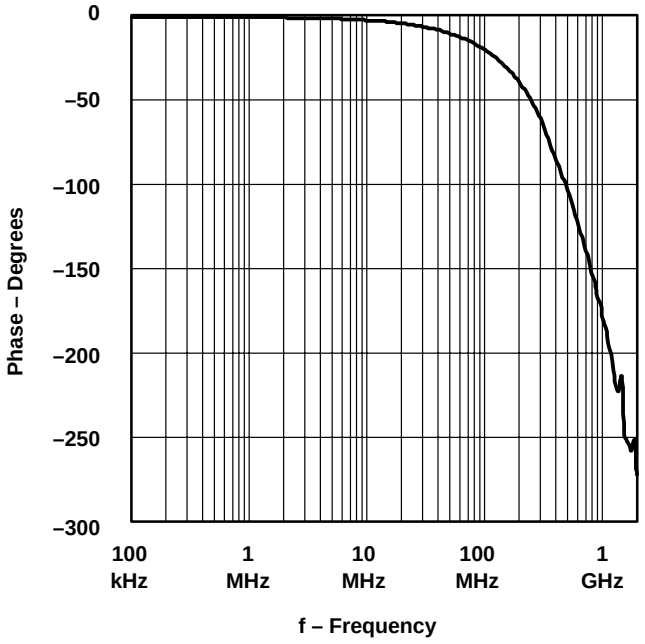


Figure 13–15. Phase vs Frequency for THS3201

13.9 References

[1] Rabiner, L.R., and Gold. B., *Theory and Application of Digital signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey, 1975

[2] Parks, T.W., and Burrus, C.S., *Digital Filter Design*, Wiley, New York, NY, 1987

[3] Miller, P., *Aspects of Data Acquisition System Design*, Analog Application Journal, Texas Instruments, August 1999

[4] Proakis, J.G., *Digital Communications*, McGraw–Hill Book Company, New York, 1983

[5] Miller, P., and Moore, D*., Precision Voltage References*, Analog Application Journal, Texas Instruments, November 1999

[6] Miller, P., and Moore, D*., The Design and Performance of a Precision Voltage Reference Circuit for 14-Bit and 16-Bit A-to-D and D-to-A Converters*, Analog Application Journal, Texas Instruments, May 2000

[7] Daniel H. Sheingold, *Analog–Digital Conversion Handbook*, 3rd Edition, Prentice Hall, Englewood Cliffs, New Jersey 07632

[8] Graeme J.G., Tobey G.E., and Huelsman L.P., Burr–Brown, *Operation Amplifiers Design and Application*, McGraw–Hill Book Company, New York, 1971

[9] Schartz, M., *Information Transmission, Modulation and Noise*, McGraw–Hill Book Company, New York, 1970

[10] Papoulis A., *On The Approximation Problem on Filter Design*, IRE conv. Record, Pt.2, pp. 175 – 1957

Chapter 14

Interfacing D/A Converters to Loads

Bruce Carter

14.1 Introduction

A digital to analog converter, or D/A, is a component that takes a digital word and converts it to a corresponding analog voltage. It has the opposite function of an A/D converter. The D/A is only capable of producing a quantized representation of an analog voltage, not an infinite range of output voltages.

The application will almost always dictate the selection of the D/A converter, leaving the designer the task of interfacing that converter with the output load.

A D/A converter interfaces with a buffer op amp. Most D/A converters are manufactured with a process that is incompatible with op amps. Therefore, the op amp cannot be manufactured on the same IC. It must be external, and its characteristics are an integral part of the conversion process. In most cases, the data sheet will make a recommendation for the selection of a buffer op amp. Follow the recommendation, unless there is a compelling reason not to do so. Performance can be improved only if you know exactly what op amp specifications need to be optimized.

Signal conditioning — low pass filtering, DC offsets, and power stages — should all be placed after the recommended op amp buffer. Do not attempt to combine these functions with the buffer unless you are an experienced designer with a good grasp of all of the implications.

14.2 Load Characteristics

There are two main types of loads that a D/A may have to drive — ac and dc. Each has different characteristics, and will require different interface circuitry.

14.2.1 DC Loads

These include linear actuators such as those used on positioning tables, motors, programmable power supplies, outdoor displays, and lighting systems. Large load currents

and/or high voltages characterize some of these loads. DC accuracy is important, because it is related to a series of desired mechanical positions or intensities in the load device.

14.2.2 AC Loads

These include linear actuators such as those used on positioning tables, motors, programmable power supplies, outdoor displays, and lighting systems. Large load currents and/or high voltages characterize some of these loads. DC accuracy is important, because it is related to a series of desired mechanical positions or intensities in the load device.

14.3 Understanding the D/A Converter and its Specifications

It is important to understand the D/A converter and its specifications before discussing interfaces.

14.3.1 Types of D/A Converters — Understanding the Tradeoffs

D/A converters are available in several types, the most common of which is the resistor ladder type. There are several variations on the resistor ladder technique, with the R/2R configuration being the most common.

14.3.2 The Resistor Ladder D/A Converter

In this type of converter, a precision voltage reference is divided into 2^{N-1} parts in an internal voltage divider, where N is the number of bits specified for the converter. One switch at a time turns on, corresponding to the correct dc level (Figure 14–1).

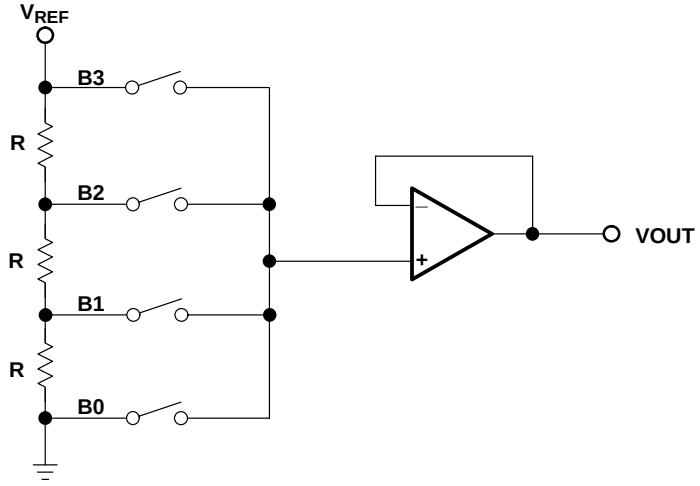


Figure 14–1. Resistor Ladder D/A Converter

Unfortunately, the number of resistors and switches doubles for each additional bit of resolution. This means that an 8-bit D/A converter would have 255 resistors and 256 switches, and a 16-bit D/A converter would have 65535 resistors and 65536 switches. For this reason, this architecture is almost never used for higher resolution D/A converters.

14.3.3 The Weighted Resistor D/A Converter

This type of converter is very similar to the Resistor Ladder D/A converter. In this case, however, each resistor in the string is given a value proportional to the binary value of the bit it represents. Currents are then summed from each active bit to achieve the output (Figure 14–2).

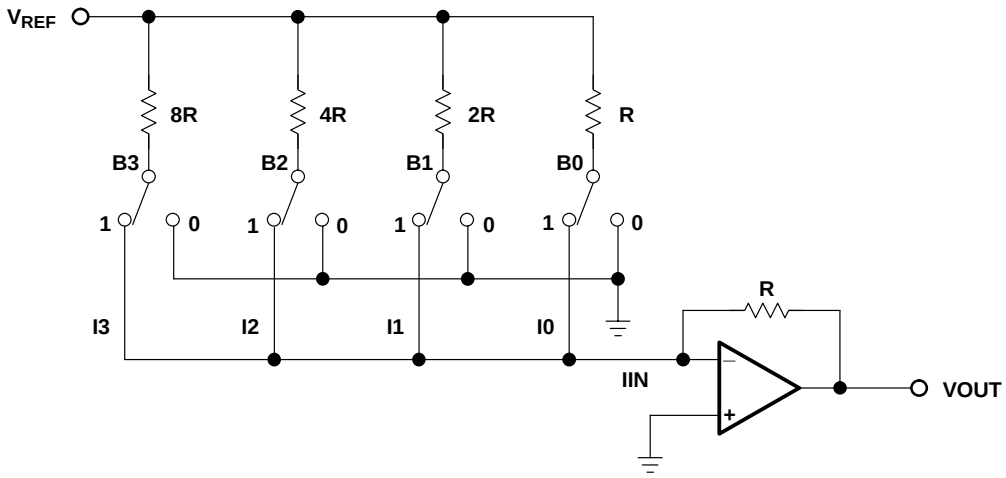


Figure 14–2. Binary Weighted D/A Converter

The number of resistors and switches reduced to one per bit, but the range of the resistors is extremely wide for high-resolution converters, making it hard to fabricate all of them on the IC. The resistor used for B3 is the limiting factor for power dissipation from V_{REF} to ground.

This converter architecture is often used to make logarithmic converters. In this case, the R, 2R, 4R, 8R … resistors are replaced with logarithmically-weighted resistors.

This type of converter, and the R/2R converter described in the next paragraph, use a feedback resistor fabricated on the D/A IC itself. This feedback resistor is not an optional convenience for the designer — it is crucial to the accuracy of the D/A. It is fabricated on the same silicon as the resistor ladder. Therefore, it experiences the same thermal drift as the resistor ladder. The gain of the buffer amplifier is fixed, with a full-scale output voltage limited to V_{REF}. If a different full-scale D/A output voltage is needed, change V_{REF}.

If the full scale V_{OUT} must exceed the maximum rating of the D/A reference voltage, use a gain stage after the buffer op amp (see Paragraph 14.7.2).

The op amp must be selected carefully, because it will be operated in much less than unity gain mode for some combinations of bits. This is probably one of the main reasons why this architecture is not popular, as well as the requirement for a wide range of resistor values for high precision converters.

14.3.4 The R/2R D/A Converter

An R/R2 network can be used to make a D/A converter (Figure 14–3).

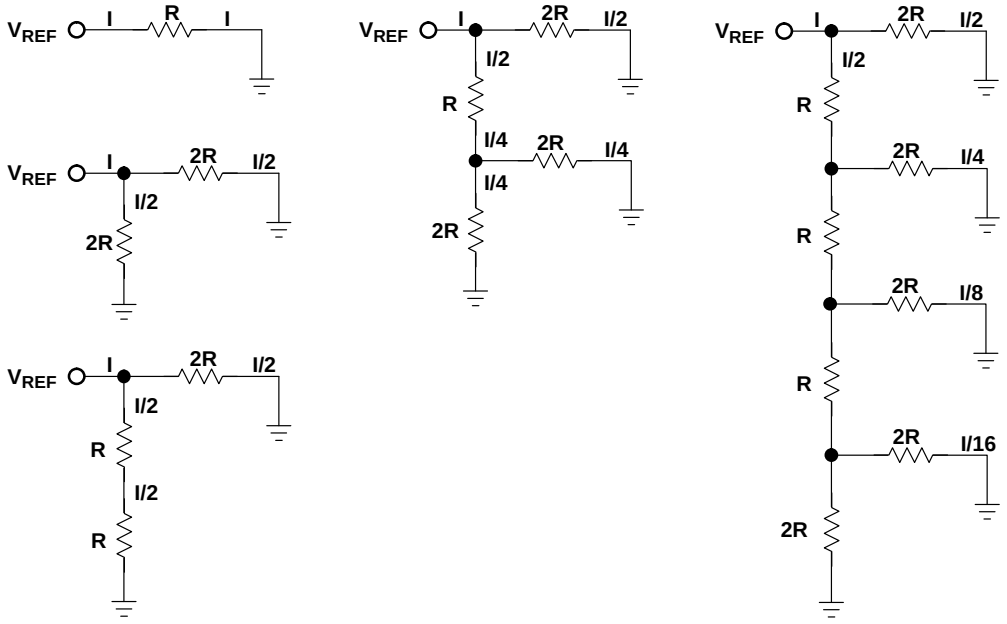


Figure 14–3. R/2R Resistor Array

For a given reference voltage V_{REF}, a current I flows through resistor R. If two resistors, each the same value (2R) are connected from V_{REF} to ground, a current I/2 flows through each leg of the circuit. But the same current will flow if one leg is made up of two resistors, each with the value of R. If two resistors in parallel whose value is 2R replace the bottom resistor, the parallel combination is still R. I/4 flows through both legs, adding to I/2. Extending the network for 4 bits as shown on the right, the total current on the bottom leg is I/4 plus I/8 plus I/16 plus I/16 in the resistor to ground. Kirchoff's current law is satisfied, and convenient tap points have been established to construct a D/A converter (Figure 14–4):

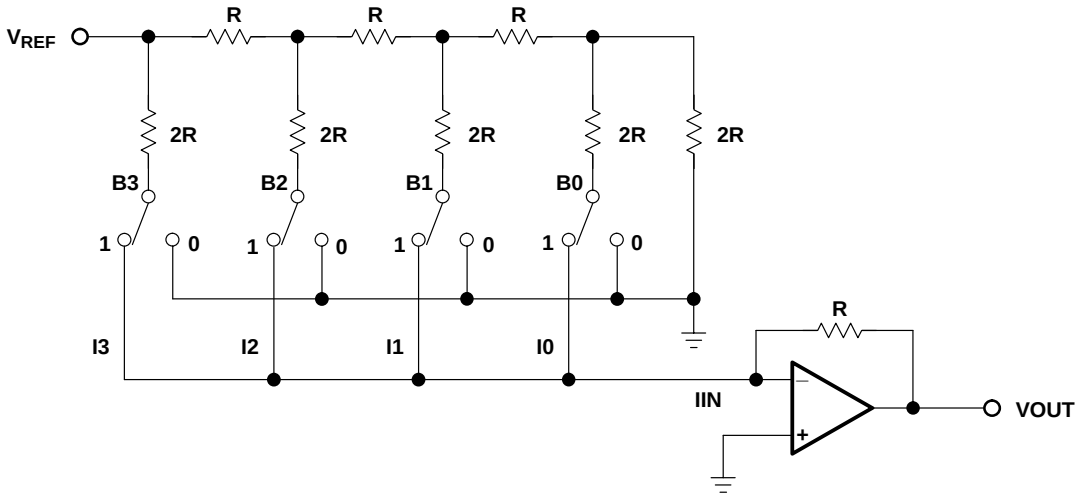


Figure 14–4. R/2R D/A Converter

This converter architecture has advantages over the types previously mentioned. The number of resistors has doubled from the number required for the current-summing type, but there are only two values. Usually, the 2R resistors are composed of two resistors in series, each with a value of R. The feedback resistor for the buffer amplifier is again fabricated on the converter itself for maximum accuracy. Although the op amp is still not operated in unity gain mode for all combinations of bits, it is much closer to unity gain with this architecture.

The important op amp parameters for all resistor ladder D/As are:

- Input offset voltage — the lower the better. It adds to the converter offset error.
- Input bias current — the lower the better. The product of the bias current and the feedback resistance creates an output offset error.
- Output voltage swing — it must meet or preferably exceed the zero to full-scale swing from the D/A.
- Settling time and slew rate — must be fast enough to allow the op amp to settle before the next digital bit combination is presented to the D/A input register.

14.3.5 The Sigma Delta D/A Converter

The sigma delta D/A converter takes advantage of the speed of advanced IC processes to do a conversion as a series of approximations summed together. A phase-locked loop-derived (PLL) sample clock operates at many times the overall conversion frequency — in the case shown in Figure 14–5, it is 128×. The PLL is used to drive an interpolation filter, a digital modulator, and a 1-bit D/A converter. The conversion is done by using the density

ratio of the voltage out of the 1-bit D/A as the analog signal. As the pattern of 1s and 0s is presented to the 1-bit converter, their time average at the sample frequency recreates the analog waveform.

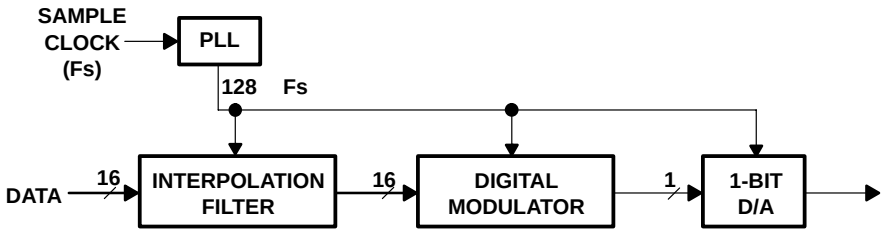


Figure 14–5. Sigma Delta D/A Converter

Sigma delta converters are popular for audio frequencies, particularly CD players. The primary limiting factor is the sample clock. CD players operate at a sample rate of 44.1 kHz, which means that according to Nyquist sampling theory, the maximum audio frequency that can be reproduced is 22.05 kHz. If an audio frequency of 23.05 kHz is present in the recorded material, it will alias back into the audio output at 1 kHz — producing an annoying whistle. This places a tremendous constraint on the low-pass filter following the D/A in a CD player. It must reject all audio frequencies above 22.05 kHz while passing those up to 20 kHz, the commonly accepted upper limit of human hearing. While this can be done in conventional filter topologies, they are extremely complex (9 or more poles). Inevitably, phase shift and amplitude rolloff or ripple will start far below 20 kHz. The original CD players often sounded a bit “harsh” or “dull” because of this.

The solution was to overclock the sample clock. To keep things simple, designers made it a binary multiple of the original sampling frequency. Today, 8× or even higher oversampling is standard in CD players. Little do the audio enthusiasts know that the primary reason why this was done was to substantially reduce the cost of the CD player! A faster sample clock is very cheap. Nine-pole audio filters are not. At 8× oversampling, the CD player only needs to achieve maximum rolloff at 352.8 kHz — a very easy requirement. Instead of the filter having to roll off in a mere 2 kHz of bandwidth, now it has 332 kHz of bandwidth to accomplish the rolloff. The sound of an oversampled CD player really is better, but it comes at the cost of increased radiated RFI, coming from the sample clock.

Sigma delta converters introduce a great deal of noise onto the power rails, because the internal digital circuitry is continually switching to the power supply rails at the sample clock frequency F_S.

14.4 D/A Converter Error Budget

The system designer must do an error budget to know how many bits are actually needed to meet the system requirements — how much “graininess” or what step size is acceptable in the output signal.

14.4.1 Accuracy versus Resolution

It is important for the designer to understand the difference between converter accuracy and converter resolution. The number of bits determines resolution of a converter. Insufficient resolution is not error — it is a design characteristic of the D/A. If a given converter's resolution is insufficient, use a converter with better resolution (more bits).

Accuracy is the error in the analog output from the theoretical value for a given digital input. Errors are described in the next paragraph. A very common method of compensating for D/A error is to use a converter that has one or two bits more resolution than the application requires. With the cost of converters coming down, and more advanced models being introduced every day, this may be cost effective.

14.4.2 DC Application Error Budget

DC applications will depend on the value of dc voltage coming out of the converter. THD and signal-to-noise will not be important because the frequency coming out of the converter is almost dc.

The resolution of a converter is ±1/2 LSB, where an LSB is defined as:

$$1\ \text{LSB} = \frac{V_{FS}}{2^N - 1} \qquad (14\text{–}1)$$

Where:

V_{FS} Full-scale output voltage
N Number of converter bits

The number of bits in a dc system determines the dc step size that corresponds to a bit. Table 14–1 shows the number of bits, and the corresponding voltage step size for three popular voltages:

Table 14–1. DC Step Size for D/A converters

BITS	STATES	3 V	5 V	10 V
4	16	0.1875	0.3125	0.625
8	256	0.011719	0.019531	0.039063
10	1024	0.00293	0.004883	0.009766
12	4096	0.000732	0.001221	0.002441
14	16384	0.000183	0.000305	0.00061
16	65536	4.58 E–05	7.63 E–05	0.000153
18	262144	1.14 E–05	1.91 E–05	3.81 E–05
20	1048576	2.86 E–06	4.77 E–06	9.54 E–06
22	4194304	7.15 E–07	1.19 E–06	2.38 E–06
24	16777216	1.79 E–07	2.98 E–07	5.96 E–07

The bit step size can get critical, especially for portable equipment. There is a requirement to operate off of low voltage, to minimize the number of batteries. The buffer amplifier, if it includes gain, will use large resistor values, lowering its noise immunity. Fortunately, the vast majority of dc applications are not portable; they are in an industrial environment.

For example: A converter is used to position a drill on a table used to drill PCB holes. The positions of the holes are specified as 0.001 inches, ±0.0003 inches. The actuators are centered on the table at zero volts, with full negative position of –12 inches occurring at –5 V, and full positive position of +12 inches occurring at +5 V. There are two actuators, one for vertical, and one for horizontal.

This example has several aspects. The first is that the positioning voltage has to swing both positive and negative. In the real world, it may be necessary to add (or subtract in this case) a fixed offset to the D/A output. The output voltage has to swing over a 10-V range, which may mean that the output of the D/A has to be amplified. The actuators themselves probably operate off of higher current than the D/A is designed to provide. Section 14.7 covers some methods for meeting these requirements.

Assume, for now, that the D/A has the necessary offset and gain. A ±12 inch position is 24 inches total, which corresponds to ±5 V from the D/A circuitry. The 24-inch range must be divided into equal 0.0003-inch steps to meet the resolution requirement, which is 80,000 steps. From Table 14–1, an 18-bit D/A converter is required. The actual system will be able to position with a step size of 0.0000916 inches. Two independent conversion systems are needed, one for horizontal and one for vertical.

14.4.3 AC Application Error Budget

The error budget for an ac application will most likely be specified as total harmonic distortion, dynamic range, or signal-to-noise ratio. Assuming no internal noise, and no noise in the buffer op amp circuitry, the inverse of the dynamic range is the signal-to-noise ratio of the converter D/A. Of course, noise is always present, and is measured with all input data set to zero. Noise will make the S/N ratio decrease.

The number of converter bits, however, is the overwhelming factor determining these parameters. Technically, they are not “errors”, because the design of the converter sets them. If the designer cannot live with these design limits, the only choice is to specify a converter with better resolution (more bits).

14.4.3.1 Total Harmonic Distortion

The total harmonic distortion of an ideal D/A converter is the quantization noise due to the converter resolution. The number of bits of the converter determines the lowest possible total harmonic distortion. The greater the number of bits, the lower the amplitude of the harmonics, as shown in Figure 14–6.

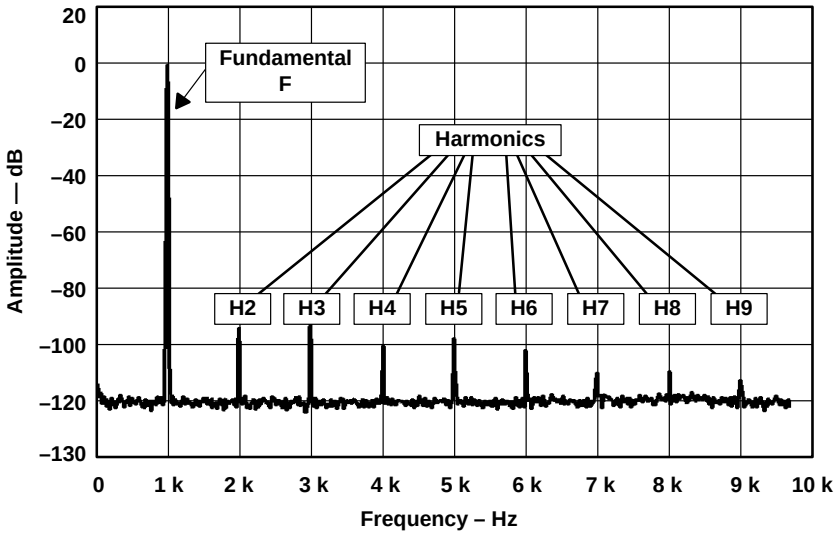


Figure 14–6. Total Harmonic Distortion

Assuming ideal D/A conversion, there is a direct relationship between the number of bits and the THD caused by the resolution:

$$\text{THD}(\%) = \frac{1}{2^N} \times 100 \qquad (14\text{–}2)$$

Where N is the number of converter bits. Of course, this is the limit for ideal conversion.

14.4.3.2 Dynamic Range

There is also a direct relationship between the number of bits (n) and the maximum dynamic range of the D/A (Equation 14–2 and Table 14–2):

$$\text{Dynamic Range} = 6.02 \times n + 1.76 \qquad (14\text{–}3)$$

Notice that there is approximately a 6-dB improvement in dynamic range per bit. This is an easy way to figure out what improvement can be realized by increasing the number of bits from one value to another.

Table 14–2. Converter Bits, THD, and Dynamic Range

BITS	STATES	THD	Dynamic Range
4	16	6.25%	25.8
8	256	0.390625%	49.9
10	1024	0.097656%	62.0
12	4096	0.024414%	74.0
14	16384	0.006104%	86.0
16	65536	0.001526%	98.1
18	262144	0.000381%	110.1
20	1048576	0.000095%	122.2
22	4194304	0.000024%	134.2
24	16777216	0.000006%	146.2

For example, if the designers of a CD player want to have a 90-dB signal-to-noise ratio, they would pick a 16-bit converter from Table 14–2. The THD is 0.0015% minimum.

14.4.4 RF Application Error Budget

RF applications are a high-frequency subset of ac applications. RF applications may be concerned with the position and relative amplitude of various harmonics. Minimizing one harmonic at the expense of another may be acceptable if the overall RF spectrum is within specified limits.

14.5 D/A Converter Errors and Parameters

The D/A errors described in this section will add to the errors caused by the resolution of the converter.

This section is divided into dc and ac sections, but many of the dc errors masquerade as ac errors. A given D/A may or may not include either dc or ac error specifications. This should give the designer a clue that the device is optimized for dc or ac applications. Like any component, D/A converters are designed with tradeoffs. It is possible to misapply a converter meant for high-frequency ac operation in a dc application, etc.

14.5.1 DC Errors and Parameters

The following paragraphs describe D/A dc errors and parameters.

14.5.1.1 Offset Error

The analog output voltage range for the complete range of input bits may be shifted linearly from the ideal 0 to full-scale value (Figure 14–7). The offset error is the $\pm\Delta$ V from 0 V that results when a digital code is entered that is supposed to produce 0.

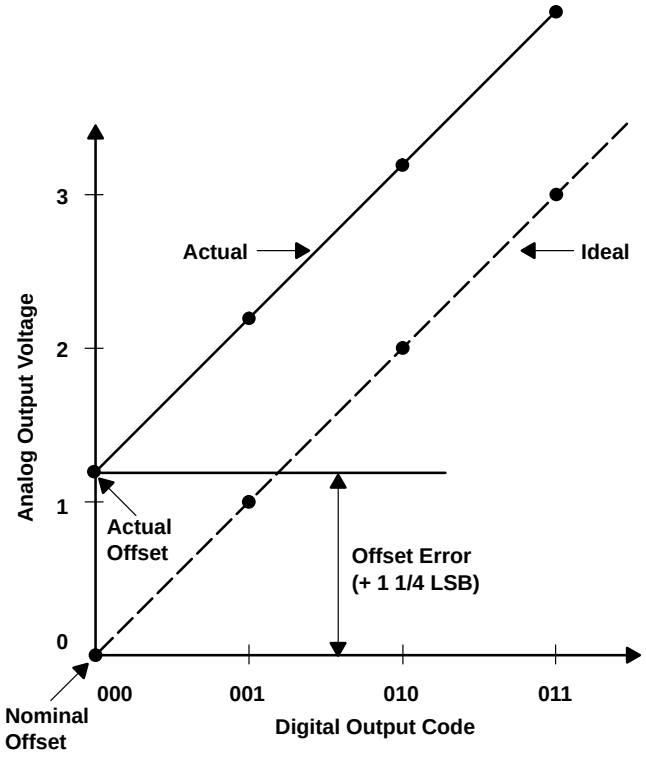


Figure 14–7. D/A Offset Error

Related to the offset error is the offset error temperature coefficient, which is the change in offset over temperature. This is usually specified in ppm/°C.

Offset error is critical in dc applications. For this reason, a buffer op amp must be selected that does not contribute to the problem — its own offset voltages should be much less than that of the converter. In ac applications, the offset error is not important and can be ignored. The buffer op amp can be selected for low THD, high slew rate, or whatever other parameters are important for the application.

14.5.1.2 Gain Error

The gain of the D/A converter may greater than or less than the gain needed to produce the desired full-scale analog voltage (Figure 14–8). The gain error is the difference in slope between the ideal D/A output gain and the actual gain.

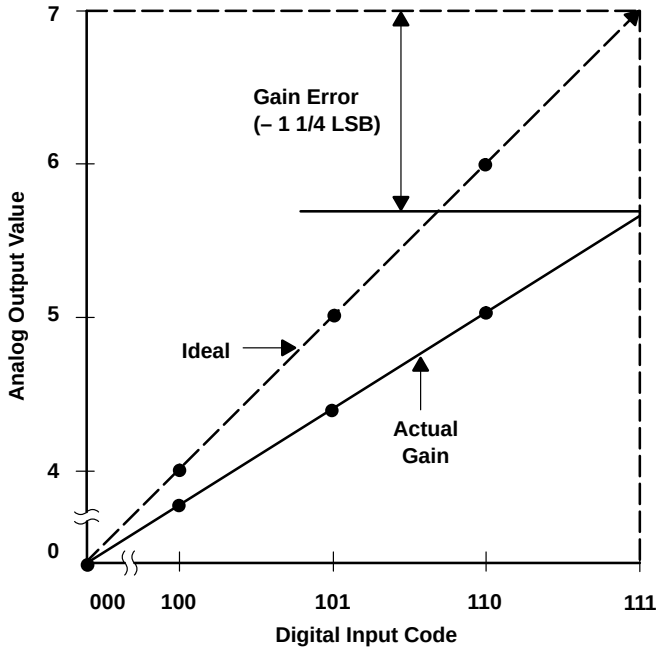


Figure 14–8. D/A Gain Error

Related to the gain error is the gain error temperature coefficient, which is the change in gain over temperature.

Gain error can be critical in both ac and dc applications. For example:

- An RF predriver must not cause the output stage to exceed FCC license requirements.
- A mechanical positioner must not stop short of or go past its intended position.

The op amp buffer should be operated with the internal feedback resistor. If possible, full-scale amplitude adjustments should be made to V_{REF}. This way, tolerances and thermal drift in external resistors do not contribute to the gain error.

14.5.1.3 Differential Nonlinearity Error

When the increase in output voltage (ΔV) is not the same for every combination of bits, the converter has a differential nonlinearity error (DNL). If the DNL exceeds 1 LSB, the converter is nonmonotonic. This can cause a problem for some servo control loops. A nonmonotonic D/A would appear in Figure 14–9 as a momentary dip in the analog output characteristic.

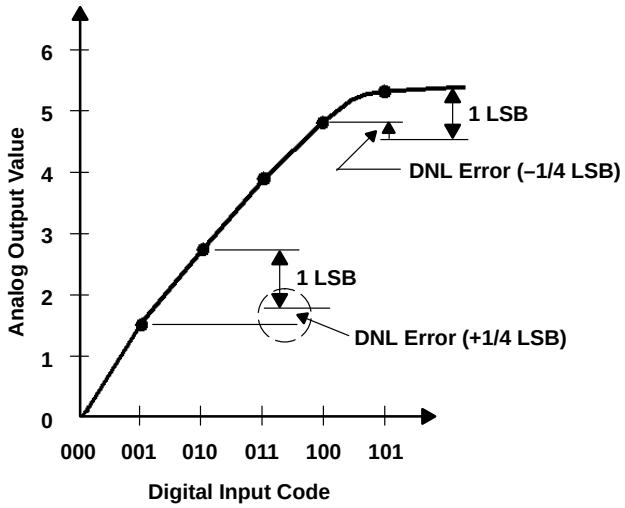


Figure 14–9. Differential Nonlinearity Error

14.5.1.4 Integral Nonlinearity Error

The integral nonlinearity error is similar to the differential nonlinearity error, except it is a first-order effect that stretches across the entire range from 0 to full scale output voltage (Figure 14–10).

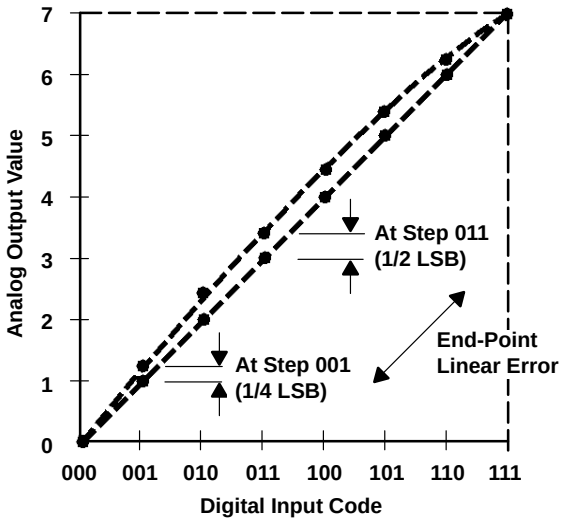


Figure 14–10. Integral Nonlinearity Error

Both the INL and DNL errors affect ac applications as distortion and spectral harmonics (spurs). In dc applications, they will result in an error in the dc output voltage. The mechanical steps of a positioning table, for instance, may not be exact increments.

14.5.1.5 Power Supply Rejection Ratio

The power supply rejection ratio is sometimes called the *power supply sensitivity*. It is the ability of the converter to reject ripple and noise on its power inputs. DC applications may not be adversely affected. Poor power supply rejection can cause spurs and harmonic distortion in ac applications, as external frequency components leak into the output and modulate with it. The designer must decouple the D/A and buffer op amp carefully to combat these problems.

14.5.2 AC Application Errors and Parameters

The following paragraphs describe D/A ac errors and parameters.

14.5.2.1 THD+N

There will always be some noise that is generated internally in the converter and buffer amp. A useful specification for audio and communication system designers is the THD+N (total harmonic distortion plus noise). The distortion plus noise (THD+N) is the ratio of the sum of the harmonic distortion and noise to the rms power of the input signal. As was the case with op amp parameters (Chapter 11), the noise sources add according to the root-mean-square law. The distortion and noise are measured separately and then added together to form the ratio. The noise voltage relates to the measured bandwidth.

14.5.2.2 SINAD

The signal-to-noise and distortion (SINAD) is the ratio of the input signal to the sum of the harmonic distortion and noise. The distortion and noise are measured separately and than added together to form the ratio. The SINAD is the reciprocal to the THD+N. The SINAD and THD+N are a good indication of the overall dynamic performance of the ADC, because all components of noise and distortion are included.

14.5.2.3 ENOB

The SINAD is used to determine the *effective number of bits* (ENOB) of accuracy the converter displays at that frequency. For example, a nominal 8-bit resolution D/A may be specified as having 45-dB SNR at a particular input frequency. The number of effective bits is defined as

$$\text{ENOB} = \frac{\text{SNR}_{\text{REAL}} - 1.76}{6.02} = 7.2 \text{ bits} \qquad (14\text{–}4)$$

The actual performance of the device is therefore less than its nominal resolution at this frequency.

14.5.2.4 Spurious Free Dynamic Range

Spurious free dynamic range is the difference in dB between the maximum signal component and the largest distortion component (Figure 114–11):

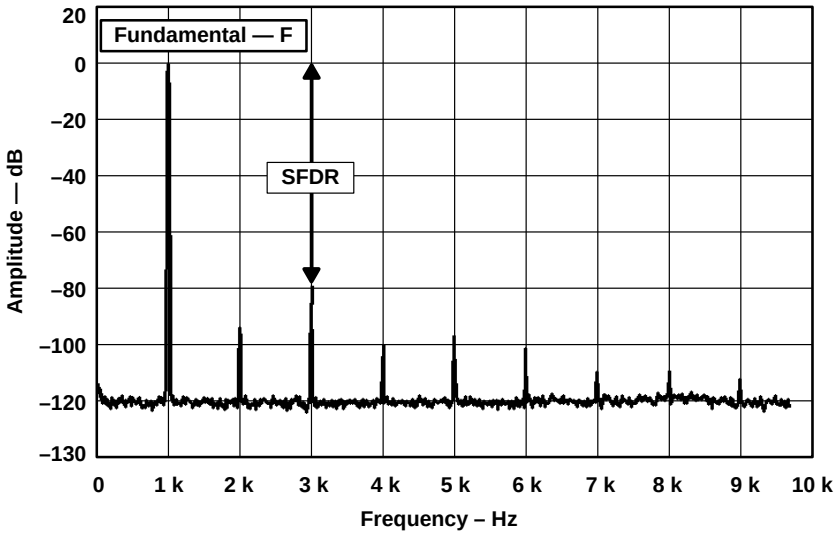


Figure 14–11. Spurious Free Dynamic Range

It is an important specification in RF applications, where FCC regulations specify the magnitude of spurs.

Improper decoupling may cause spurs. A notch filter can be used to eliminate a spur, but many RF applications are RF agile — changing the frequency of the spur as well. The notch must catch all spur frequencies, or it is useless.

14.5.2.5 Intermodulation Distortion

The differential and integral nonlinearity errors described previously appear in a high-frequency ac application as intermodulation distortion (Figure 14–12).

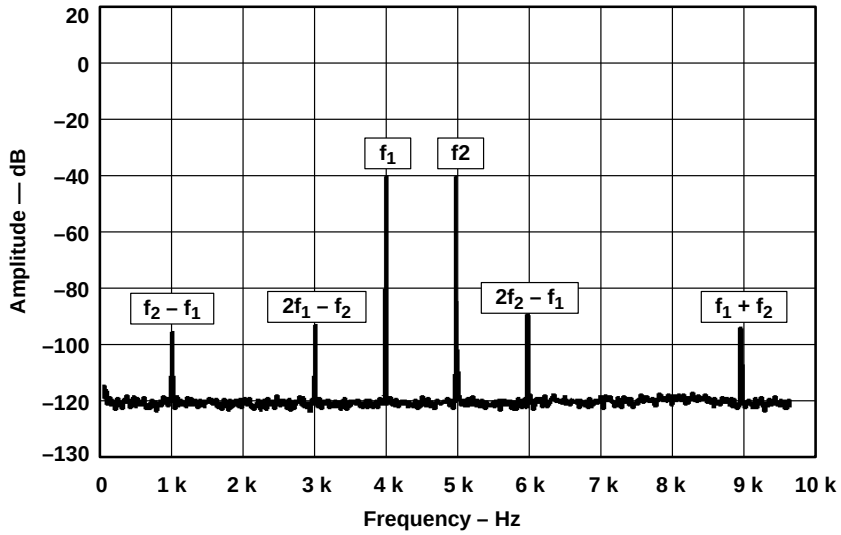


Figure 14–12. Intermodulation Distortion

The best method of combating intermodulation distortion is to make the buffer amplifier system as linear as possible (beware of rail-to-rail op amps that may not be linear near the voltage rails). Try to limit current through the internal feedback resistor in the DAC. See Paragraph 14.7.1 on increasing the voltage rail for suggestions about reducing internal feedback resistor power dissipation.

14.5.2.6 Settling Time

The settling time of a D/A converter is the time between the switching of the digital inputs of the converter and the time when the output reaches its final value and remains within a specified error band (Figure 14–13). Settling time is the reciprocal of the maximum D/A conversion rate.

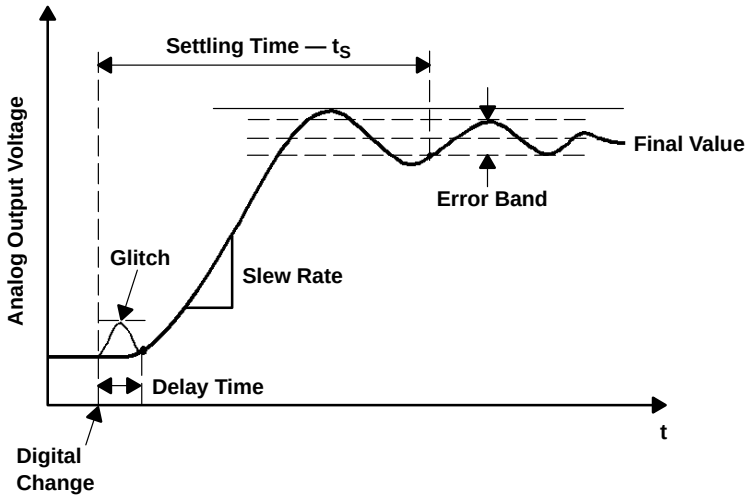


Figure 14–13. D/A Settling Time

When an output buffer op amp is used with a D/A, it becomes a part of the settling time/ conversion rate calculations.

Related to the settling time is a glitch response that occurs when the digital code changes state. Even though this effect is transitory in nature, it can result in noise or harmonics when used in fast AC applications. The best way of reducing the glitch is to properly decouple the D/A and op amp buffer (see Chapter 17). In extreme cases, a deglitching circuit may be needed (Figure 14–14).

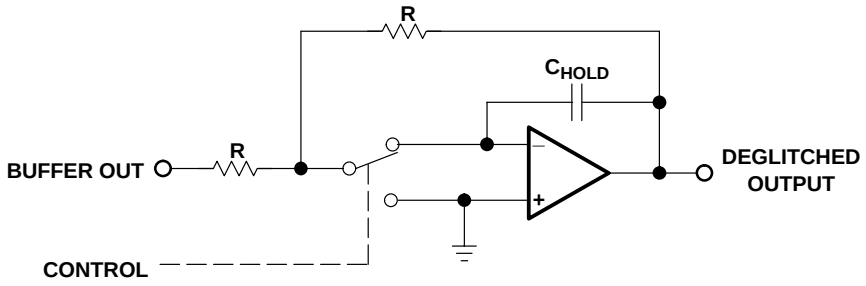


Figure 14–14. D/A Deglitch Circuit

This technique relies on the software designer to balance the timing of the control signal so it activates the hold function right before the D/A input code changes, then releases the hold right after the code has changed. The selection of C_{HOLD} is critical — it must hold the buffer output without droop and without compromising system bandwidth.

14.6 Compensating For DAC Capacitance

D/A converters are constructed of either bipolar or CMOS technology, with CMOS being the more common. CMOS transistors, however, have a lot of capacitance. This capacitance will add in D/A converters, depending on the number of resistors switched on or off. Capacitance at an inverting op amp input is a good way to cause it to oscillate, especially since some buffer amplifiers will be operated at less than unity gain. The converter capacitance C_O must be compensated for externally (Figure 14–15).

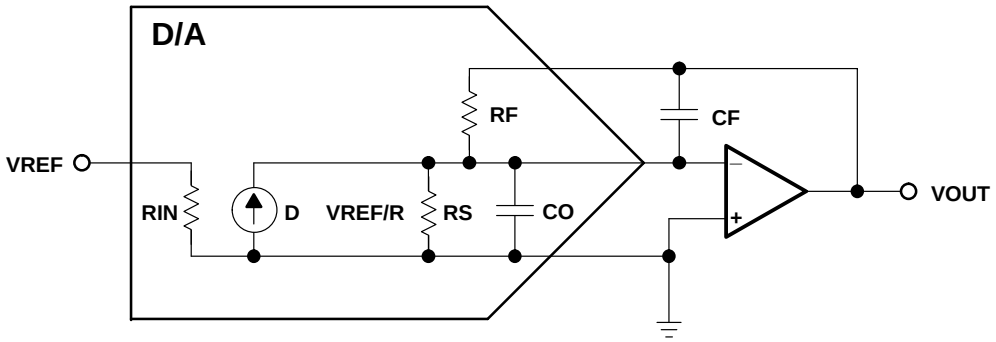


Figure 14–15. Compensating for CMOS DAC Output Capacitance

The normal technique for compensating the buffer amplifier for output capacitance is to add a feedback capacitor C_F. C_F is calculated by the following:

$$C_F = 2 \times \sqrt{\frac{C_O}{2\pi R_F} \times \frac{1}{G_{BW}}} \qquad (14–5)$$

Where:

C_O	the output capacitance from the D/A data sheet
R_F	the feedback resistance from the D/A data sheet
G_{BW}	the small signal unity gain bandwidth product of the output amplifier

Unfortunately, the feedback capacitors C_F and the internal D/A capacitance C_O will both limit the conversion speed of the D/A. If faster conversion is needed, a D/A with a lower output capacitance, and therefore a lower feedback compensation capacitor will be needed. The overall settling time with the external capacitance is:

$$T_S \approx \sqrt{\frac{R_F\left(C_O + C_F\right)}{2G_{BW}}} \qquad (14\text{–}6)$$

Where:

C_O	the D/A internal capacitance
R_F	the feedback resistor
C_F	the compensation capacitance
G_{BW}	the small signal unity gain bandwidth product of the output amplifier

14.7 Increasing Op Amp Buffer Amplifier Current and Voltage

Process limitations of op amps limit the power that can be dissipated at the output. Unfortunately, there are applications that will require the DAC to interface to loads that dissipate considerable power. These include actuators, position solenoids, stepper motors, loudspeakers, vibration tables, positioning tables — the possibilities are endless.

While several “power op amps” are available that can drive heavy loads, they usually compromise several other specifications to achieve the high power operation. Input voltage offset, input current, and input capacitance can be decades higher than the designer is accustomed to, and make these power op amps unsuitable for direct interface with a DAC as a replacement for the buffer op amp.

The power booster stage can be designed discretely, or a prepackaged amplifier of some sort, depending on what is needed for the application. Sometimes high current is required for driving loads such as actuators and stepper motors. Audio applications can require a lot of wattage to drive loudspeakers. This implies a higher voltage rail than op amps commonly operate at. This and other high voltage applications can operate off of, and generate lethal voltages. The designer needs to be extremely careful not to create an unsafe product, or be electrocuted while developing it.

The power stage is most often included in the feedback loop of the op amp circuit, so that the closed loop can compensate for power stage errors. This is not always possible if the voltage swing of the output exceeds that of the op amp voltage rails. In these cases, a voltage divided version of the output should be used.

There are three broad categories of booster, the current booster, the voltage booster, and boosters that do both. All of them work on the same principle: anything that is put inside the feedback loop of the op amp will be compensated for — the output voltage will swing to whatever voltage it needs to make the voltage at the buffer op amp inputs equal.

14.7.1 Current Boosters

These usually use some variation of the class-B push-pull amplifier topology (Figure 14–16).

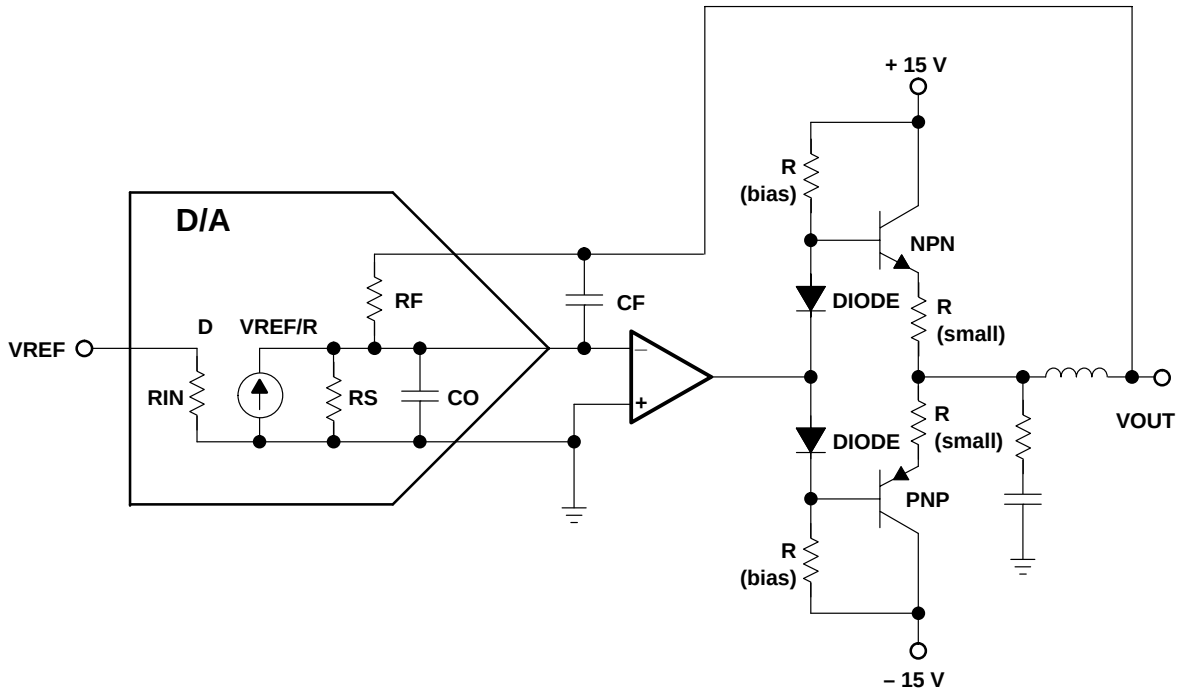


Figure 14–16. D/A Output Current Booster

The circuit in Figure 14–16 has been employed for decades — many resources are available that can be used to design exact component values. It boosts current because the output impedance of the op amp has been bypassed, and used as the driver for the base of the NPN and PNP power transistors. The two diodes compensate for the VBE drop in the transistors, whose bases are biased by two resistors off of the supplies. The output of the booster stage is fed back to the feedback resistor in the D/A to complete the feedback loop. The output impedance of the stage is only limited by the characteristics of the output transistors and small emitter resistors. Modern power transistors have such high frequency response that this circuit may oscillate. The RC snubber network and a small inductor in series with the load can be used to damp the oscillation — or be omitted if oscillation is not a problem. Beware of varying transistor betas, however.

14.7.2 Voltage Boosters

If even more current is needed, or the output voltage swing must be more than ±15 V, the booster stage can be operated at voltages higher than the buffer amplifier potentials. A designer might be tempted to try the circuit of Figure 14–17.

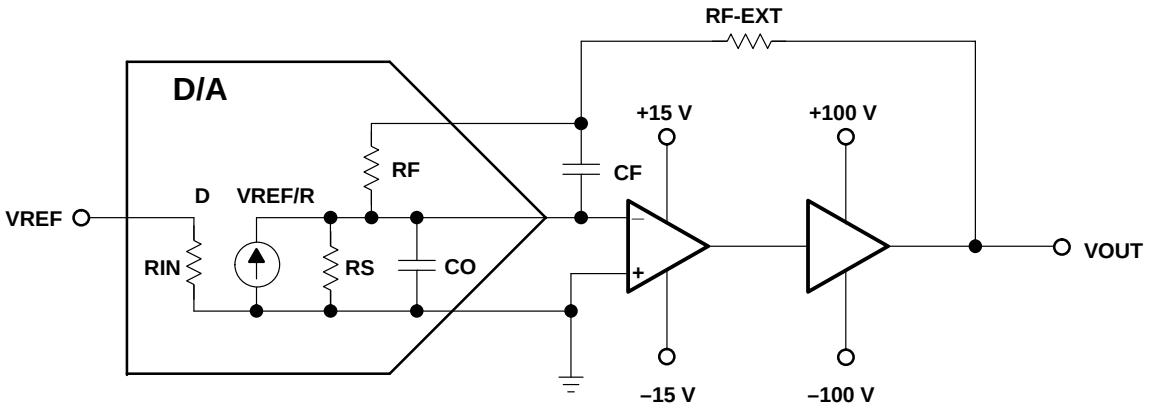


Figure 14–17. Incorrect Method of Increasing Voltage Swing of D/A Converters

Any time there are higher voltage rails on the output section, there are potential hazards. The circuit above illustrates a common misapplication.

- The whole reason for using the booster amp is to allow the V_{OUT} to swing to a ±100-V rail. If this circuit was operated in the unity gain mode (external $R_F = 0$), the VOUT will only swing ±15 V, maximum. There would be no need for the ±100-V rail. That voltage rail is there to allow voltage gain.
- If the circuit is operated with a gain (external R larger than 0), the external R_F adds to the internal R_F to create the gain:

$$\text{Gain} = \frac{R_{F-EXT} + R_{F-INT}}{R_S} \qquad (14\text{–}7)$$

The problem with this is that the wattage of the resistors increases as the external voltage rail increases. The designer has control over the wattage of the external R_F, but has no control whatsoever over internal R_F or R_S. Because these resistors are fabricated on the IC, their wattage is limited. Even if the wattage rating of the internal resistors is meticulously observed, they may have undesirable thermal coefficients if allowed to dissipate that wattage. Resistor self-heating will change the resistance according to its rated temperature coefficient (maximum). The external resistor is sure to have a different thermal coefficient from the internal resistors, causing a gain error. The designer may never have encountered the effects of resistor self-heating before, because through-hole and surface-mount devices have enough bulk to minimize the effect of self-heating. At the geometries present on IC D/As, resistor self-heating is a much more pronounced effect. It will produce a nonlinearity error in the D/A output.

This effect is most pronounced in high-resolution converters, where the geometry is the smallest. The designer, therefore, must limit the current in the feedback resistor if at all possible. Figure 14–18 shows a method of achieving gain control while keeping the high current path out of the internal feedback resistor:

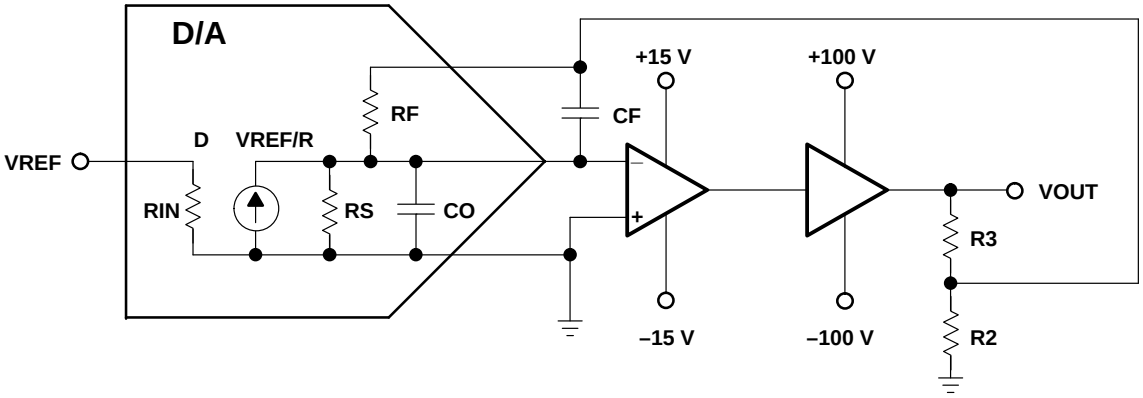


Figure 14–18. Correct Method of Increasing Voltage Range

In Figure 14–18:

- R3 and R2 are selected to ensure that the feedback voltage to the D/A internal R_F can never exceed the D/A rated limits.
- R3 and R2, of course, have to be the correct power rating. R2, in particular, has to be carefully selected. If it burns out, the feedback loop will present hazardous voltages to the D/A. R3, which drops the bulk of the voltage, will have to dissipate considerable wattage.

If the combination of voltage swing and power ratings cannot be balanced to achieve a working design, the only choice left to the designer will be to break the feedback loop and live with the loss of accuracy. For ac applications, this may be acceptable.

14.7.3 Power Boosters

The two types of boosters above can, of course, be combined to produce more power. In audio applications, for example, a ±15-V power supply limits the output power to 112.5 W, absolute maximum, into an 8-Ω load. In order to increase the power, the voltage rails must also be increased, with all of the cautions of the previous paragraph observed.

14.7.4 Single-Supply Operation and DC Offsets

A D/A power circuit is not the right place to try to apply single-supply design techniques. In audio applications, a single-supply design would force a large coupling capacitor, which would distort and limit low frequency response. In dc applications, a dc offset will continually drive a load — which will have to dissipate the excess voltage through its internal resistance as heat.

Nevertheless, there may be applications that require a dc offset. The designer is fortunate in that there is already a precision reference available in the circuit. The reference drives

the resistor network in the D/A, and may be external or internal to it. In most cases, an internal reference is brought out to a pin on the device. It is important for the designer not to excessively load the reference, as that would directly affect D/A accuracy (Figure 14–19).

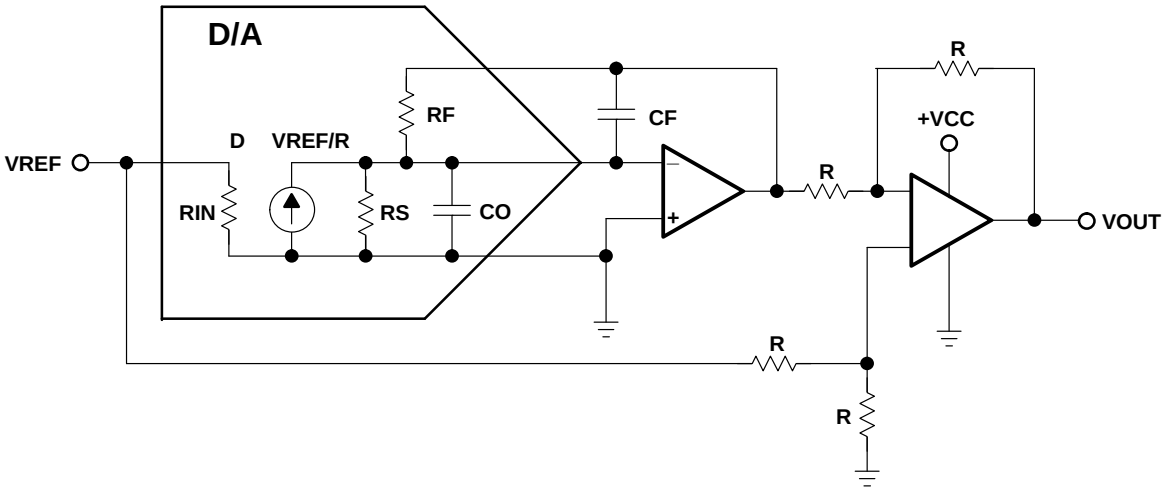


Figure 14–19. Single-Supply DAC Operation

In the circuit in Figure 14–19, the output of the buffer amplifier is shifted up in dc level by ½ VREF (not ½ VCC). VREF was selected because it is much more stable and accurate than VCC. The four resistors in the level shifter circuit must be highly accurate and matched, or this circuit will contribute to gain and offset errors. Thermal errors, however, cannot be compensated for, because the external resistors are probably going to have a different thermal drift than those on the IC. This technique is limited to applications that will see only a small change in ambient temperature.

Chapter 15

Sine Wave Oscillators

Ron Mancini and Richard Palmer

15.1 What is a Sine Wave Oscillator?

Op amp oscillators are circuits that are unstable — not the type that are sometimes unintentionally designed or created in the lab — but circuits intentionally designed to remain in an unstable state. Oscillators are useful for creating uniform signals that are used as a reference in applications such as audio, function generators, digital systems, and communication systems.

Two general classes of oscillators exist: sinusoidal and relaxation. Sinusoidal oscillators consist of amplifiers with RC or LC circuits that have adjustable oscillation frequencies, or crystals that have a fixed oscillation frequency. Relaxation oscillators generate triangular, sawtooth, square, pulse, or exponential waveforms, and they are not discussed here.

Op amp sine wave oscillators operate without an externally applied input signal. Some combination of positive and negative feedback is used to drive the op amp into an unstable state, causing the output to transition back and forth at a continuous rate. The amplitude and the oscillation frequency are set by the arrangement of passive and active components around a central op amp.

Op amp oscillators are restricted to the lower end of the frequency spectrum because op amps do not have the required bandwidth to achieve low phase shift at high frequencies. Voltage-feedback op amps are limited to a the low kHz range since their dominant, open loop pole may be as low as 10 Hz. The new current-feedback op amps have a much wider bandwidth, but they are very hard to use in oscillator circuits because they are sensitive to feedback capacitance and are beyond the scope of this chapter. Crystal oscillators are used in high frequency applications up to the hundreds of MHz range.

15.2 Requirements for Oscillation

The canonical, or simplest form, of a negative feedback system is used to demonstrate the requirements for oscillation to occur. The block diagram of this system is shown in Fig-

ure 15–1, and the corresponding classic expression for a feedback system is shown in Equation 15–1. The derivation and explanation of the block diagram and equation can be found in Chapter 5.

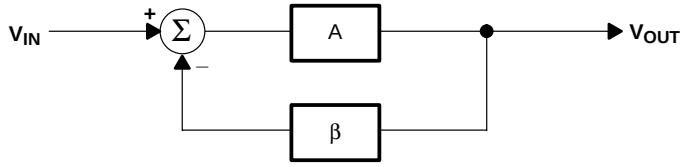


Figure 15–1. Canonical Form of a Feedback System with Positive or Negative Feedback

$$\frac{V_{OUT}}{V_{IN}} = \frac{A}{1 + A\beta} \qquad (15–1)$$

Oscillators do not require an externally applied input signal, but instead use some fraction of the output signal created by the feedback network as the input signal. It is the noise voltage that provides the inital boost signal to the circuit when positive feedback is employed. Over a period of time, the output builds up, oscillating at the frequency set by the circuit components[1].

Oscillation results when the feedback system is not able to find a stable state because its transfer function can not be satisfied. The system becomes unstable when the denominator in Equation 15–1 is 0. When (1 +Aβ) = 0, Aβ = –1. The key to designing an oscillator, then, is to ensure that Aβ = –1. This is called the Barkhausen criterion. This constraint requires the magnitude of the loop gain be 1 with a corresponding phase shift of 180° as indicated by the minus sign. An equivalent expression using complex math is $A\beta = 1\angle{-180}°$ for a negative feedback system. For a positive feedback system, the expression becomes $A\beta = 1\angle 0°$ and the sign is negative in Equation 15–1.

Once the phase shift is 180° and Aβ = |1|, the output voltage of the unstable system heads for infinite voltage in an attempt to destroy the world, and is only prevented from succeeding by an energy-limited power supply. When the output voltage approaches either power rail, the active devices in the amplifiers change gain, causing the value of A to change so the value of $A\beta \neq 1$; thus the charge to infinite voltage slows down and eventually halts. At this point, one of three things can occur. First, nonlinearity in saturation or cutoff can cause the system to become stable and lock up at the power rail. Second, the initial charge can cause the system to saturate (or cutoff) and stay that way for a long time before it becomes linear and heads for the opposite power rail. Third, the system stays linear and reverses direction heading for the opposite power rail. Alternative two produces highly distorted oscillations (usually quasi square waves), and the resulting oscillators are called relaxation oscillators. Alternative three produces sine wave oscillators.

15.3 Phase Shift in the Oscillator

The 180° phase shift in the equation $A\beta = 1\angle -180°$ is introduced by active and passive components. Like any well-designed feedback circuit, oscillators are made dependent on passive component phase shift because it is accurate and almost drift-free. The phase shift contributed by active components is minimized because it varies with temperature, has a wide initial tolerance, and is device dependent. Amplifiers are selected such that they contribute little or no phase shift at the oscillation frequency. These constraints limit the op amp oscillator to relatively low frequencies.

A single pole RL or RC circuit contributes up to 90° phase shift per pole, and because 180° of phase shift is required for oscillation, at least two poles must be used in the oscillator design. An LC circuit has two poles, thus it contributes up to 180° phase shift per pole pair. But LC and LR oscillators are not considered here because low frequency inductors are expensive, heavy, bulky, and very nonideal. LC oscillators are designed in high frequency applications, beyond the frequency range of voltage feedback op amps, where the inductor size, weight, and cost are less significant. Multiple RC sections are used in low frequency oscillator design in lieu of inductors.

Phase shift determines the oscillation frequency because the circuit oscillates at the frequency that accumulates 180° phase shift. The rate of change of phase with frequency, $d\phi/d\omega$, determines frequency stability. When buffered RC sections (an op amp buffer provides high input and low output impedance) are cascaded, the phase shift multiplies by the number of sections, *n* (see Figure 15–2).

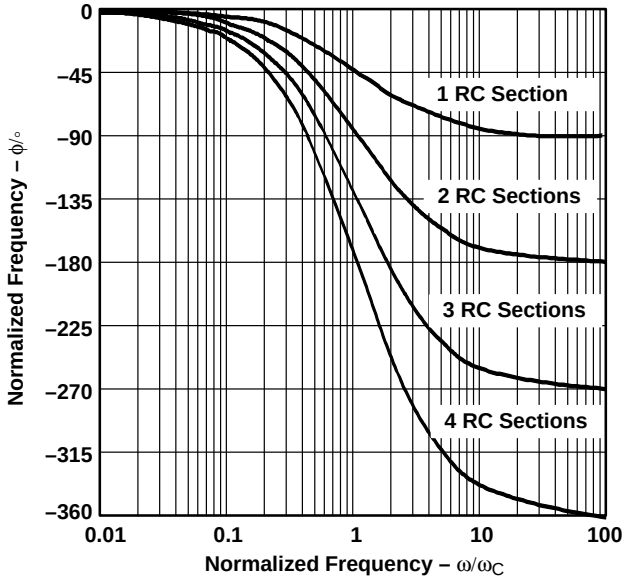


Figure 15–2. Phase Plot of RC Sections

The frequency of oscillation is very dependent upon the change in phase at the point where the phase shift is 180°. A tight frequency specification will require a large change in phase shift, $d\phi$, for a small change in frequency, $d\omega$, at 180°. Figure 15–2 demonstrates that, although two cascaded RC sections eventually provide 180° phase shift, $d\phi/d\omega$ at the oscillator frequency is unacceptably low. Thus, oscillators made with two cascaded RC sections have poor frequency stability. Three equal cascaded RC filter sections have a much higher $d\phi/d\omega$ (see Figure 15–2), and the resulting oscillator has improved frequency stability. Adding a fourth RC section produces an oscillator with an excellent $d\phi/d\omega$ (see Figure 15–2); thus this is the most stable RC oscillator configuration. Four sections are the maximum number used because op amps come in quad packages, and the four-section oscillator section yields four sine waves 45° phase shifted relative to each other. This oscillator can be used to obtain sine/cosine or quadrature sine waves.

Crystal or ceramic resonators make the most stable oscillators because resonators have an extremely high $d\phi/d\omega$ resulting from their nonlinear properties. Resonators are used for high frequency oscillators, but low frequency oscillators do not use resonators because of size, weight, and cost restrictions. Op amps are not generally used with crystal or ceramic resonator oscillators because op amps have low bandwidth. Experience shows that it is more cost effective to build a high frequency crystal oscillator, count the output down, and filter the output to obtain a low frequency than it is to use a low frequency resonator.

15.4 Gain in the Oscillator

The oscillator gain must equal 1 ($A\beta = 1\angle -180°$) at the oscillation frequency. Under normal conditions, the circuit becomes stable when the gain exceeds 1 and oscillations cease. However, when the gain exceeds 1 with a phase shift of –180°, the active device nonlinearity reduces the gain to 1 and the circuit oscillates. The nonlinearity happens when the amplifier swings close to either power rail because cutoff or saturation reduces the active device (transistor) gain. The paradox is that worst case design practice requires nominal gains exceeding 1 for manufacturability, but excess gain causes more distortion of the output sine wave.

When the gain is too low, oscillations cease under worst case conditions, and when the gain is too high, the output wave form looks more like a square wave than a sine wave. Distortion is a direct result of excess gain overdriving the amplifier; thus gain must be carefully controlled in low-distortion oscillators. Phase shift oscillators have distortion, but they achieve low-distortion output voltages because cascaded RC sections act as distortion filters. Also, buffered phase shift oscillators have low distortion because the gain is controlled and distributed among the buffers.

Most circuit configurations require an auxiliary circuit for gain adjustment when low-distortion outputs are desired. Auxiliary circuits range from inserting a nonlinear component in

the feedback loop, to automatic gain control (AGC) loops, to limiting by external components such as resistors and diodes. Consideration must also be given to the change in gain due to temperature variations and component tolerances, and the level of circuit complexity will be determined based on the required stability of the gain. The more stable the gain, the better the purity of the sine wave output.

15.5 Active Element (Op Amp) Impact on the Oscillator

Up to now, it has been assumed that the op amp has an infinite bandwidth and the output is not frequency dependent. In reality, the op amp has many poles, but it has been compensated so that they are dominated by a single pole over the specified bandwidth. Thus $A\beta$, is now frequency dependant through the op amp gain term, A. Equation 15–2 shows this dependance, where a is the maximum open loop gain, ω_a is the dominant pole frequency, and ω is the frequency of the signal. Figure 15–3 graphically depicts the frequency dependence of the op amp gain and phase. The closed loop gain, $A_{CL} = 1/\beta$ does not contain any poles or zeros and its therefore constant over frequency to the point where it impacts the op amp open-loop gain at ω_{3dB}. Here, the signal amplitude is attenuated by 3 dB and the phase shift introduced by the op amp is 45°. The amplitude and phase really begin to change one decade below this point at 0.1 ω_{3dB}, and the phase will continue to shift until it has reached 90° at 10 ω_{3dB}, one decade beyond the 3–dB point. The gain continues to roll off at –20 dB/decade until other poles and zeros come into play. The higher the closed-loop gin, the earlier A_{CL} intercepts the op amp gain.

$$A_{CL} = \frac{A}{1 + A\beta} = \frac{\frac{1}{\beta}}{1 + \frac{1}{A\beta}} = \frac{A_{CLideal}}{1 + \frac{A_{CLideal}}{A_{OL}}} \qquad (15–2)$$

The phase shift contributed by the op amp will impact the performance of the oscillator circuit by lowering the oscillation frequency, and the reduction in A_{CL} can make $A\beta < 1$ and the oscillator will not oscillate.

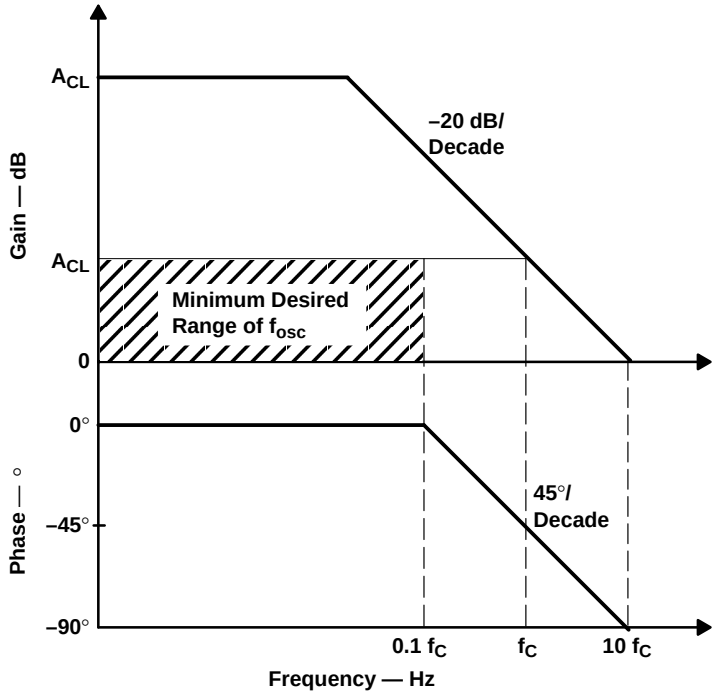


Figure 15–3. Op Amp Frequency Response

Most op amps are compensated and may have more than the 45° of phase shift at ω_{3dB}. The op amp should therefore be chosen with a gain bandwidth that is at least one decade above the oscillation frequency, as shown by the shaded area of Figure 15–3. The Wien bridge requires a gain bandwidth greater than $43\omega_{OSC}$ to maintain the gain and frequency within 10% of the ideal values [2]. Figure 15–4 compares the output distortion versus frequency of an LM328, a TLV247x, and a TLC071 op amp, which have bandwidths of 0.4 MHz and 2.8 MHz, and 10 MHz, respectively, in a Wien bridge oscillator with nonlinear feedback (see Section 15.7.1 for the circuit and transfer function) The oscillation frequency ranged from 16 Hz to 160 kHz. The graph illustrates the importance of choosing the correct op smp for the application. The LM328 achieved a maximum oscillation of 72 kHz and was attenuated more than 75%, while the TLV247x achieved 125 kHz with 18% attenuation. The wide bandwidth of the TLC071 provided a 138 kHz oscillation frequency with a mere 2% attenuation. The op amp must be chosen with the proper bandwidth or the output may oscillate at a frequency well below the design specification.

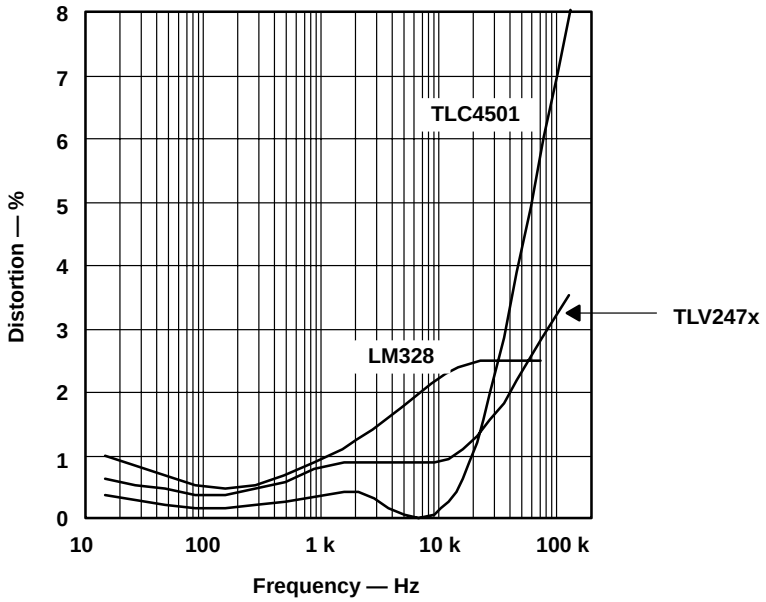


Figure 15–4. Op Amp Bandwidth and Oscillator Output

Care must be taken when using large feedback resistors since they interact with the input capacitance of the op amp to create poles with negative feedback, and both poles and zeros with positive feedback. Large resistor values can move these poles and zeros into the proximity of the oscillation frequency and impact the phase shift [3].

A final consideration is given to the slew-rate limitation of the op amp. The slew rate must be greater than $2\pi V_P f_0$, where V_P is the peak output voltage and f_0 is the oscillation frequency, or distortion of the output signal will result.

15.6 Analysis of the Oscillator Operation (Circuit)

Oscillators are created using various combinations of positive and negative feedback. Figure 15–5 shows the basic negative feedback amplifier block diagram with a positive feedback loop added. When positive and negative feedback are used, the gain of the negative feedback path is combined into one gain term (representing the closed loop gain) and Figure 15–5 reduces to Figure 15–1. The positive feedback network is then represented by $\beta = \beta_2$ and subsequent analysis is simplified. When negative feedback is used, then the positive feedback loop can be ignored since β_2 is 0. The case of positive and negative feedback combined is covered here since the negative feedback case was reviewed in Chapters 5 and 6.

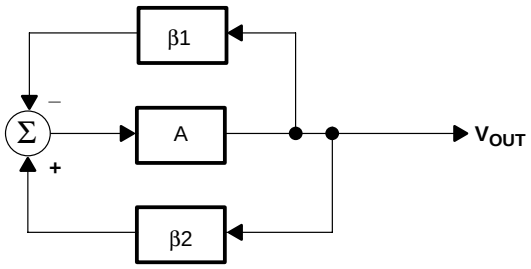
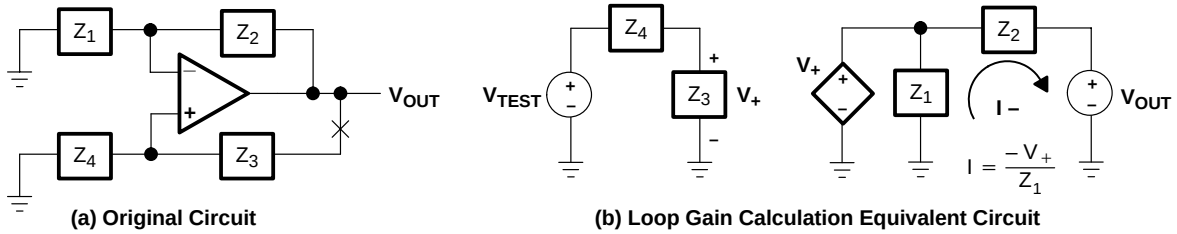


Figure 15–5. Block Diagram of an Oscillator: a) Positive and Negative Feedback Loops, and b) Simplified Diagram w

A general form of an op amp with positive and negative feedback is shown in Figure 15–6a. The first step is to break the loop at some point without altering the gain of the circuit. The positive feedback loop is broken at the point marked with an X. A test signal (V_{TEST}) is applied to the broken loop and the resulting output voltage (V_{OUT}) is measured with the equivalent circuit shown in Figure 15–6b.



(a) Original Circuit **(b) Loop Gain Calculation Equivalent Circuit**

Figure 15–6. Amplifier with Positive and Negative Feedback

V_+ is calculated first in Equation 15–3, and then is treated as an input signal to a noninverting amplifier, resulting in Equation 15–4. Equation 15–3 is substituted for V_+ into Equation 15–4 to get the transfer function in Equation 15–5. The actual circuit elements are then substituted for each impedance and the equation is simplified. These equations are valid when the op amp open-loop gain is large and the oscillation frequency is < 0.1 ω_{3dB}.

$$V_+ = V_{TEST}\left(\frac{Z_4}{Z_3 + Z_4}\right) \tag{15–3}$$

$$V_{OUT} = V_+\left(\frac{Z_1 + Z_2}{Z_1}\right) \tag{15–4}$$

$$\frac{V_{OUT}}{V_{TEST}} = \left(\frac{Z_3}{Z_3 + Z_4}\right)\left(\frac{Z_1 + Z_2}{Z_1}\right) \tag{15–5}$$

Phase shift oscillators generally use negative feedback, so the positive feedback factor (β_2) becomes zero. Oscillator circuits such as the Wien bridge use both negative (β_1) and positive (β_2) feedback to achieve a constant state of oscillation. This circuit is analyzed in detain in Section 15.7.1 using Equation 15–5.

15.7 Sine Wave Oscillator Circuits

There are many types of sine wave oscillator circuits and variations of these circuits — the choice depends upon the frequency and the desired purity of the output waveform. The focus of this section is on the more prominent oscillator circuits: Wien bridge, phase shift, and quadrature. The transfer function is derived for each case using the techniques described in Section 15.6 of this chapter and in Chapters 3, 5, and 6.

15.7.1 Wien Bridge Oscillator

The Wien bridge is one of the simplest and best known oscillators and is used extensively in circuits for audio applications. Figure 15–7 shows the basic Wien bridge circuit configuration. This circuit has only a few components and good frequency stability. The major drawback of the circuit is that the output amplitude is at the rails, saturating the op amp output transistors and causing high output distortion. Taming this distortion is more of a challenge than getting the circuit to oscillate. There are a couple of ways to minimize this effect, which will be covered later. It is now time to analyze this circuit and come up with the transfer function.

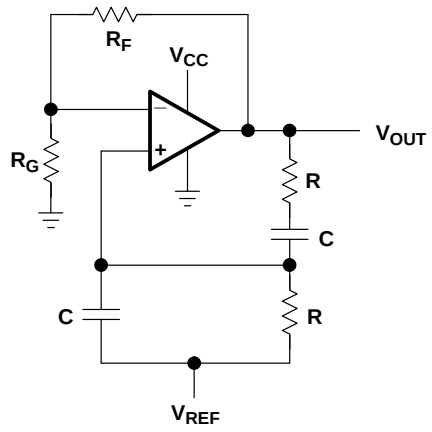


Figure 15–7. Wien Bridge Circuit Schematic

The Wien bridge circuit is of the form that is detailed in Section 15.6. The transfer function for the circuit is created using the technique described in that section. It is readily apparent that $Z_1 = R_G$, $Z_2 = R_F$, $Z_3 = (R_1 + 1/sC_1)$ and $Z_4 = (R_2 \| 1/sC_2)$. The loop is broken between

the output and Z_1, V_{TEST} is applied to Z_1, and V_{OUT} is calculated. The positive feedback voltage, V_+, is calculated first in Equations 15–6 through 15–8. Equation 15–6 shows the simple voltage divider at the noninverting input. Each term is then multiplied by (R_2C_2s + 1) and divided by R_2 to get Equation 15–7.

$$V_+ = V_{TEST}\left(\frac{Z_4}{Z_3 + Z_4}\right) = V_{TEST}\left[\frac{\left(\frac{R_2}{R_2C_2s+1}\right)}{\left(\frac{R_2}{R_2C_2s+1}\right) + \left(R_1 + \frac{1}{C_1s}\right)}\right] \quad (15–6)$$

$$\frac{V_+}{V_{TEST}} = \frac{1}{1 + R_1C_2S + \frac{R_1}{R_2} + \frac{1}{R_2C_1s} + \frac{C_2}{C_1}} \quad (15–7)$$

Substitute $s = j\omega_0$, where ω_0 is the oscillation frequency, $\omega_1 = 1/R_1C_2$, and $\omega_2 = 1/R_2C_1$ to get Equation 15–8.

$$\frac{V_+}{V_{TEST}} = \frac{1}{1 + \frac{R_1}{R_2} + \frac{C_2}{C_1} + j\left(\frac{\omega_0}{\omega_1} - \frac{\omega_2}{\omega_0}\right)} \quad (15–8)$$

Some interesting relationships now become apparent. The capacitor in the zero, represented by ω_1, and the capacitor in the pole, represented by ω_2, must each contribute 90° of phase shift toward the 180° required for oscillation at ω_0. This requires that $C_1 = C_2$ and $R_1 = R_2$. Setting ω_1 and ω_2 equal to ω_0 cancels the frequency terms, ideally removing any change in amplitude with frequency since the pole and zero negate one another. An overall feedback factor of $\beta = 1/3$ is the result (Equation 15–9).

$$\frac{V_+}{V_{TEST}} = \frac{1}{1 + \frac{R}{R} + \frac{C}{C} + j\left(\frac{\omega_0}{\omega} - \frac{\omega}{\omega_0}\right)} = \frac{1}{3 + j\left(\frac{\omega_0}{\omega_0} - \frac{\omega_0}{\omega_0}\right)} = \frac{1}{3} \quad (15–9)$$

The gain of the negative feedback portion, A, of the circuit must then be set such that $|A\beta| = 1$, requiring A = 3. R_F must be set to twice the value of R_G to satisfy the condition. The op amp in Figure 15–7 is single supply, so a dc reference voltage, V_{REF}, must be applied to bias the output for full-scale swing and minimal distortion. Applying V_{REF} to the positive input through R_2 restricts dc current flow to the negative feedback leg of the circuit. V_{REF} was set at 0.833V to bias the output at the midrail of the single supply, rail-to-rail input and output amplifier, or 2.5 V. See Cahpter 4 for details on dc biasing single-supply op amps. V_{REF} is shorted to ground for split supply applications.

The final circuit is shown in Figure 15–8, with component values selected to provide an oscillation frequency of $\omega_0 = 2\pi f_0$, where $f_0 = 1/(2\pi RC) = 15.9$ kHz. The circuit oscillated at 1.57 kHz due to slightly varying component values with 2% distortion. This high value is due to the extensive clipping of the output signal at both supply rails, producing several

large odd and even harmonics. The feedback resistor was then adjusted ±1%. Figure 15–9 shows the output voltage waveforms. The distortion grew as the saturation increased with increasing R_F, and oscillations ceased when R_F was decreased by more than 0.8%.

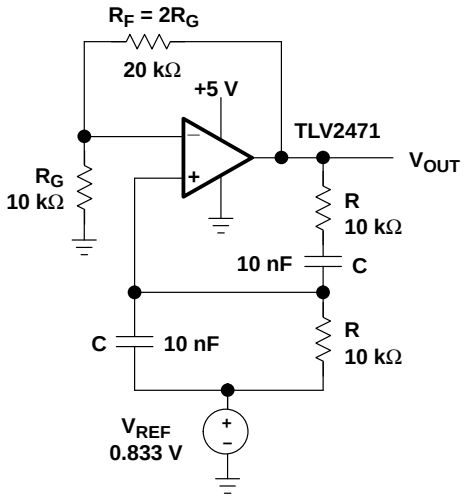


Figure 15–8. Final Wien Bridge Oscillator Circuit

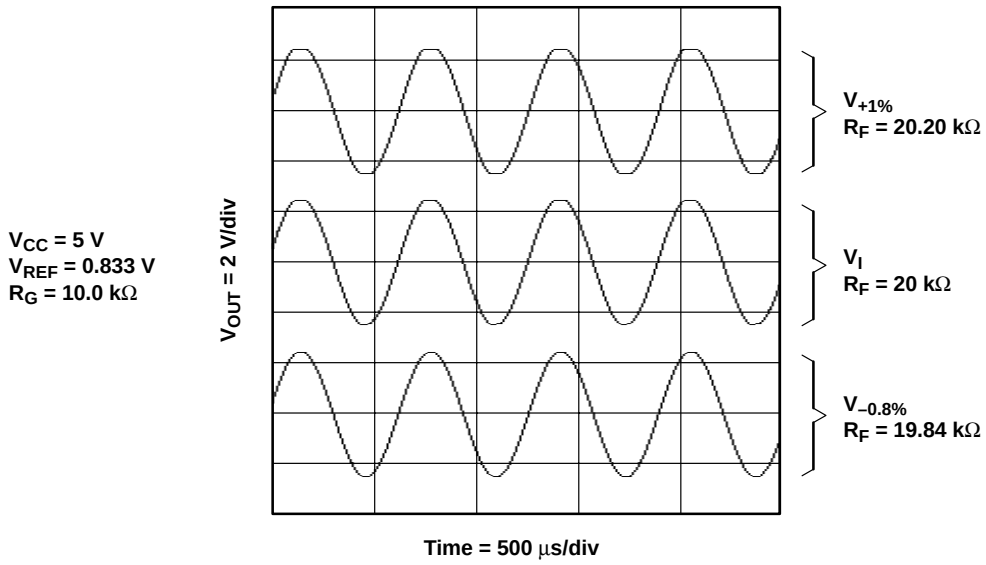


Figure 15–9. Wien Bridge Output Waveforms

Applying nonlinear feedback can minimize the distortion inherent in the basic Wien bridge circuit. A nonlinear component such as an incandescent lamp can be substituted into

the circuit for R_G as shown in Figure 15–10. The lamp resistance, R_{LAMP}, is nominally selected as half the feedback resistance, R_F, at the lamp current established by R_F and R_{LAMP}. When the power is first applied the lamp is cool and its resistance is small, so the gain is large (> 3). The current heats up the filament and the resistance increases, lowering the gain. The nonlinear relationship between the lamp current and resistance keeps output voltage changes small. Figure 15–11 shows the output of this amplifier with a distortion of 1% for f_{OSC} = 1.57 kHz. The distortion for this variation is reduced over the basic circuit by avoiding hard saturation of the op amp transistors.

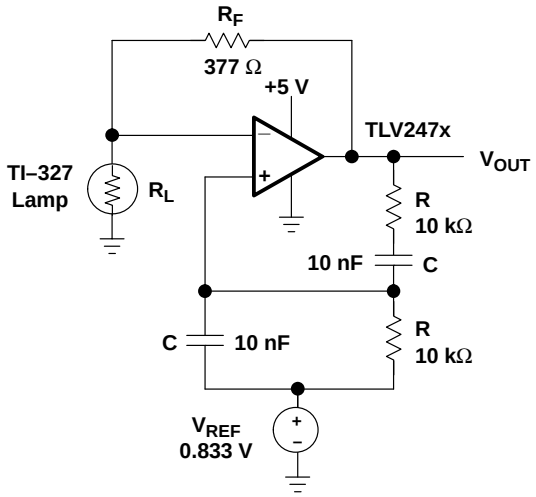


Figure 15–10. Wien Bridge Oscillator with Nonlinear Feedback

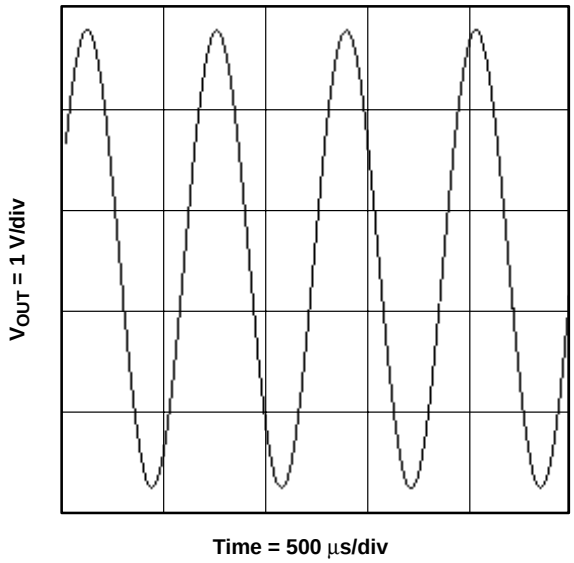


Figure 15–11. Output of the Circuit in Figure 15–10

The impedance of the lamp is mostly due to thermal effects. The output amplitude is then very temperature sensitive and will tend to drift. The gain must be set higher than 3 to compensate for any temperature variations, which increases the distortion in the circuit [4]. This type of circuit is useful when the temperature does not fluctuate over a wide range or when used in conjunction with an amplitude limiting circuit.

The lamp has an effective low frequency thermal time constant, $t_{thermal}$ (4). As f_{OSC} approaches $t_{thermal}$, distortion is greatly increased. Several lamps can be placed in series to increase $t_{thermal}$ and reduce distortion. The drawbacks are that the time required for oscillations to stabilize is increased and the output amplitude is reduced.

An automatic gain control (AGC) circuit must be used when neither of the two previous circuits yield low distortion. A typical Wien bridge oscillator with an AGC circuit is shown in Figure 15–12, with the output waveform of the circuit shown in Figure 15–13. The AGC is used to stabilize the magnitude of the sinusoidal output to an optimum gain level. The JFET serves as the AGC element, providing excellent control because of the wide range of the drain to source resistance (R_{DS}), which is controlled by the gate voltage. The JFET gate voltage is 0 V when the power is applied, and the JFET turns on with low R_{DS}. This places $R_{G2}+R_S+R_{DS}$ in parallel with R_{G1}, raising the gain to 3.05, and oscillations begin and gradually build up. As the output voltage gets large, the negative swing turns the diode on and the sample is stored on C_1, which provides a dc potential to the gate of Q_1. Resistor R_1 limits the current and establishes the time constant for charging C_1, which should be much greater than f_{OSC}. When the output voltage drifts high, R_{DS} increases, lowering the gain to a minimum of 2.87 ($1+R_F/R_{G1}$). The output stabilizes when the gain reaches 3. The distortion of the AGC is 0.8%, which is due to slight clipping at the positive rail.

The circuit of Figure 15–12 is biased with V_{REF} for a single-supply amplifier. A zener diode can be placed in series with D1 to limit the positive swing of the output and reduce distortion. A split supply can be easily implemented by grounding all points connected to V_{REF}. There are a wide variety of Wien bridge variations that exist to more precisely control the amplitude and allow selectable or even variable oscillation frequencies. Some circuits use diode limiting in place of a nonlinear feedback component. The diodes reduce the distortion by providing a soft limit for the output voltage.

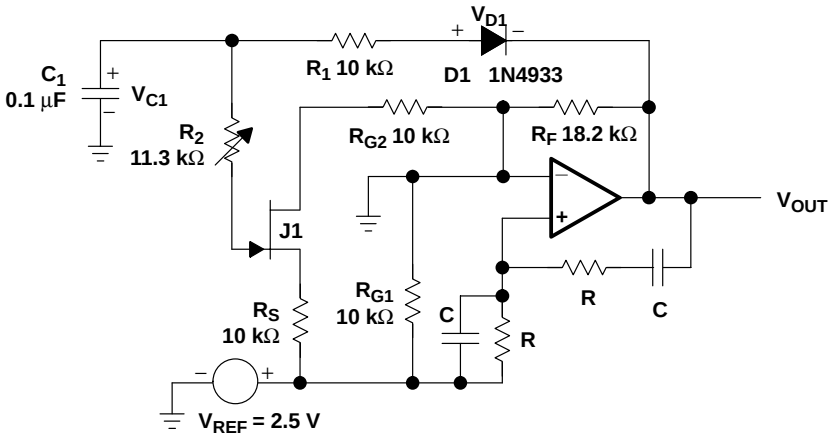


Figure 15–12. Wien Bridge Oscillator with AGC

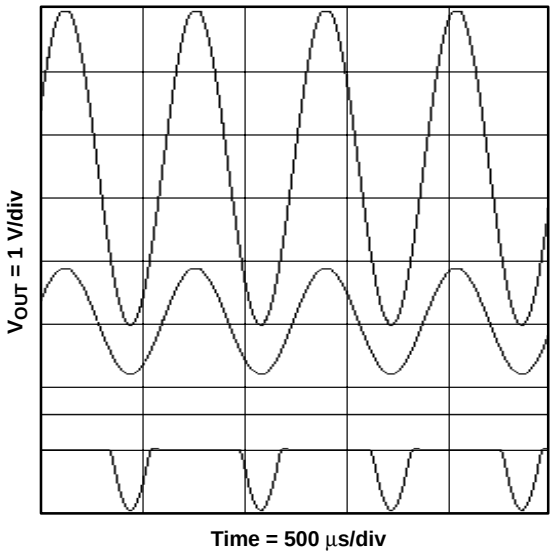


Figure 15–13. Output of the Circuit in Figure 15–12

15.7.2 Phase Shift Oscillator, Single Amplifier

Phase shift oscillators have less distortion than the Wien bridge oscillator, coupled with good frequency stability. A phase shift oscillator can be built with one op amp as shown in figure 15–14. Three RC sections are cascaded to get the steep $d\phi/d\omega$ slope as described in Section 15–3 to get a stable oscillation frequency. Any less and the oscillation frequency is high and interferes with the op amp BW limitations.

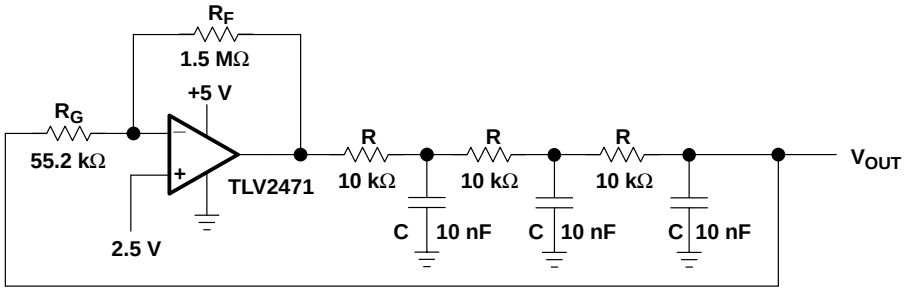


Figure 15–14. Phase Shift Oscillator (Single Op Amp)

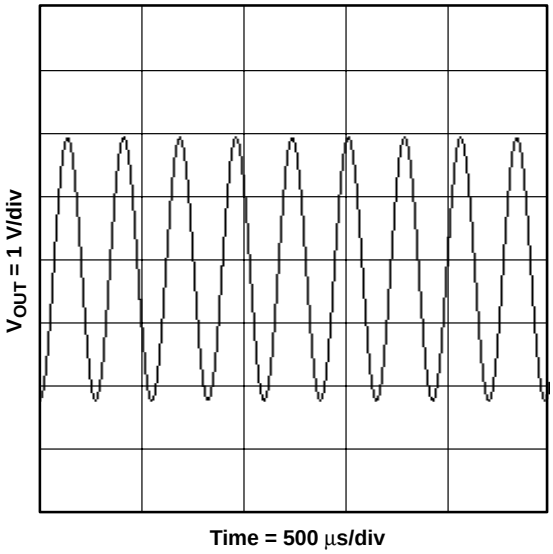


Figure 15–15. Output of the Circuit in Figure 15–14

$$A\beta = A\left(\frac{1}{RCs + 1}\right)^3 \quad (15–10)$$

The normal assumption is that the phase shift sections are independent of each other. Then Equation 15–10 is written. The loop phase shift is –180° when the phase shift of each section is –60°, and this occurs when $\omega = 2\pi f = 1.732/RC$ because the tangent of 60° = 1.732. The magnitude of β at this point is $(1/2)^3$, so the gain, A, must be equal to 8 for the system gain to be equal to one.

The oscillation frequency with the component values shown in Figure 15–14 is 3.76 kHz rather than the calculated oscillation frequency of 2.76 kHz as shown in Figure 15–14. Also, the gain required to start oscillation is 27 rather than the calculated gain of 8. These

discrepancies are partially due to component variations, but the biggest contributing factor is the incorrect assumption that the RC sections do not load each other. This circuit configuration was very popular when active components were large and expensive. But now op amps are inexpensive, small, and come four in a package, so the single op amp phase shift oscillator is losing popularity. The output distortion is a low 0.46%, considerably less than the Wein bridge circuit without amplitude stabilization.

15.7.3 Phase Shift Oscillator, Buffered

The buffered phase shift oscillator is much improved over the unbuffered version, the cost being a higher component count. The buffered phase shift oscillator is shown in Figure 15–16 and the resulting output waveform in Figure 15–17. The buffers prevent the RC sections from loading each other, hence the buffered phase shift oscillator performs closer to the calculated frequency and gain. The gain setting resistor, R_G, loads the third RC section. If the fourth buffer in a quad op amp buffers this RC section, the performance becomes ideal. Low-distortion sine waves can be obtained from either phase shift oscillator design, but the purest sine wave is taken from the output of the last RC section. This is a high-impedance node, so a high impedance input is mandated to prevent loading and frequency shifting with load variations.

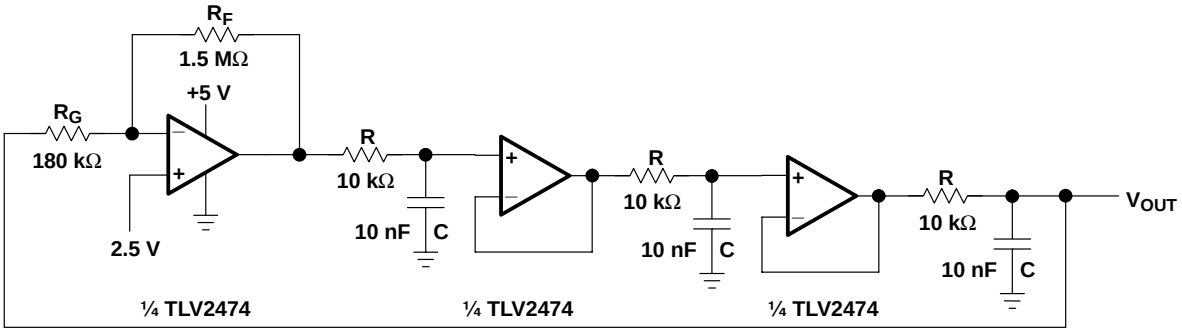


Figure 15–16. Phase Shift Oscillator, Buffered

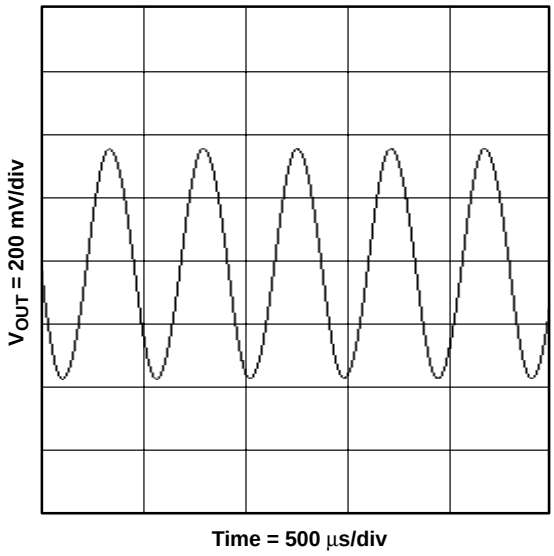


Figure 15–17. Output of the Circuit Figure 15–16

15.7.4 Bubba Oscillator

The bubba oscillator in Figure 15–18 is another phase shift oscillator, but it takes advantage of the quad op amp package to yield some unique advantages. Four RC sections require 45° phase shift per section, so this oscillator has an excellent $d\phi/dt$ resulting in minimized frequency drift. The RC sections each contribute 45° phase shift, so taking outputs from alternate sections yields low impedance quadrature outputs. When an output is taken from each op amp, the circuit delivers four 45° phase shifted sine waves. The loop equation is given in Equation 15–11. When $\omega = 1/RCs$, Equation 15–11 reduces to Equations 15–12 and 15–13.

$$A\beta = A\left(\frac{1}{RCs + 1}\right)^4 \tag{15–11}$$

$$|\beta| = \left|\left(\frac{1}{j + 4}\right)^4 = \frac{1}{\sqrt{2}^4} = \frac{1}{4}\right| \tag{15–12}$$

$$\phi = \tan^{-1}(1) = 45^\circ \tag{15–13}$$

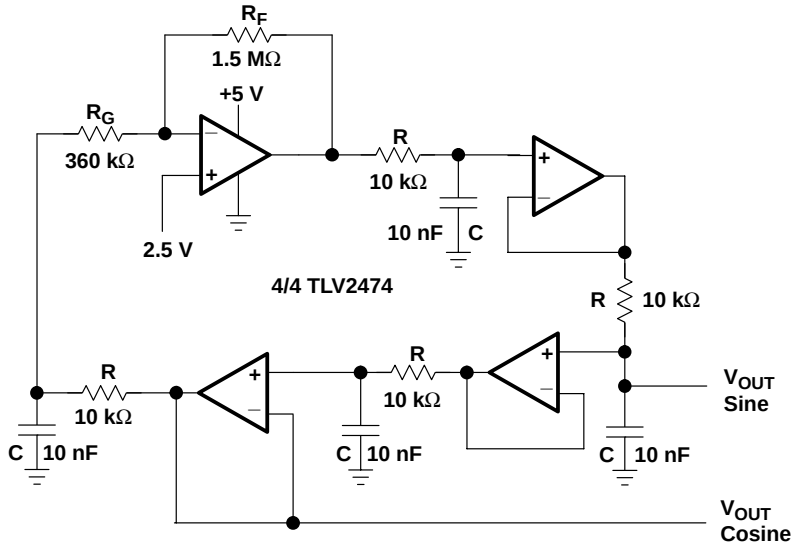


Figure 15–18. Bubba Oscillator

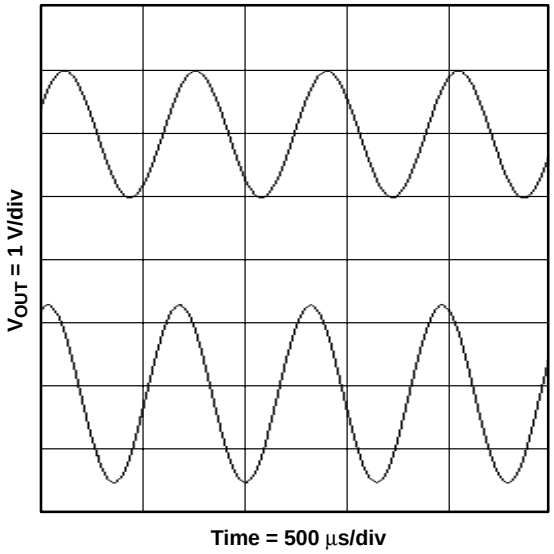


Figure 15–19. Output of the Circuit in Figure 15–18.

The gain, A, must equal 4 for oscillation to occur. The test circuit oscillated at 1.76 kHz rather than the ideal frequency of 1.72 kHz when the gain was 4.17 rather than the ideal gain 4. The output waveform is shown in Figure 15–19. Distortion is 1% for $V_{OUTSINE}$ and 0.1% for $V_{OUTCOSINE}$. With low gain, A, and low bias current op amps, the gain setting

resistor, R_G, does not load the last RC section, thus ensuring oscillator frequency accuracy. Very low distortion sine waves can be obtained from the junction of R and R_G. When low-distortion sine waves are required at all outputs, the gain should be distributed between all of the op amps. The noninverting input of the gain op amp is biased at 0.5 V to set the quiescent output voltage at 2.5 V for single supply operation, and should be ground for split supply op amps. Gain distribution requires biasing of the other op amps, but it has no effect on the oscillator frequency.

15.7.5 Quadrature Oscillator

The quadrature oscillator shown in Figure 15–20 is another type of phase shift oscillator, but the three RC sections are configured so each section contributes 90° of phase shift. This provides both sine and cosine waveform outputs (the outputs are quadrature, or 90° apart), which is a distinct advantage over other phase shift oscillators. The idea of the quadrature oscillator is to use the fact that the double integral of a sine wave is a negative sine wave of the same frequency and phase. The phase of the second integrator is then inverted and applied as positive feedback to induce oscillation [6].

The loop gain is calculated in Equation 15–14. When $R_1C_1 = R_2C_2 = R_3C_3$, Equation 15–14 reduces to Equation 15–15. When $\omega = 1/RC$, Equation 15–14 reduces to $1\angle -180$, so oscillation occurs at $\omega = 2\pi f = 1/RC$. The test circuit oscillated at 1.65 kHz rather than the calculated 1.59 kHz, as shown in Figure 15–21. This discrepancy is attributed to component variations. Both outputs have relatively high distortion that can be reduced with a gain stabilizing circuit. The sine output had 0.846% distortion and the cosine output had 0.46% distortion. Adjusting the gain can increase the amplitudes. The cost is bandwidth.

$$A\beta = A\left(\frac{1}{R_1C_1s}\right)\left(\frac{R_3C_3s + 1}{R_3C_3s\left(R_2C_2s + 1\right)}\right) \tag{15–14}$$

$$A\beta = A\left(\frac{1}{RCs}\right)^2 \tag{15–15}$$

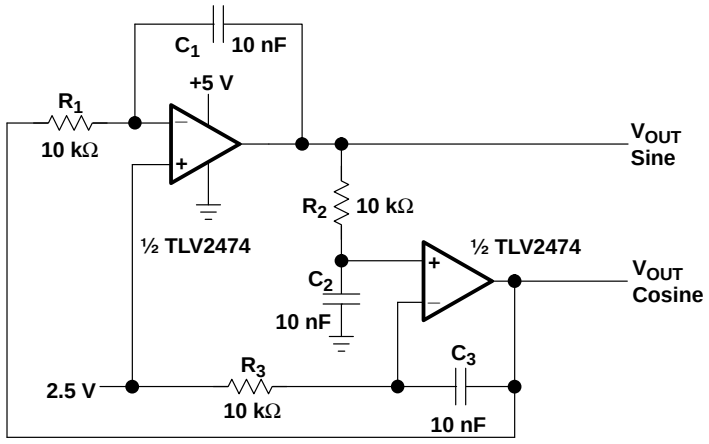


Figure 15–20. Quadrature Oscillator

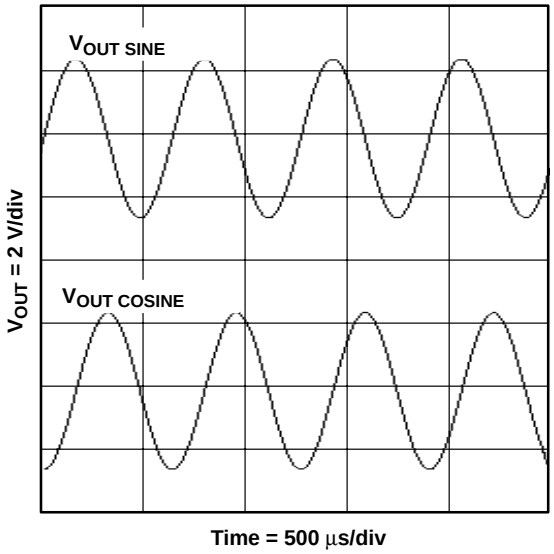


Figure 15–21. Output of the Circuit in Figure 15–20

15.7.6 Conclusion

Op amp oscillators are restricted to the lower end of the frequency spectrum because they do not have the required bandwidth to achieve low phase shift at high frequencies. The new current-feedback op amps have a much greater bandwidth than the voltage-feedback op amps, but are very difficult to use in oscillator circuits because of their sensitivity to feedback capacitance. Voltage-feedback op amps are limited to tens of Hz (at the most!) because of their low frequency rolloff. The bandwidth is reduced when op amps are cascaded due to the multiple contribution of phase shift.

The Wien bridge oscillator has few parts and good frequency stability, but the basic circuit has a high output distortion. AGC improves the distortion considerably, particularly at the lower frequency range. Nonlinear feedback offers the best performance over the mid and upper frequency ranges. The phase shift oscillator has lower output distortion and, without buffering, requires a high gain, which limits the use to very low frequencies. Decreasing cost of op amps and components has reduced the popularity of the phase shift oscillators. The quadrature oscillator only requires two op amps, has reasonable distortion, and offers both sine and cosine waveforms. The drawback is the low amplitude, which may require a higher gain and a reduction in bandwidth, or an additional gain stage.

May your oscillators always oscillate, and your amplifiers always amplify.

15.8 References

[1] Gottlieb, Irving M., *Practical Oscillator Handbook*, Newnes, 1997.

[2] Kennedy, E. J., *Operational Amplifier Circuits, Theory and Applications,* Holt Rhienhart and Winston, 1988.

[3] Graeme, Jerald, *Optimizing Op Amp Performance*, McGraw Hill Book Company, 1997.

[4] Philbrick Researches, Inc., *Applications Manual for Computing Amplifiers*, Nimrod Press, Inc., 1966.

[5] Graf, Rudolf F., *Oscillator Circuits*, Newnes, 1997.

[6] Graeme, Jerald, *Applications of Operational Amplifiers, Third Generation Techniques*, McGraw Hill Book Company, 1973.

Chapter 16

Active Filter Design Techniques

Thomas Kugelstadt

16.1 Introduction

What is a filter?

> *A filter is a device that passes electric signals at certain frequencies or frequency ranges while preventing the passage of others*. — Webster.

Filter circuits are used in a wide variety of applications. In the field of telecommunication, band-pass filters are used in the audio frequency range (0 kHz to 20 kHz) for modems and speech processing. High-frequency band-pass filters (several hundred MHz) are used for channel selection in telephone central offices. Data acquisition systems usually require anti-aliasing low-pass filters as well as low-pass noise filters in their preceding signal conditioning stages. System power supplies often use band-rejection filters to suppress the 60-Hz line frequency and high frequency transients.

In addition, there are filters that do not filter any frequencies of a complex input signal, but just add a linear phase shift to each frequency component, thus contributing to a constant time delay. These are called all-pass filters.

At high frequencies (> 1 MHz), all of these filters usually consist of passive components such as inductors (L), resistors (R), and capacitors (C). They are then called LRC filters.

In the lower frequency range (1 Hz to 1 MHz), however, the inductor value becomes very large and the inductor itself gets quite bulky, making economical production difficult.

In these cases, active filters become important. Active filters are circuits that use an operational amplifier (op amp) as the active device in combination with some resistors and capacitors to provide an LRC-like filter performance at low frequencies (Figure 16–1).

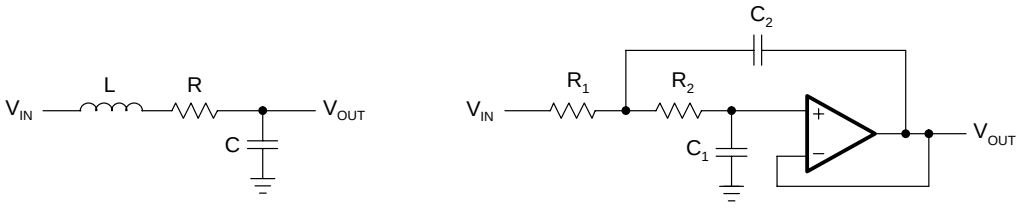


Figure 16–1. Second-Order Passive Low-Pass and Second-Order Active Low-Pass

This chapter covers active filters. It introduces the three main filter optimizations (Butterworth, Tschebyscheff, and Bessel), followed by five sections describing the most common active filter applications: low-pass, high-pass, band-pass, band-rejection, and all-pass filters. Rather than resembling just another filter book, the individual filter sections are written in a cookbook style, thus avoiding tedious mathematical derivations. Each section starts with the general transfer function of a filter, followed by the design equations to calculate the individual circuit components. The chapter closes with a section on practical design hints for single-supply filter designs.

16.2 Fundamentals of Low-Pass Filters

The most simple low-pass filter is the passive RC low-pass network shown in Figure 16–2.

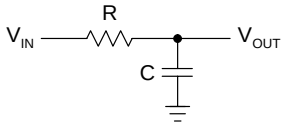


Figure 16–2. First-Order Passive RC Low-Pass

Its transfer function is:

$$A(s) = \frac{\frac{1}{RC}}{s + \frac{1}{RC}} = \frac{1}{1 + sRC}$$

where the complex frequency variable, $s = j\omega+\sigma$, allows for any time variable signals. For pure sine waves, the damping constant, σ, becomes zero and $s = j\omega$.

For a normalized presentation of the transfer function, *s* is referred to the filter's corner frequency, or –3 dB frequency, $\omega_{C,}$ and has these relationships:

$$s = \frac{s}{\omega_C} = \frac{j\omega}{\omega_C} = j\frac{f}{f_C} = j\Omega$$

With the corner frequency of the low-pass in Figure 16–2 being $f_C = 1/2\pi RC$, *s* becomes $s = sRC$ and the transfer function A(s) results in:

$$A(s) = \frac{1}{1 + s}$$

The magnitude of the gain response is:

$$|A| = \frac{1}{\sqrt{1 + \Omega^2}}$$

For frequencies $\Omega >> 1$, the rolloff is 20 dB/decade. For a steeper rolloff, ***n*** filter stages can be connected in series as shown in Figure 16–3. To avoid loading effects, op amps, operating as impedance converters, separate the individual filter stages.

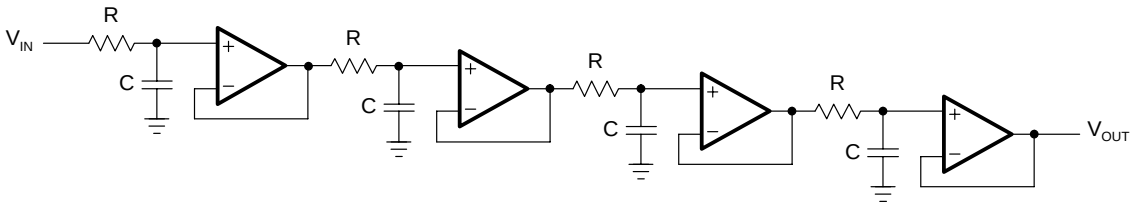


Figure 16–3. Fourth-Order Passive RC Low-Pass with Decoupling Amplifiers

The resulting transfer function is:

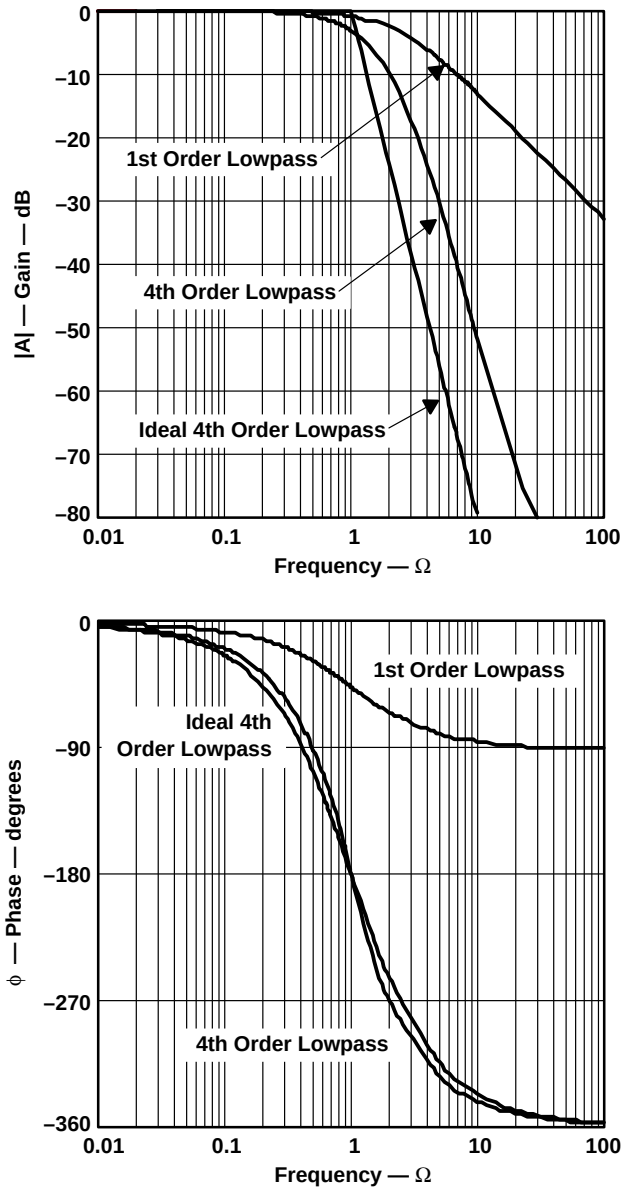
$$A(s) = \frac{1}{(1 + \alpha_1 s)(1 + \alpha_2 s) \ldots (1 + \alpha_n s)}$$

In the case that all filters have the same cut-off frequency, f_C, the coefficients become

$\alpha_1 = \alpha_2 = \ldots \alpha_n = \alpha = \sqrt{\sqrt{2} - 1}$, and f_C of each partial filter is $1/\alpha$ times higher than f_C of the overall filter.

Figure 16–4 shows the results of a fourth-order RC low-pass filter. The rolloff of each partial filter (Curve 1) is –20 dB/decade, increasing the roll-off of the overall filter (Curve 2) to 80 dB/decade.

> **Note:**
>
> **Filter response graphs plot gain versus the normalized frequency axis Ω ($\Omega = f/f_C$).**



Note: Curve 1: 1st-order partial low-pass filter, Curve 2: 4th-order overall low-pass filter, Curve 3: Ideal 4th-order low-pass filter

Figure 16–4. Frequency and Phase Responses of a Fourth-Order Passive RC Low-Pass Filter

The corner frequency of the overall filter is reduced by a factor of $\alpha \approx 2.3$ times versus the –3 dB frequency of partial filter stages.

In addition, Figure 16–4 shows the transfer function of an ideal fourth-order low-pass function (Curve 3).

In comparison to the ideal low-pass, the RC low-pass lacks in the following characteristics:

- The passband gain varies long before the corner frequency, f_C, thus amplifying the upper passband frequencies less than the lower passband.
- The transition from the passband into the stopband is not sharp, but happens gradually, moving the actual 80-dB roll off by 1.5 octaves above f_C.
- The phase response is not linear, thus increasing the amount of signal distortion significantly.

The gain and phase response of a low-pass filter can be optimized to satisfy *one* of the following three criteria:

1) A maximum passband flatness,
2) An immediate passband-to-stopband transition,
3) A linear phase response.

For that purpose, the transfer function must allow for complex poles and needs to be of the following type:

$$A(s) = \frac{A_0}{(1 + a_1 s + b_1 s^2)(1 + a_2 s + b_2 s^2) \ldots (1 + a_n s + b_n s^2)} = \frac{A_0}{\prod_i (1 + a_i s + b_i s^2)}$$

where A_0 is the passband gain at dc, and a_i and b_i are the filter coefficients.

Since the denominator is a product of quadratic terms, the transfer function represents a series of cascaded second-order low-pass stages, with a_i and b_i being positive real coefficients. These coefficients define the complex pole locations for each second-order filter stage, thus determining the behavior of its transfer function.

The following three types of predetermined filter coefficients are available listed in table format in Section 16.9:

- The Butterworth coefficients, optimizing the passband for maximum flatness
- The Tschebyscheff coefficients, sharpening the transition from passband into the stopband
- The Bessel coefficients, linearizing the phase response up to f_C

The transfer function of a passive RC filter does not allow further optimization, due to the lack of complex poles. The only possibility to produce conjugate complex poles using pas-

sive components is the application of LRC filters. However, these filters are mainly used at high frequencies. In the lower frequency range (< 10 MHz) the inductor values become very large and the filter becomes uneconomical to manufacture. In these cases active filters are used.

Active filters are RC networks that include an active device, such as an operational amplifier (op amp).

Section 16.3 shows that the products of the RC values and the corner frequency must yield the predetermined filter coefficients a_i and b_i, to generate the desired transfer function.

The following paragraphs introduce the most commonly used filter optimizations.

16.2.1 Butterworth Low-Pass FIlters

The Butterworth low-pass filter provides maximum passband flatness. Therefore, a Butterworth low-pass is often used as anti-aliasing filter in data converter applications where precise signal levels are required across the entire passband.

Figure 16–5 plots the gain response of different orders of Butterworth low-pass filters versus the normalized frequency axis, Ω ($\Omega = f / f_C$); the higher the filter order, the longer the passband flatness.

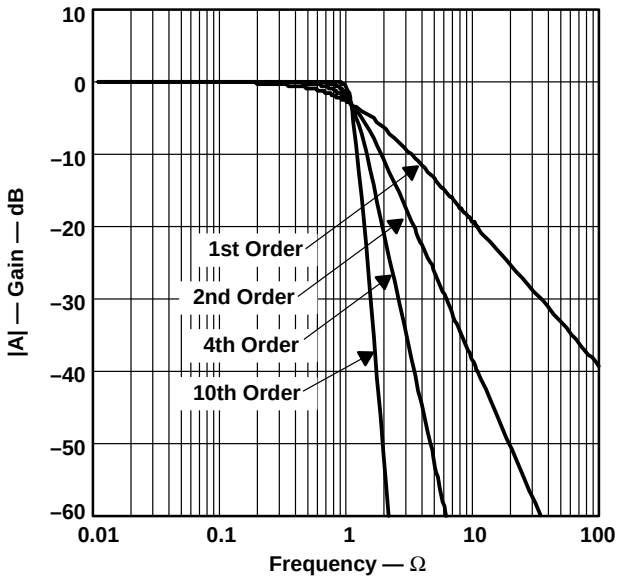


Figure 16–5. Amplitude Responses of Butterworth Low-Pass Filters

16.2.2 Tschebyscheff Low-Pass Filters

The Tschebyscheff low-pass filters provide an even higher gain rolloff above f_C. However, as Figure 16–6 shows, the passband gain is not monotone, but contains ripples of constant magnitude instead. For a given filter order, the higher the passband ripples, the higher the filter's rolloff.

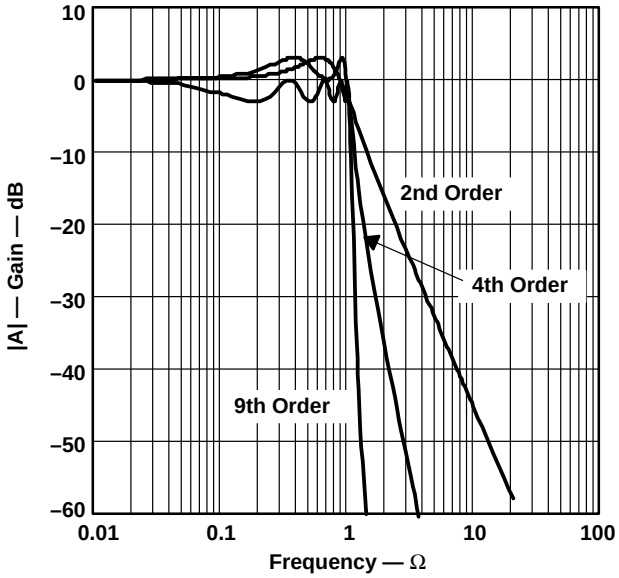


Figure 16–6. Gain Responses of Tschebyscheff Low-Pass Filters

With increasing filter order, the influence of the ripple magnitude on the filter rolloff diminishes.

Each ripple accounts for one second-order filter stage. Filters with even order numbers generate ripples above the 0-dB line, while filters with odd order numbers create ripples below 0 dB.

Tschebyscheff filters are often used in filter banks, where the frequency content of a signal is of more importance than a constant amplification.

16.2.3 Bessel Low-Pass Filters

The Bessel low-pass filters have a linear phase response (Figure 16–7) over a wide frequency range, which results in a constant group delay (Figure 16–8) in that frequency range. Bessel low-pass filters, therefore, provide an optimum square-wave transmission behavior. However, the passband gain of a Bessel low-pass filter is not as flat as that of the Butterworth low-pass, and the transition from passband to stopband is by far not as sharp as that of a Tschebyscheff low-pass filter (Figure 16–9).

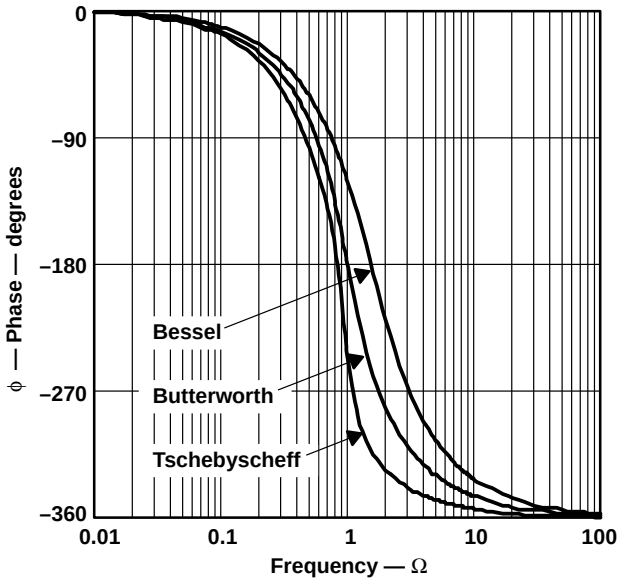


Figure 16–7. Comparison of Phase Responses of Fourth-Order Low-Pass Filters

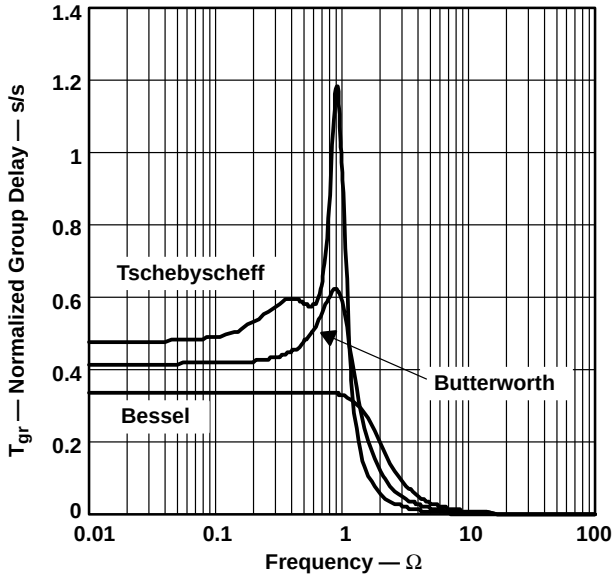


Figure 16–8. Comparison of Normalized Group Delay (Tgr) of Fourth-Order Low-Pass Filters

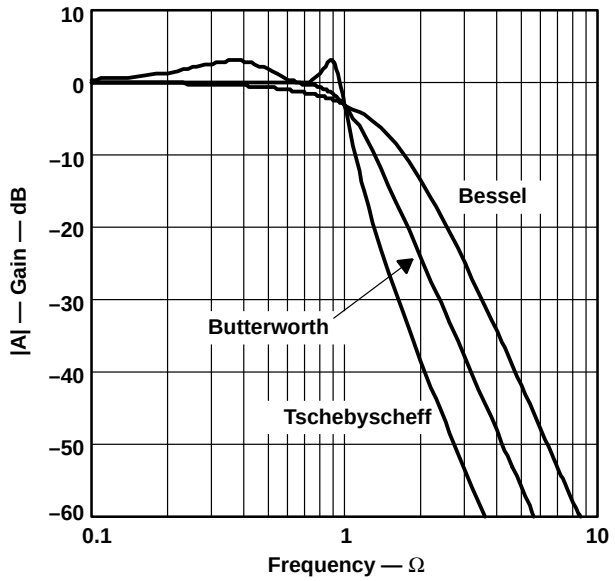


Figure 16–9. Comparison of Gain Responses of Fourth-Order Low-Pass Filters

16.2.4 Quality Factor *Q*

The quality factor *Q* is an equivalent design parameter to the filter order *n.* Instead of designing an n^{th} order Tschebyscheff low-pass, the problem can be expressed as designing a Tschebyscheff low-pass filter with a certain Q.

For band-pass filters, Q is defined as the ratio of the mid frequency, $f_{m,}$ to the bandwidth at the two –3 dB points:

$$Q = \frac{f_m}{(f_2 - f_1)}$$

For low-pass and high-pass filters, Q represents the pole quality and is defined as:

$$Q = \frac{\sqrt{b_i}}{a_i}$$

High Qs can be graphically presented as the distance between the 0-dB line and the peak point of the filter's gain response. An example is given in Figure 16–10, which shows a tenth-order Tschebyscheff low-pass filter and its five partial filters with their individual Qs.

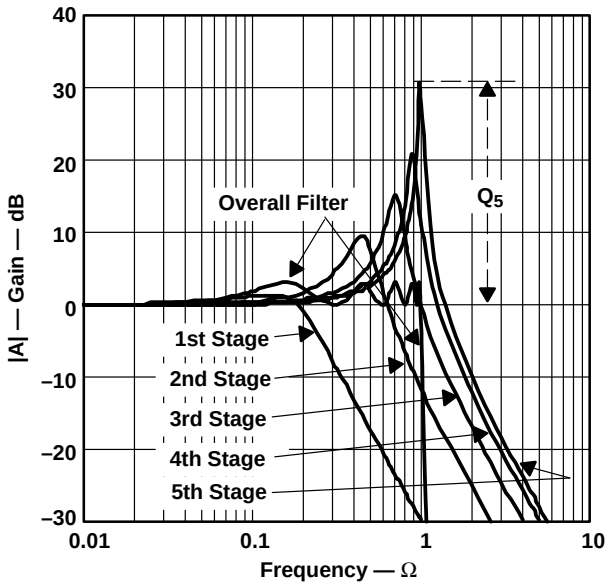


Figure 16–10. Graphical Presentation of Quality Factor Q on a Tenth-Order Tschebyscheff Low-Pass Filter with 3-dB Passband Ripple

The gain response of the fifth filter stage peaks at 31 dB, which is the logarithmic value of Q_5:

$$Q_5[\text{dB}] = 20 \cdot \log Q_5$$

Solving for the numerical value of Q5 yields:

$$Q_5 = 10^{\frac{31}{20}} = 35.48$$

which is within 1% of the theoretical value of Q = 35.85 given in Section 16.9, Table 16–9, last row.

The graphical approximation is good for Q > 3. For lower Qs, the graphical values differ from the theoretical value significantly. However, only higher Qs are of concern, since the higher the Q is, the more a filter inclines to instability.

16.2.5 Summary

The general transfer function of a low-pass filter is :

$$A(s) = \frac{A_0}{\prod_i \left(1 + a_i s + b_i s^2\right)} \qquad (16\text{–}1)$$

The filter coefficients a_i and b_i distinguish between Butterworth, Tschebyscheff, and Bessel filters. The coefficients for all three types of filters are tabulated down to the tenth order in Section 16.9, Tables 16–4 through 16–10.

The multiplication of the denominator terms with each other yields an n^{th} order polynomial of S, with n being the filter order.

While n determines the gain rolloff above f_C with $-n \cdot 20\ dB/decade$, a_i and b_i determine the gain behavior in the passband.

In addition, the ratio $\frac{\sqrt{b_i}}{a_i} = Q$ is defined as the pole quality. The higher the Q value, the more a filter inclines to instability.

16.3 Low-Pass Filter Design

Equation 16–1 represents a cascade of second-order low-pass filters. The transfer function of a single stage is:

$$A_i(s) = \frac{A_0}{\left(1 + a_i s + b_i s^2\right)} \qquad (16–2)$$

For a first-order filter, the coefficient b is always zero (b_1=0), thus yielding:

$$A(s) = \frac{A_0}{1 + a_1 s} \qquad (16–3)$$

The first-order and second-order filter stages are the building blocks for higher-order filters.

Often the filters operate at unity gain (A_0=1) to lessen the stringent demands on the op amp's open-loop gain.

Figure 16–11 shows the cascading of filter stages up to the sixth order. A filter with an even order number consists of second-order stages only, while filters with an odd order number include an additional first-order stage at the beginning.

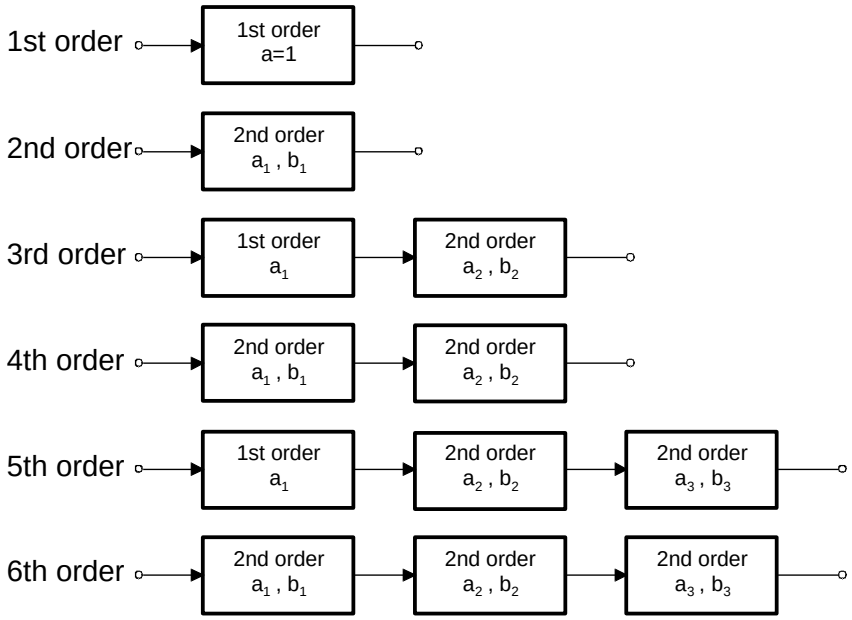


Figure 16–11. Cascading Filter Stages for Higher-Order Filters

Figure 16–10 demonstrated that the higher the corner frequency of a partial filter, the higher its Q. Therefore, to avoid the saturation of the individual stages, the filters need to be placed in the order of rising Q values. The Q values for each filter order are listed (in rising order) in Section 16–9, Tables 16–4 through 16–10.

16.3.1 First-Order Low-Pass Filter

Figures 16–12 and 16–13 show a first-order low-pass filter in the inverting and in the noninverting configuration.

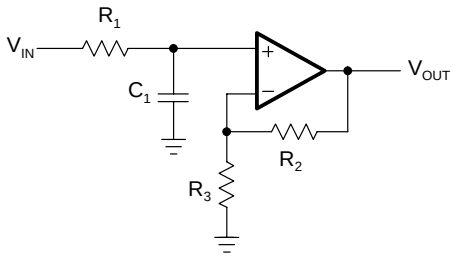


Figure 16–12. First-Order Noninverting Low-Pass Filter

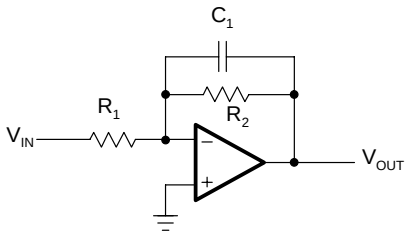


Figure 16–13. First-Order Inverting Low-Pass Filter

The transfer functions of the circuits are:

$$A(s) = \frac{1 + \frac{R_2}{R_3}}{1 + \omega_c R_1 C_1 s} \quad \text{and} \quad A(s) = \frac{-\frac{R_2}{R_1}}{1 + \omega_c R_2 C_1 s}$$

The negative sign indicates that the inverting amplifier generates a 180° phase shift from the filter input to the output.

The coefficient comparison between the two transfer functions and Equation 16–3 yields:

$$A_0 = 1 + \frac{R_2}{R_3} \quad \text{and} \quad A_0 = -\frac{R_2}{R_1}$$

$$a_1 = \omega_c R_1 C_1 \quad \text{and} \quad a_1 = \omega_c R_2 C_1$$

To dimension the circuit, specify the corner frequency (f_C), the dc gain (A_0), and capacitor C_1, and then solve for resistors R_1 and R_2:

$$R_1 = \frac{a_1}{2\pi f_c C_1} \quad \text{and} \quad R_2 = \frac{a_1}{2\pi f_c C_1}$$

$$R_2 = R_3(A_0 - 1) \quad \text{and} \quad R_1 = -\frac{R_2}{A_0}$$

The coefficient a_1 is taken from one of the coefficient tables, Tables 16–4 through 16–10 in Section 16.9.

Note, that all filter types are identical in their first order and $a_1 = 1$. For higher filter orders, however, $a_1 \neq 1$ because the corner frequency of the first-order stage is different from the corner frequency of the overall filter.

Example 16–1. First-Order Unity-Gain Low-Pass Filter

For a first-order unity-gain low-pass filter with f_C = 1 kHz and C_1 = 47 nF, R_1 calculates to:

$$R_1 = \frac{a_1}{2\pi f_c C_1} = \frac{1}{2\pi \cdot 10^3 \text{Hz} \cdot 47 \cdot 10^{-9} \text{F}} = 3.38\ \text{k}\Omega$$

However, to design the first stage of a third-order unity-gain Bessel low-pass filter, assuming the same values for f_C and C_1, requires a different value for R_1. In this case, obtain a_1 for a third-order Bessel filter from Table 16–4 in Section 16.9 (Bessel coefficients) to calculate R_1:

$$R_1 = \frac{a_1}{2\pi f_c C_1} = \frac{0.756}{2\pi \cdot 10^3 \text{Hz} \cdot 47 \cdot 10^{-9} \text{F}} = 2.56\ \text{k}\Omega$$

When operating at unity gain, the noninverting amplifier reduces to a voltage follower (Figure 16–14), thus inherently providing a superior gain accuracy. In the case of the inverting amplifier, the accuracy of the unity gain depends on the tolerance of the two resistors, R_1 and R_2.

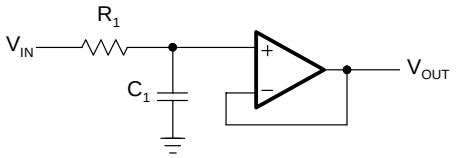


Figure 16–14. First-Order Noninverting Low-Pass Filter with Unity Gain

16.3.2 Second-Order Low-Pass Filter

There are two topologies for a second-order low-pass filter, the Sallen-Key and the Multiple Feedback (MFB) topology.

16.3.2.1 Sallen-Key Topology

The general Sallen-Key topology in Figure 16–15 allows for separate gain setting via $A_0 = 1+R_4/R_3$. However, the unity-gain topology in Figure 16–16 is usually applied in filter designs with high gain accuracy, unity gain, and low Qs (Q < 3).

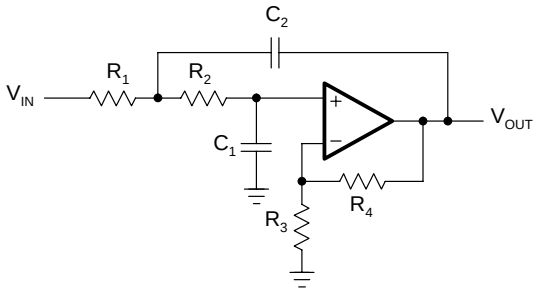


Figure 16–15. General Sallen-Key Low-Pass Filter

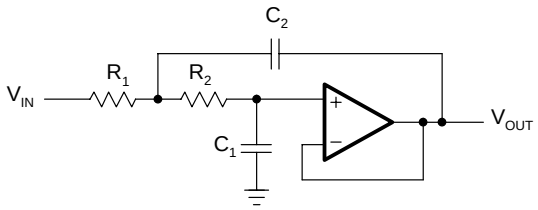


Figure 16–16. Unity-Gain Sallen-Key Low-Pass Filter

The transfer function of the circuit in Figure 16–15 is:

$$A(s) = \frac{A_0}{1 + \omega_c\left[C_1(R_1 + R_2) + (1 - A_0)\,R_1C_2\right]s + \omega_c{}^2\,R_1R_2C_1C_2s^2}$$

For the unity-gain circuit in Figure 16–16 (A_0=1), the transfer function simplifies to:

$$A(s) = \frac{1}{1 + \omega_cC_1(R_1 + R_2)\,s + \omega_c{}^2\,R_1R_2C_1C_2s^2}$$

The coefficient comparison between this transfer function and Equation 16–2 yields:

$$A_0 = 1$$
$$a_1 = \omega_cC_1(R_1 + R_2)$$
$$b_1 = \omega_c{}^2R_1R_2C_1C_2$$

Given C_1 and C_2, the resistor values for R_1 and R_2 are calculated through:

$$R_{1,2} = \frac{a_1C_2 \mp \sqrt{a_1{}^2C_2{}^2 - 4b_1C_1C_2}}{4\pi f_cC_1C_2}$$

In order to obtain real values under the square root, C_2 must satisfy the following condition:

$$C_2 \geq C_1 \frac{4b_1}{a_1^2}$$

Example 16–2. Second-Order Unity-Gain Tschebyscheff Low-Pass Filter

The task is to design a second-order unity-gain Tschebyscheff low-pass filter with a corner frequency of f_C = 3 kHz and a 3-dB passband ripple.

From Table 16–9 (the Tschebyscheff coefficients for 3-dB ripple), obtain the coefficients a_1 and b_1 for a second-order filter with a_1 = 1.0650 and b_1 = 1.9305.

Specifying C_1 as 22 nF yields in a C_2 of:

$$C_2 \geq C_1 \frac{4b_1}{a_1^2} = 22 \cdot 10^{-9}\text{nF} \cdot \frac{4 \cdot 1.9305}{1.065^2} \cong 150\ \text{nF}$$

Inserting a_1 and b_1 into the resistor equation for $R_{1,2}$ results in:

$$R_1 = \frac{1.065 \cdot 150 \cdot 10^{-9} - \sqrt{\left(1.065 \cdot 150 \cdot 10^{-9}\right)^2 - 4 \cdot 1.9305 \cdot 22 \cdot 10^{-9} \cdot 150 \cdot 10^{-9}}}{4\pi \cdot 3 \cdot 10^3 \cdot 22 \cdot 10^{-9} \cdot 150 \cdot 10^{-9}} = 1.26\ \text{k}\Omega$$

and

$$R_2 = \frac{1.065 \cdot 150 \cdot 10^{-9} + \sqrt{\left(1.065 \cdot 150 \cdot 10^{-9}\right)^2 - 4 \cdot 1.9305 \cdot 22 \cdot 10^{-9} \cdot 150 \cdot 10^{-9}}}{4\pi \cdot 3 \cdot 10^3 \cdot 22 \cdot 10^{-9} \cdot 150 \cdot 10^{-9}} = 1.30\ \text{k}\Omega$$

with the final circuit shown in Figure 16–17.

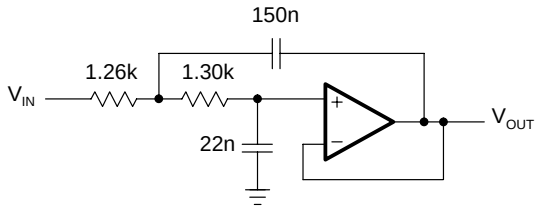


Figure 16–17. Second-Order Unity-Gain Tschebyscheff Low-Pass with 3-dB Ripple

A special case of the general Sallen-Key topology is the application of equal resistor values and equal capacitor values: $R_1 = R_2 = R$ and $C_1 = C_2 = C$.

The general transfer function changes to:

$$A(s) = \frac{A_0}{1 + \omega_c RC(3 - A_0)s + (\omega_c RC)^2 s^2} \qquad \text{with} \qquad A_0 = 1 + \frac{R_4}{R_3}$$

The coefficient comparison with Equation 16–2 yields:

$$a_1 = \omega_c RC(3 - A_0)$$

$$b_1 = (\omega_c RC)^2$$

Given C and solving for R and A_0 results in:

$$R = \frac{\sqrt{b_1}}{2\pi f_c C} \qquad \text{and} \qquad A_0 = 3 - \frac{a_1}{\sqrt{b_1}} = 3 - \frac{1}{Q}$$

Thus, A_0 depends solely on the pole quality Q and vice versa; Q, and with it the filter type, is determined by the gain setting of A_0:

$$Q = \frac{1}{3 - A_0}$$

The circuit in Figure 16–18 allows the filter type to be changed through the various resistor ratios R_4/R_3.

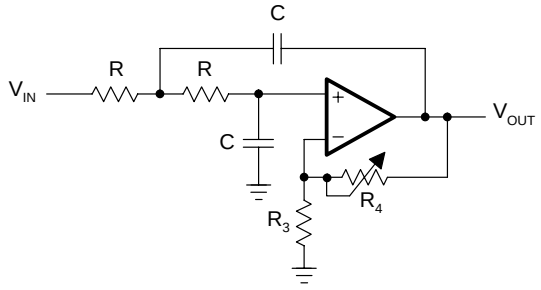


Figure 16–18. Adjustable Second-Order Low-Pass Filter

Table 16–1 lists the coefficients of a second-order filter for each filter type and gives the resistor ratios that adjust the Q.

Table 16–1. Second-Order FIlter Coefficients

SECOND-ORDER	BESSEL	BUTTERWORTH	3-dB TSCHEBYSCHEFF
a_1	1.3617	1.4142	1.065
b_1	0.618	1	1.9305
Q	0.58	0.71	1.3
R_4/R_3	0.268	0.568	0.234

16.3.2.2 Multiple Feedback Topology

The MFB topology is commonly used in filters that have high Qs and require a high gain.

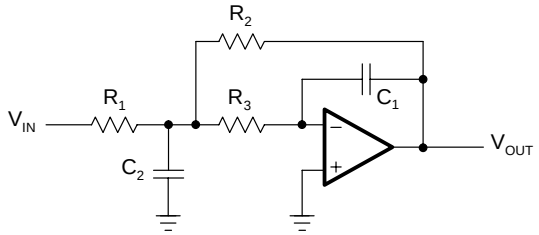


Figure 16–19. Second-Order MFB Low-Pass Filter

The transfer function of the circuit in Figure 16–19 is:

$$A(s) = -\frac{\frac{R_2}{R_1}}{1 + \omega_c C_1\left(R_2 + R_3 + \frac{R_2 R_3}{R_1}\right)s + \omega_c{}^2\, C_1 C_2 R_2 R_3 s^2}$$

Through coefficient comparison with Equation 16–2 one obtains the relation:

$$A_0 = -\frac{R_2}{R_1}$$

$$a_1 = \omega_c C_1\left(R_2 + R_3 + \frac{R_2 R_3}{R_1}\right)$$

$$b_1 = \omega_c{}^2\, C_1 C_2 R_2 R_3$$

Given C_1 and C_2, and solving for the resistors R_1–R_3:

$$R_2 = \frac{a_1 C_2 - \sqrt{a_1{}^2\, C_2{}^2 - 4 b_1 C_1 C_2 (1 - A_0)}}{4\pi f_c C_1 C_2}$$

$$R_1 = \frac{R_2}{-A_0}$$

$$R_3 = \frac{b_1}{4\pi^2 f_c{}^2\, C_1 C_2 R_2}$$

In order to obtain real values for R_2, C_2 must satisfy the following condition:

$$C_2 \geq C_1 \frac{4b_1\left(1 - A_0\right)}{{a_1}^2}$$

16.3.3 Higher-Order Low-Pass Filters

Higher-order low-pass filters are required to sharpen a desired filter characteristic. For that purpose, first-order and second-order filter stages are connected in series, so that the product of the individual frequency responses results in the optimized frequency response of the overall filter.

In order to simplify the design of the partial filters, the coefficients a_i and b_i for each filter type are listed in the coefficient tables (Tables 16–4 through 16–10 in Section 16.9), with each table providing sets of coefficients for the first 10 filter orders.

Example 16–3. Fifth-Order Filter

The task is to design a fifth-order unity-gain Butterworth low-pass filter with the corner frequency f_C = 50 kHz.

First the coefficients for a fifth-order Butterworth filter are obtained from Table 16–5, Section 16.9:

	a_i	b_i
Filter 1	$a_1 = 1$	$b_1 = 0$
Filter 2	$a_2 = 1.6180$	$b_2 = 1$
Filter 3	$a_3 = 0.6180$	$b_3 = 1$

Then dimension each partial filter by specifying the capacitor values and calculating the required resistor values.

First Filter

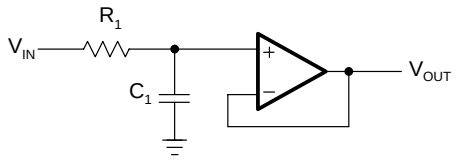


Figure 16–20. First-Order Unity-Gain Low-Pass

With C_1 = 1nF,

$$R_1 = \frac{a_1}{2\pi f_c C_1} = \frac{1}{2\pi \cdot 50 \cdot 10^3 \text{Hz} \cdot 1 \cdot 10^{-9} \text{F}} = 3.18\ \text{k}\Omega$$

The closest 1% value is 3.16 kΩ.

Second Filter

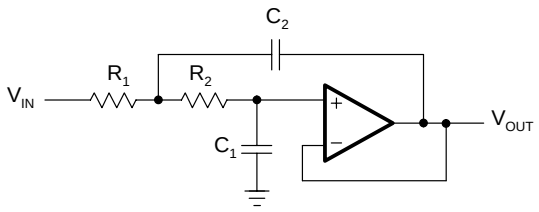


Figure 16–21. Second-Order Unity-Gain Sallen-Key Low-Pass Filter

With C_1 = 820 pF,

$$C_2 \geq C_1 \frac{4b_2}{a_2^2} = 820 \cdot 10^{-12} \text{F} \cdot \frac{4 \cdot 1}{1.618^2} = 1.26 \text{ nF}$$

The closest 5% value is 1.5 nF.

With C_1 = 820 pF and C_2 = 1.5 nF, calculate the values for R1 and R2 through:

$$R_1 = \frac{a_2 C_2 - \sqrt{a_2^2 C_2^2 - 4b_2 C_1 C_2}}{4\pi f_c C_1 C_2} \quad \text{and} \quad R_1 = \frac{a_2 C_2 + \sqrt{a_2^2 C_2^2 - 4b_2 C_1 C_2}}{4\pi f_c C_1 C_2}$$

and obtain

$$R_1 = \frac{1.618 \cdot 1.5 \cdot 10^{-9} - \sqrt{\left(1.618 \cdot 1.5 \cdot 10^{-9}\right)^2 - 4 \cdot 1 \cdot 820 \cdot 10^{-12} \cdot 1.5 \cdot 10^{-9}}}{4\pi \cdot 50 \cdot 10^3 \cdot 820 \cdot 10^{-12} \cdot 1.5 \cdot 10^{-9}} = 1.87 \text{ k}\Omega$$

$$R_2 = \frac{1.618 \cdot 1.5 \cdot 10^{-9} + \sqrt{\left(1.618 \cdot 1.5 \cdot 10^{-9}\right)^2 - 4 \cdot 1 \cdot 820 \cdot 10^{-12} \cdot 1.5 \cdot 10^{-9}}}{4\pi \cdot 50 \cdot 10^3 \cdot 820 \cdot 10^{-12} \cdot 1.5 \cdot 10^{-9}} = 4.42 \text{ k}\Omega$$

R_1 and R_2 are available 1% resistors.

Third Filter

The calculation of the third filter is identical to the calculation of the second filter, except that a_2 and b_2 are replaced by a_3 and b_3, thus resulting in different capacitor and resistor values.

Specify C_1 as 330 pF, and obtain C_2 with:

$$C_2 \geq C_1 \frac{4b_3}{a_3^2} = 330 \cdot 10^{-12} \text{F} \cdot \frac{4 \cdot 1}{0.618^2} = 3.46 \text{ nF}$$

The closest 10% value is 4.7 nF.

With C_1 = 330 pF and C_2 = 4.7 nF, the values for R1 and R2 are:

- R_1 = 1.45 kΩ, with the closest 1% value being 1.47 kΩ
- R_2 = 4.51 kΩ, with the closest 1% value being 4.53 kΩ

Figure 16–22 shows the final filter circuit with its partial filter stages.

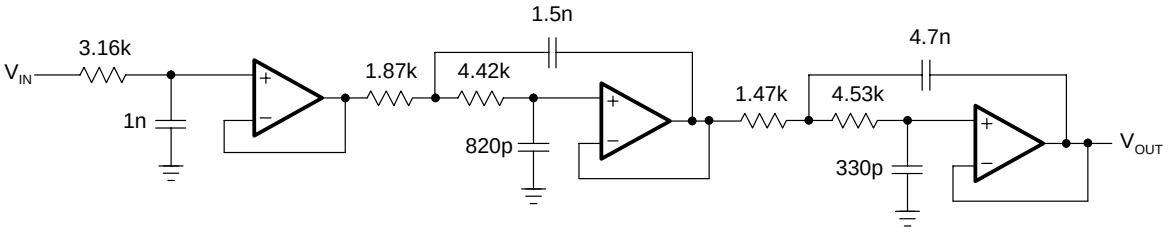


Figure 16–22. Fifth-Order Unity-Gain Butterworth Low-Pass Filter

16.4 High-Pass Filter Design

By replacing the resistors of a low-pass filter with capacitors, and its capacitors with resistors, a high-pass filter is created.

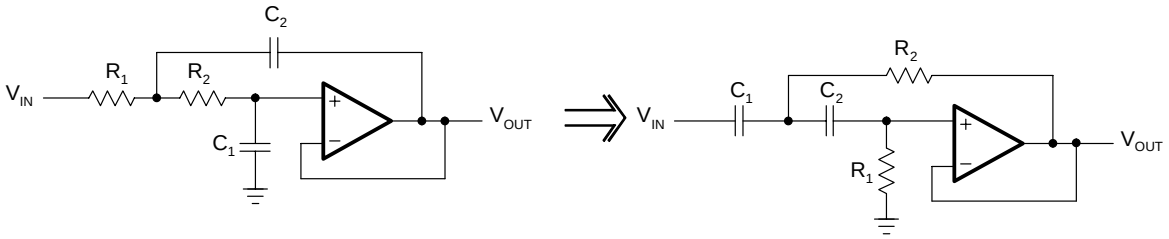


Figure 16–23. Low-Pass to High-Pass Transition Through Components Exchange

To plot the gain response of a high-pass filter, mirror the gain response of a low-pass filter at the corner frequency, Ω=1, thus replacing Ω with 1/Ω and S with 1/S in Equation 16–1.

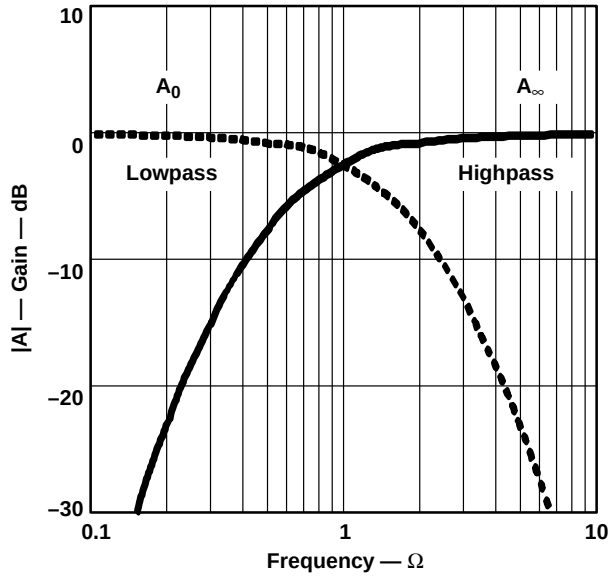


Figure 16–24. Developing The Gain Response of a High-Pass Filter

The general transfer function of a high-pass filter is then:

$$A(s) = \frac{A_\infty}{\prod_i \left(1 + \frac{a_i}{s} + \frac{b_i}{s^2}\right)} \qquad (16–4)$$

with A_∞ being the passband gain.

Since Equation 16–4 represents a cascade of second-order high-pass filters, the transfer function of a single stage is:

$$A_i(s) = \frac{A_\infty}{\left(1 + \frac{a_i}{s} + \frac{b_i}{s^2}\right)} \qquad (16–5)$$

With b=0 for all first-order filters, the transfer function of a first-order filter simplifies to:

$$A(s) = \frac{A_0}{1 + \frac{a_i}{s}} \qquad (16–6)$$

16.4.1 First-Order High-Pass Filter

Figure 16–25 and 16–26 show a first-order high-pass filter in the noninverting and the inverting configuration.

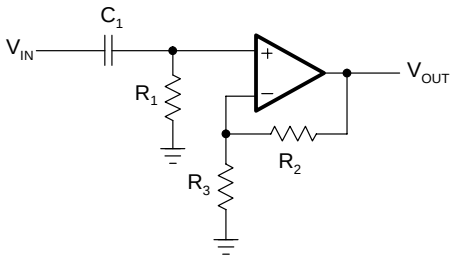


Figure 16–25. First-Order Noninverting High-Pass Filter

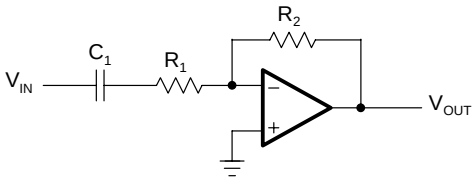


Figure 16–26. First-Order Inverting High-Pass Filter

The transfer functions of the circuits are:

$$A(s) = \frac{1 + \frac{R_2}{R_3}}{1 + \frac{1}{\omega_c R_1 C_1} \cdot \frac{1}{s}} \qquad \text{and} \qquad A(s) = -\frac{\frac{R_2}{R_1}}{1 + \frac{1}{\omega_c R_1 C_1} \cdot \frac{1}{s}}$$

The negative sign indicates that the inverting amplifier generates a 180° phase shift from the filter input to the output.

The coefficient comparison between the two transfer functions and Equation 16–6 provides two different passband gain factors:

$$A_\infty = 1 + \frac{R_2}{R_3} \qquad \text{and} \qquad A_\infty = -\frac{R_2}{R_1}$$

while the term for the coefficient a_1 is the same for both circuits:

$$a_1 = \frac{1}{\omega_c R_1 C_1}$$

To dimension the circuit, specify the corner frequency (f_C), the dc gain (A_∞), and capacitor (C_1), and then solve for R_1 and R_2:

$$R_1 = \frac{1}{2\pi f_c a_1 C_1}$$

$$R_2 = R_3(A_\infty - 1) \qquad \text{and} \qquad R_2 = -R_1 A_\infty$$

16.4.2 Second-Order High-Pass Filter

High-pass filters use the same two topologies as the low-pass filters: Sallen-Key and Multiple Feedback. The only difference is that the positions of the resistors and the capacitors have changed.

16.4.2.1 Sallen-Key Topology

The general Sallen-Key topology in Figure 16–27 allows for separate gain setting via $A_0 = 1+R_4/R_3$.

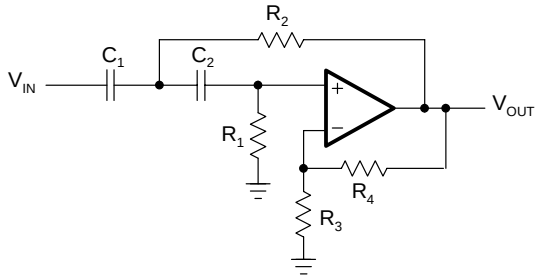


Figure 16–27. General Sallen-Key High-Pass Filter

The transfer function of the circuit in Figure 16–27 is:

$$A(s) = \frac{\alpha}{1 + \frac{R_2(C_1+C_2)+R_1C_2(1-\alpha)}{\omega_c R_1 R_2 C_1 C_2}\cdot\frac{1}{s} + \frac{1}{\omega_c^2 R_1 R_2 C_1 C_2}\cdot\frac{1}{s^2}} \qquad \text{with} \qquad \alpha = 1 + \frac{R_4}{R_3}$$

The unity-gain topology in Figure 16–28 is usually applied in low-Q filters with high gain accuracy.

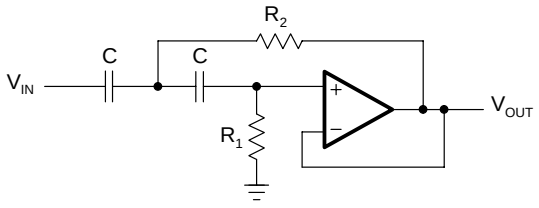


Figure 16–28. Unity-Gain Sallen-Key High-Pass Filter

To simplify the circuit design, it is common to choose unity-gain ($\alpha = 1$), and $C_1 = C_2 = C$. The transfer function of the circuit in Figure 16–28 then simplifies to:

$$A(s) = \frac{1}{1 + \frac{2}{\omega_c R_1 C} \cdot \frac{1}{s} + \frac{1}{\omega_c^2 R_1 R_2 C^2} \cdot \frac{1}{s^2}}$$

The coefficient comparison between this transfer function and Equation 16–5 yields:

$$A_\infty = 1$$

$$a_1 = \frac{2}{\omega_c R_1 C}$$

$$b_1 = \frac{1}{\omega_c^2 R_1 R_2 C^2}$$

Given C, the resistor values for R_1 and R_2 are calculated through:

$$R_1 = \frac{1}{\pi f_c C a_1}$$

$$R_2 = \frac{a_1}{4\pi f_c C b_1}$$

16.4.2.2 Multiple Feedback Topology

The MFB topology is commonly used in filters that have high Qs and require a high gain.

To simplify the computation of the circuit, capacitors C_1 and C_3 assume the same value ($C_1 = C_3 = C$) as shown in Figure 16–29.

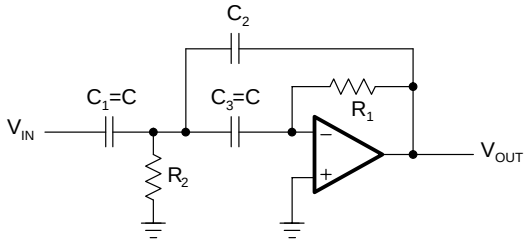


Figure 16–29. Second-Order MFB High-Pass Filter

The transfer function of the circuit in Figure 16–29 is:

$$A(s) = \frac{-\frac{C}{C_2}}{1 + \frac{2C_2 + C}{\omega_c R_1 C_2 C} \cdot \frac{1}{s} + \frac{1}{\omega_c^2 R_2 R_1 C_2 C} \cdot \frac{1}{s^2}}$$

Through coefficient comparison with Equation 16–5, obtain the following relations:

$$A_\infty = \frac{C}{C_2}$$

$$a_1 = \frac{2C + C_2}{\omega_c R_1 C C_2}$$

$$b_1 = \frac{2C + C_2}{\omega_c R_1 C C_2}$$

Given capacitors C and C_2, and solving for resistors R_1 and R_2:

$$R_1 = \frac{1 - 2A_\infty}{2\pi f_c \cdot C \cdot a_1}$$

$$R_2 = \frac{a_1}{2\pi f_c \cdot b_1 C_2 (1 - 2A_\infty)}$$

The passband gain (A_∞) of a MFB high-pass filter can vary significantly due to the wide tolerances of the two capacitors C and C_2. To keep the gain variation at a minimum, it is necessary to use capacitors with tight tolerance values.

16.4.3 Higher-Order High-Pass Filter

Likewise, as with the low-pass filters, higher-order high-pass filters are designed by cascading first-order and second-order filter stages. The filter coefficients are the same ones used for the low-pass filter design, and are listed in the coefficient tables (Tables 16–4 through 16–10 in Section 16.9).

Example 16–4. Third-Order High-Pass Filter with f_C = 1 kHz

The task is to design a third-order unity-gain Bessel high-pass filter with the corner frequency f_C = 1 kHz. Obtain the coefficients for a third-order Bessel filter from Table 16–4, Section 16.9:

	a_i	b_i
Filter 1	$a_1 = 0.756$	$b_1 = 0$
Filter 2	$a_2 = 0.9996$	$b_2 = 0.4772$

and compute each partial filter by specifying the capacitor values and calculating the required resistor values.

First Filter

With C_1 = 100 nF,

$$R_1 = \frac{1}{2\pi f_c a_1 C_1} = \frac{1}{2\pi \cdot 10^3 Hz \cdot 0.756 \cdot 100 \cdot 10^{-9} F} = 2.105\ k\Omega$$

Closest 1% value is 2.1 kΩ.

Second Filter

With C = 100nF,

$$R_1 = \frac{1}{\pi f_c C a_1} = \frac{1}{\pi \cdot 10^3 \cdot 100 \cdot 10^{-9} \cdot 0.756} = 3.18\ k\Omega$$

Closest 1% value is 3.16 kΩ.

$$R_2 = \frac{a_1}{4\pi f_c C b_1} = \frac{0.9996}{4\pi \cdot 10^3 \cdot 100 \cdot 10^{-9} \cdot 0.4772} = 1.67\ k\Omega$$

Closest 1% value is 1.65 kΩ.

Figure 16–30 shows the final filter circuit.

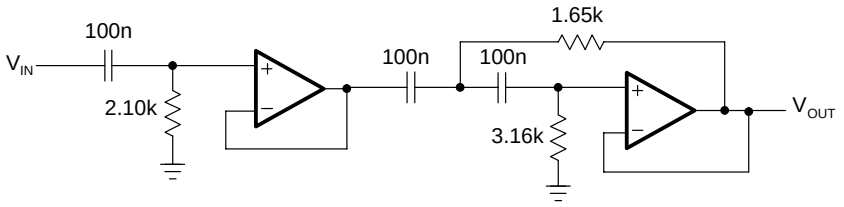


Figure 16–30. Third-Order Unity-Gain Bessel High-Pass

16.5 Band-Pass Filter Design

In Section 16.4, a high-pass response was generated by replacing the term *S* in the low-pass transfer function with the transformation *1/S*. Likewise, a band-pass characteristic is generated by replacing the S term with the transformation:

$$\frac{1}{\Delta\Omega}\left(s + \frac{1}{s}\right) \qquad (16–7)$$

In this case, the passband characteristic of a low-pass filter is transformed into the upper passband half of a band-pass filter. The upper passband is then mirrored at the mid frequency, f_m (Ω=1), into the lower passband half.

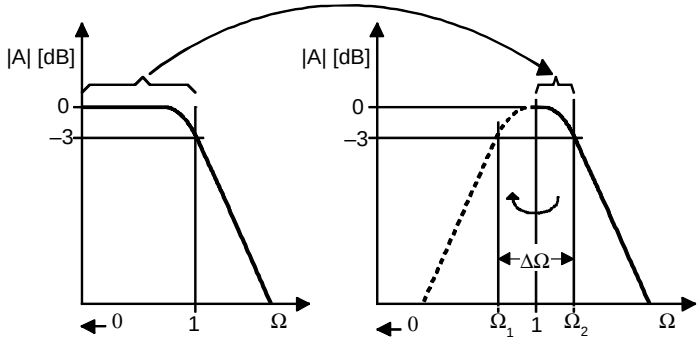


Figure 16–31. Low-Pass to Band-Pass Transition

The corner frequency of the low-pass filter transforms to the lower and upper –3 dB frequencies of the band-pass, Ω_1 and Ω_2. The difference between both frequencies is defined as the normalized bandwidth $\Delta\Omega$:

$$\Delta\Omega = \Omega_2 - \Omega_1$$

The normalized mid frequency, where Q = 1, is:

$$\Omega_m = 1 = \Omega_2 \cdot \Omega_1$$

In analogy to the resonant circuits, the quality factor Q is defined as the ratio of the mid frequency (f_m) to the bandwidth (B):

$$Q = \frac{f_m}{B} = \frac{f_m}{f_2 - f_1} = \frac{1}{\Omega_2 - \Omega_1} = \frac{1}{\Delta\Omega} \qquad (16\text{–}8)$$

The simplest design of a band-pass filter is the connection of a high-pass filter and a low-pass filter in series, which is commonly done in wide-band filter applications. Thus, a first-order high-pass and a first-order low-pass provide a second-order band-pass, while a second-order high-pass and a second-order low-pass result in a fourth-order band-pass response.

In comparison to wide-band filters, narrow-band filters of higher order consist of cascaded second-order band-pass filters that use the Sallen-Key or the Multiple Feedback (MFB) topology.

16.5.1 Second-Order Band-Pass Filter

To develop the frequency response of a second-order band-pass filter, apply the transformation in Equation 16–7 to a first-order low-pass transfer function:

$$A(s) = \frac{A_0}{1 + s}$$

Replacing s with $\frac{1}{\Delta\Omega}\left(s + \frac{1}{s}\right)$

yields the general transfer function for a second-order band-pass filter:

$$A(s) = \frac{A_0 \cdot \Delta\Omega \cdot s}{1 + \Delta\Omega \cdot s + s^2} \qquad (16–9)$$

When designing band-pass filters, the parameters of interest are the gain at the mid frequency (A_m) and the quality factor (Q), which represents the selectivity of a band-pass filter.

Therefore, replace A_0 with A_m and $\Delta\Omega$ with 1/Q (Equation 16–7) and obtain:

$$A(s) = \frac{\frac{A_m}{Q} \cdot s}{1 + \frac{1}{Q} \cdot s + s^2} \qquad (16–10)$$

Figure 16–32 shows the normalized gain response of a second-order band-pass filter for different Qs.

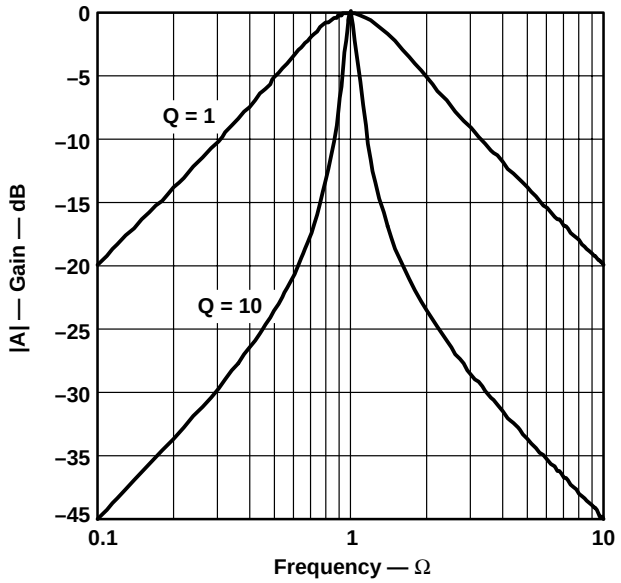


Figure 16–32. Gain Response of a Second-Order Band-Pass Filter

The graph shows that the frequency response of second-order band-pass filters gets steeper with rising Q, thus making the filter more selective.

16.5.1.1 Sallen-Key Topology

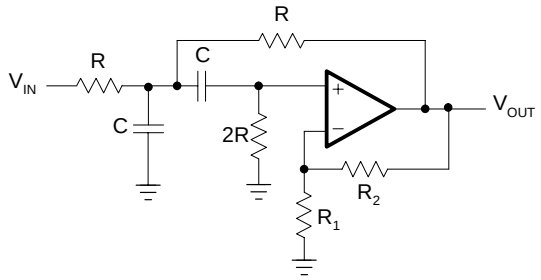


Figure 16–33. Sallen-Key Band-Pass

The Sallen-Key band-pass circuit in Figure 16–33 has the following transfer function:

$$A(s) = \frac{G \cdot RC\omega_m \cdot s}{1 + RC\omega_m(3 - G) \cdot s + R^2C^2\omega_m{}^2 \cdot s^2}$$

Through coefficient comparison with Equation 16–10, obtain the following equations:

$$\text{mid-frequency: } f_m = \frac{1}{2\pi RC}$$

$$\text{inner gain: } G = 1 + \frac{R_2}{R_1}$$

$$\text{gain at } f_{m:} \quad A_m = \frac{G}{3 - G}$$

$$\text{filter quality: } Q = \frac{1}{3 - G}$$

The Sallen-Key circuit has the advantage that the quality factor (Q) can be varied via the inner gain (G) without modifying the mid frequency (fm). A drawback is, however, that Q and A_m cannot be adjusted independently.

Care must be taken when G approaches the value of 3, because then A_m becomes infinite and causes the circuit to oscillate.

To set the mid frequency of the band-pass, specify f_m and C and then solve for R:

$$R = \frac{1}{2\pi f_m C}$$

Because of the dependency between Q and A_m, there are two options to solve for R_2: either to set the gain at mid frequency:

$$R_2 = \frac{2A_m - 1}{1 + A_m}$$

or to design for a specified Q:

$$R_2 = \frac{2Q - 1}{Q}$$

16.5.1.2 Multiple Feedback Topology

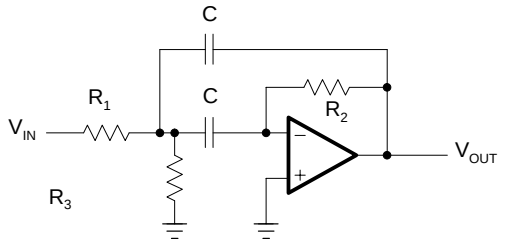


Figure 16–34. MFB Band-Pass

The MFB band-pass circuit in Figure 16–34 has the following transfer function:

$$A(s) = \frac{-\frac{R_2R_3}{R_1+R_3}C\omega_m \cdot s}{1 + \frac{2R_1R_3}{R_1+R_3}C\omega_m \cdot s + \frac{R_1R_2R_3}{R_1+R_3}C^2 \cdot \omega_m{}^2 \cdot s^2}$$

The coefficient comparison with Equation 16–9, yields the following equations:

$$\text{mid-frequency: } f_m = \frac{1}{2\pi C}\sqrt{\frac{R_1 + R_3}{R_1R_2R_3}}$$

$$\text{gain at } f_m: \quad -A_m = \frac{R_2}{2R_1}$$

$$\text{filter quality: } \quad Q = \pi f_m R_2 C$$

$$\text{bandwidth: } \quad B = \frac{1}{\pi R_2 C}$$

The MFB band-pass allows to adjust Q, A_m, and f_m independently. Bandwidth and gain factor do not depend on R_3. Therefore, R_3 can be used to modify the mid frequency with-

out affecting bandwidth, B, or gain, A_m. For low values of Q, the filter can work without R3, however, Q then depends on A_m via:

$$-A_m = 2Q^2$$

Example 16–5. Second-Order MFB Band-Pass Filter with f_m = 1 kHz

To design a second-order MFB band-pass filter with a mid frequency of f_m = 1 kHz, a quality factor of Q = 10, and a gain of A_m = –2, assume a capacitor value of C = 100 nF, and solve the previous equations for R_1 through R_3 in the following sequence:

$$R_2 = \frac{Q}{\pi f_m C} = \frac{10}{\pi \cdot 1\ \text{kHz} \cdot 100\ \text{nF}} = 31.8\ \text{k}\Omega$$

$$R_1 = \frac{R_2}{-2A_m} = \frac{31.8\ \text{k}\Omega}{4} = 7.96\ \text{k}\Omega$$

$$R_3 = \frac{-A_m R_1}{2Q^2 + A_m} = \frac{2 \cdot 7.96\ \text{k}\Omega}{200 - 2} = 80.4\ \Omega$$

16.5.2 Fourth-Order Band-Pass Filter (Staggered Tuning)

Figure 16–32 shows that the frequency response of second-order band-pass filters gets steeper with rising Q. However, there are band-pass applications that require a flat gain response close to the mid frequency as well as a sharp passband-to-stopband transition. These tasks can be accomplished by higher-order band-pass filters.

Of particular interest is the application of the low-pass to band-pass transformation onto a second-order low-pass filter, since it leads to a fourth-order band-pass filter.

Replacing the *S* term in Equation 16–2 with Equation 16–7 gives the general transfer function of a fourth-order band-pass:

$$A(s) = \frac{\frac{s^2 \cdot A_0 (\Delta\Omega)^2}{b_1}}{1 + \frac{a_1}{b_1}\Delta\Omega \cdot s + \left[2 + \frac{(\Delta\Omega)^2}{b_1}\right] \cdot s^2 + \frac{a_1}{b_1}\Delta\Omega \cdot s^3 + s^4} \qquad (16\text{–}11)$$

Similar to the low-pass filters, the fourth-order transfer function is split into two second-order band-pass terms. Further mathematical modifications yield:

$$A(s) = \frac{\frac{A_{mi}}{Q_i} \cdot \alpha s}{\left[1 + \frac{\alpha s}{Q_1} + (\alpha s)^2\right]} \cdot \frac{\frac{A_{mi}}{Q_i} \cdot \frac{s}{\alpha}}{\left[1 + \frac{1}{Q_i}\left(\frac{s}{\alpha}\right) + \left(\frac{s}{\alpha}\right)^2\right]} \qquad (16\text{–}12)$$

Equation 16–12 represents the connection of two second-order band-pass filters in series, where

- A_{mi} is the gain at the mid frequency, f_{mi}, of each partial filter
- Q_i is the pole quality of each filter
- α and $1/\alpha$ are the factors by which the mid frequencies of the individual filters, f_{m1} and f_{m2}, derive from the mid frequency, f_m, of the overall bandpass.

In a fourth-order band-pass filter with high Q, the mid frequencies of the two partial filters differ only slightly from the overall mid frequency. This method is called staggered tuning.

Factor α needs to be determined through successive approximation, using equation 16–13:

$$\alpha^2 + \left[\frac{\alpha \cdot \Delta\Omega \cdot a_1}{b_1(1 + \alpha^2)}\right]^2 + \frac{1}{\alpha^2} - 2 - \frac{(\Delta\Omega)^2}{b^1} = 0 \qquad (16\text{–}13)$$

with a_1 and b_1 being the second-order low-pass coefficients of the desired filter type.

To simplify the filter design, Table 16–2 lists those coefficients, and provides the α values for three different quality factors, Q = 1, Q = 10, and Q = 100.

Table 16–2. Values of α For Different Filter Types and Different Qs

Bessel				**Butterworth**				**Tschebyscheff**			
a_1	1.3617			**a_1**	1.4142			**a_1**	1.0650		
b_1	0.6180			**b_1**	1.0000			**b_1**	1.9305		
Q	100	10	1	**Q**	100	10	1	**Q**	100	10	1
$\Delta\Omega$	**0.01**	**0.1**	**1**	$\Delta\Omega$	**0.01**	**0.1**	**1**	$\Delta\Omega$	**0.01**	**0.1**	**1**
α	1.0032	1.0324	1.438	α	1.0035	1.036	1.4426	α	1.0033	1.0338	1.39

After α has been determined, all quantities of the partial filters can be calculated using the following equations:

The mid frequency of filter 1 is:

$$f_{m1} = \frac{f_m}{\alpha} \tag{16–14}$$

the mid frequency of filter 2 is:

$$f_{m2} = f_m \cdot \alpha \tag{16–15}$$

with f_m being the mid frequency of the overall forth-order band-pass filter.

The individual pole quality, Q_i, is the same for both filters:

$$Q_i = Q \cdot \frac{(1 + \alpha^2)b_1}{\alpha \cdot a_1} \tag{16–16}$$

with Q being the quality factor of the overall filter.

The individual gain (A_{mi}) at the partial mid frequencies, f_{m1} and f_{m2}, is the same for both filters:

$$A_{mi} = \frac{Q_i}{Q} \cdot \sqrt{\frac{A_m}{B_1}} \tag{16–17}$$

with A_m being the gain at mid frequency, f_m, of the overall filter.

Example 16–6. Fourth-Order Butterworth Band-Pass Filter

The task is to design a fourth-order Butterworth band-pass with the following parameters:

- mid frequency, f_m = 10 kHz
- bandwidth, B = 1000 Hz
- and gain, A_m = 1

From Table 16–2 the following values are obtained:

- a1 = 1.4142
- b1 = 1
- α = 1.036

In accordance with Equations 16–14 and 16–15, the mid frequencies for the partial filters are:

$$f_{mi} = \frac{10\ \text{kHz}}{1.036} = 9.653\ \text{kHz} \qquad \text{and} \qquad f_{m2} = 10\ \text{kHz} \cdot 1.036 = 10.36\ \text{kHz}$$

The overall Q is defined as $Q = f_m/B$, and for this example results in Q = 10.

Using Equation 16–16, the Q_i of both filters is:

$$Q_i = 10 \cdot \frac{(1 + 1.036^2) \cdot 1}{1.036 \cdot 1.4142} = 14.15$$

With Equation 16–17, the passband gain of the partial filters at f_{m1} and f_{m2} calculates to:

$$A_{mi} = \frac{14.15}{10} \cdot \sqrt{\frac{1}{1}} = 1.415$$

The Equations 16–16 and 16–17 show that Q_i and A_{mi} of the partial filters need to be independently adjusted. The only circuit that accomplishes this task is the MFB band-pass filter in Paragraph 16.5.1.2.

To design the individual second-order band-pass filters, specify C = 10 nF, and insert the previously determined quantities for the partial filters into the resistor equations of the MFB band-pass filter. The resistor values for both partial filters are calculated below.

Filter 1:

$$R_{21} = \frac{Q_i}{\pi f_{m1} C} = \frac{14.15}{\pi \cdot 9.653\ \text{kHz} \cdot 10\ \text{nF}} = 46.7\ \text{k}\Omega$$

$$R_{11} = \frac{R_{21}}{-2A_{mi}} = \frac{46.7\ \text{k}\Omega}{-2 \cdot -1.415} = 16.5\ \text{k}\Omega$$

$$R_{31} = \frac{-A_{mi} R_{11}}{2Q_i^2 + A_{mi}} = \frac{1.415 \cdot 16.5\ \text{k}\Omega}{2 \cdot 14.15^2 + 1.415} = 58.1\ \Omega$$

Filter 2:

$$R_{22} = \frac{Q_i}{\pi f_{m2} C} = \frac{14.15}{\pi \cdot 10.36\ \text{kHz} \cdot 10\ \text{nF}} = 43.5\ \text{k}\Omega$$

$$R_{12} = \frac{R_{22}}{-2A_{mi}} = \frac{43.5\ \text{k}\Omega}{-2 \cdot -1.415} = 15.4\ \text{k}\Omega$$

$$R_{32} = \frac{-A_{mi} R_{12}}{2Q_i^2 + A_{mi}} = \frac{1.415 \cdot 15.4\ \text{k}\Omega}{2 \cdot 14.15^2 + 1.415} = 54.2\ \Omega$$

Figure 16–35 compares the gain response of a fourth-order Butterworth band-pass filter with Q = 1 and its partial filters to the fourth-order gain of Example 16–4 with Q = 10.

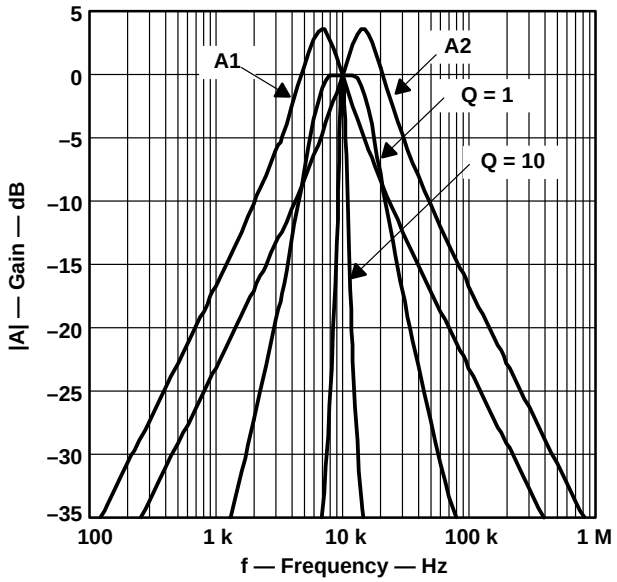


Figure 16–35. Gain Responses of a Fourth-Order Butterworth Band-Pass and its Partial Filters

16.6 Band-Rejection Filter Design

A band-rejection filter is used to suppress a certain frequency rather than a range of frequencies.

Two of the most popular band-rejection filters are the active twin-T and the active Wien-Robinson circuit, both of which are second-order filters.

To generate the transfer function of a second-order band-rejection filter, replace the *S* term of a first-order low-pass response with the transformation in 16–18:

$$\frac{\Delta\Omega}{s + \frac{1}{s}} \tag{16–18}$$

which gives:

$$A(s) = \frac{A_0(1 + s^2)}{1 + \Delta\Omega \cdot s + s^2} \tag{16–19}$$

Thus the passband characteristic of the low-pass filter is transformed into the lower passband of the band-rejection filter. The lower passband is then mirrored at the mid frequency, f_m (Ω=1), into the upper passband half (Figure 16–36).

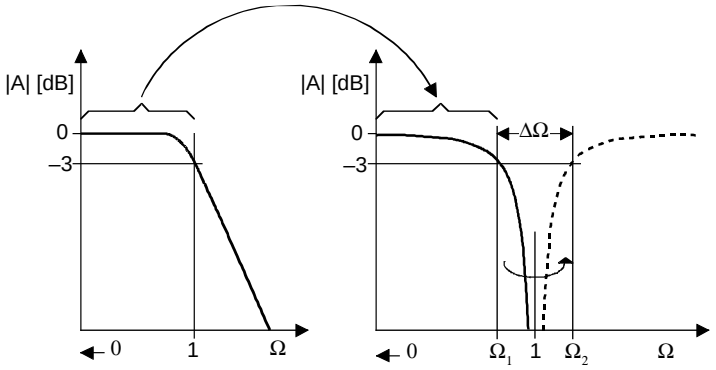


Figure 16–36. Low-Pass to Band-Rejection Transition

The corner frequency of the low-pass transforms to the lower and upper –3-dB frequencies of the band-rejection filter Ω_1 and Ω_2. The difference between both frequencies is the normalized bandwidth $\Delta\Omega$:

$$\Delta\Omega = \Omega_{max} - \Omega_{min}$$

Identical to the selectivity of a band-pass filter, the quality of the filter rejection is defined as:

$$Q = \frac{f_m}{B} = \frac{1}{\Delta\Omega}$$

Therefore, replacing $\Delta\Omega$ in Equation 16–19 with 1/Q yields:

$$A(s) = \frac{A_0(1 + s^2)}{1 + \frac{1}{Q}\cdot s + s^2} \qquad (16–20)$$

16.6.1 Active Twin-T Filter

The original twin-T filter, shown in Figure 16–37, is a passive RC-network with a quality factor of Q = 0.25. To increase Q, the passive filter is implemented into the feedback loop of an amplifier, thus turning into an active band-rejection filter, shown in Figure 16–38.

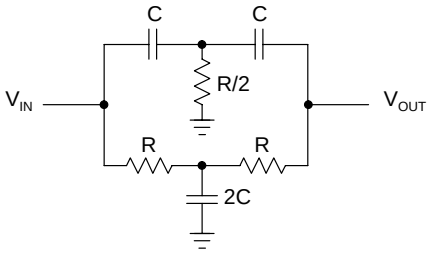


Figure 16–37. Passive Twin-T Filter

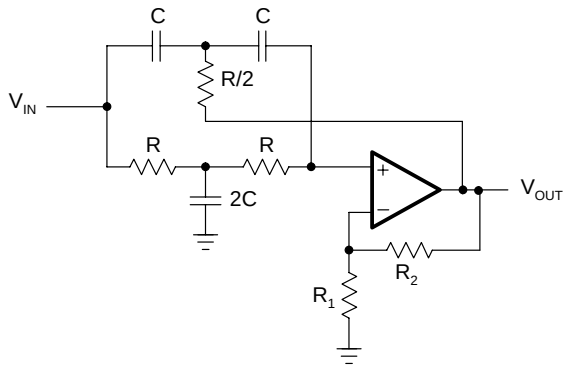


Figure 16–38. Active Twin-T Filter

The transfer function of the active twin-T filter is:

$$A(s) = \frac{k(1 + s^2)}{1 + 2(2 - k)\cdot s + s^2} \qquad (16–21)$$

Comparing the variables of Equation 16–21 with Equation 16–20 provides the equations that determine the filter parameters:

$$\text{mid-frequency:} \quad f_m = \frac{1}{2\pi RC}$$

$$\text{inner gain:} \quad G = 1 + \frac{R_2}{R_1}$$

$$\text{passband gain:} \quad A_0 = G$$

$$\text{rejection quality:} \quad Q = \frac{1}{2(2 - G)}$$

The twin-T circuit has the advantage that the quality factor (Q) can be varied via the inner gain (G) without modifying the mid frequency (f_m). However, Q and A_m cannot be adjusted independently.

To set the mid frequency of the band-pass, specify f_m and C, and then solve for R:

$$R = \frac{1}{2\pi f_m C}$$

Because of the dependency between Q and A_m, there are two options to solve for R_2: either to set the gain at mid frequency:

$$R_2 = (A_0 - 1)R_1$$

or to design for a specific Q:

$$R_2 = R_1\left(1 - \frac{1}{2Q}\right)$$

16.6.2 Active Wien-Robinson Filter

The Wien-Robinson bridge in Figure 16–39 is a passive band-rejection filter with differential output. The output voltage is the difference between the potential of a constant voltage divider and the output of a band-pass filter. Its Q-factor is close to that of the twin-T circuit. To achieve higher values of Q, the filter is connected into the feedback loop of an amplifier.

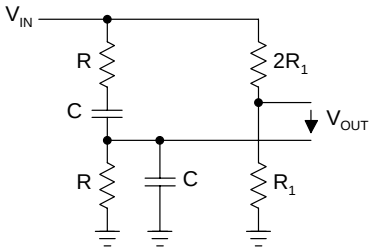


Figure 16–39. Passive Wien-Robinson Bridge

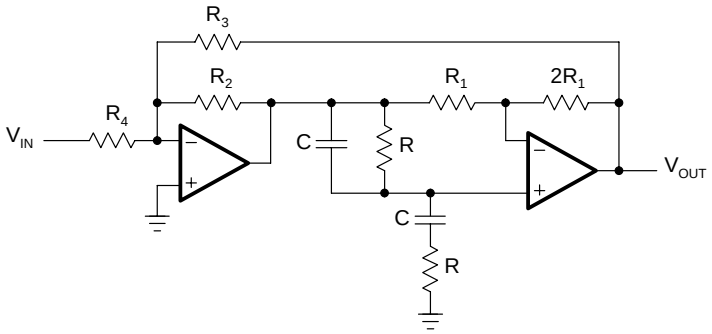


Figure 16–40. Active Wien-Robinson Filter

The active Wien-Robinson filter in Figure 16–40 has the transfer function:

$$A(s) = -\frac{\frac{\beta}{1+\alpha}\left(1 + s^2\right)}{1 + \frac{3}{1+\alpha}\cdot s + s^2} \qquad (16\text{–}22)$$

$$\text{with } \alpha = \frac{R_2}{R_3} \quad \text{and} \quad \beta = \frac{R_2}{R_4}$$

Comparing the variables of Equation 16–22 with Equation 16–20 provides the equations that determine the filter parameters:

mid-frequency: $f_m = \frac{1}{2\pi RC}$

passband gain: $A_0 = -\frac{\beta}{1+\alpha}$

rejection quality: $Q = \frac{1+\alpha}{3}$

To calculate the individual component values, establish the following design procedure:

1) Define f_m and C and calculate R with:

$$R = \frac{1}{2\pi f_m C}$$

2) Specify Q and determine α via:

$$\alpha = 3Q - 1$$

3) Specify A_0 and determine β via:

$$\beta = -A_0 \cdot 3Q$$

4) Define R_2 and calculate R_3 and R_4 with:

$$R_3 = \frac{R_2}{\alpha}$$

and

$$R_4 = \frac{R_2}{\beta}$$

In comparison to the twin-T circuit, the Wien-Robinson filter allows modification of the passband gain, A_0, without affecting the quality factor, Q.

If f_m is not completely suppressed due to component tolerances of R and C, a fine-tuning of the resistor $2R_2$ is required.

Figure 16–41 shows a comparison between the filter response of a passive band-rejection filter with Q = 0.25, and an active second-order filter with Q = 1, and Q = 10.

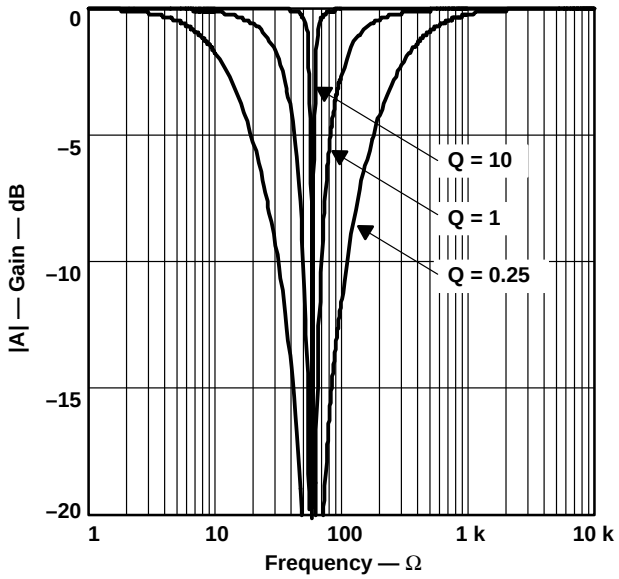


Figure 16–41. Comparison of Q Between Passive and Active Band-Rejection Filters

16.7 All-Pass Filter Design

In comparison to the previously discussed filters, an all-pass filter has a constant gain across the entire frequency range, and a phase response that changes linearly with frequency.

Because of these properties, all-pass filters are used in phase compensation and signal delay circuits.

Similar to the low-pass filters, all-pass circuits of higher order consist of cascaded first-order and second-order all-pass stages. To develop the all-pass transfer function from a low-pass response, replace A_0 with the conjugate complex denominator.

The general transfer function of an allpass is then:

$$A(s) = \frac{\prod_i \left(1 - a_i s + b_i s^2\right)}{\prod_i \left(1 + a_i s + b_i s^2\right)} \qquad (16\text{–}23)$$

with a_i and b_i being the coefficients of a partial filter. The all-pass coefficients are listed in Table 16–10 of Section 16.9.

Expressing Equation 16–23 in magnitude and phase yields:

$$A(s) = \frac{\prod_i \sqrt{\left(1 - b_i\Omega^2\right)^2 + a_i^2\Omega^2} \cdot e^{-ja}}{\prod_i \sqrt{\left(1 - b_i\Omega^2\right)^2 + a_i^2\Omega^2} \cdot e^{+ja}} \quad (16–24)$$

This gives a constant gain of 1, and a phase shift,ϕ, of:

$$\phi = -2\alpha = -2\sum_i \arctan\frac{a_i\Omega}{1 - b_i\Omega^2} \quad (16–25)$$

To transmit a signal with minimum phase distortion, the all-pass filter must have a constant group delay across the specified frequency band. The group delay is the time by which the all-pass filter delays each frequency within that band.

The frequency at which the group delay drops to $1/\sqrt{2}$–times its initial value is the corner frequency, f_C.

The group delay is defined through:

$$t_{gr} = -\frac{d\phi}{d\omega} \quad (16–26)$$

To present the group delay in normalized form, refer t_{gr} to the period of the corner frequency, T_C, of the all-pass circuit:

$$T_{gr} = \frac{t_{gr}}{T_c} = t_{gr} \cdot f_c = t_{gr} \cdot \frac{\omega_c}{2\pi} \quad (16–27)$$

Substituting t_{gr} through Equation 16–26 gives:

$$T_{gr} = -\frac{1}{2\pi} \cdot \frac{d\phi}{d\Omega} \quad (16–28)$$

Inserting the φ term in Equation 16–25 into Equation 16–28 and completing the derivation, results in:

$$T_{gr} = \frac{1}{\pi} \sum_{i} \frac{a_i\left(1 + b_i\Omega^2\right)}{1 + \left(a_1{}^2 - 2b_1\right)\cdot\Omega^2 + b_1{}^2\Omega^4} \qquad (16–29)$$

Setting $\Omega = 0$ in Equation 16–29 gives the group delay for the low frequencies, $0 < \Omega < 1$, which is:

$$T_{gr0} = \frac{1}{\pi} \sum_{i} a_i \qquad (16–30)$$

The values for T_{gr0} are listed in Table 16–10, Section 16.9, from the first to the tenth order.

In addition, Figure 16–42 shows the group delay response versus the frequency for the first ten orders of all-pass filters.

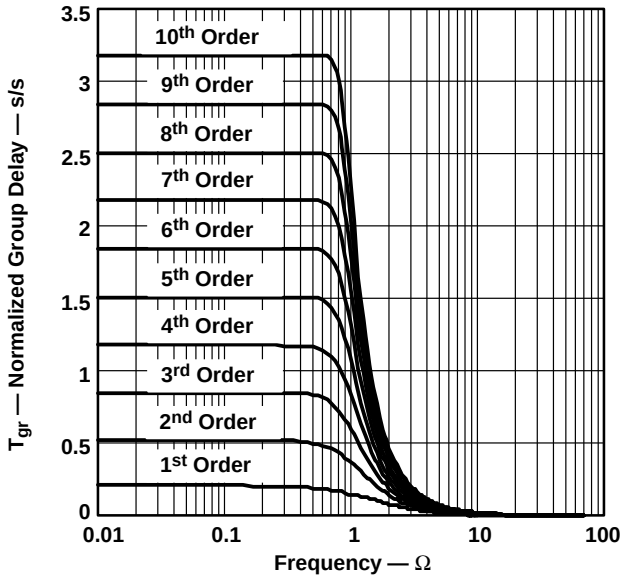


Figure 16–42. Frequency Response of the Group Delay for the First 10 Filter Orders

16.7.1 First-Order All-Pass Filter

Figure 16–43 shows a first-order all-pass filter with a gain of +1 at low frequencies and a gain of –1 at high frequencies. Therefore, the magnitude of the gain is 1, while the phase changes from 0° to –180°.

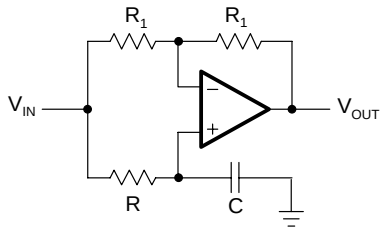


Figure 16–43. First-Order All-Pass

The transfer function of the circuit above is:

$$A(s) = \frac{1 - RC\omega_c \cdot s}{1 + RC\omega_c \cdot s}$$

The coefficient comparison with Equation 16–23 (b_1=1), results in:

$$a_i = RC \cdot 2\pi f_c \qquad (16–31)$$

To design a first-order all-pass, specify f_C and C and then solve for R:

$$R = \frac{a_i}{2\pi f_c \cdot C} \qquad (16–32)$$

Inserting Equation 16–31 into 16–30 and substituting ω_C with Equation 16–27 provides the maximum group delay of a first-order all-pass filter:

$$t_{gr0} = 2RC \qquad (16–33)$$

16.7.2 Second-Order All-Pass Filter

Figure 16–44 shows that one possible design for a second-order all-pass filter is to subtract the output voltage of a second-order band-pass filter from its input voltage.

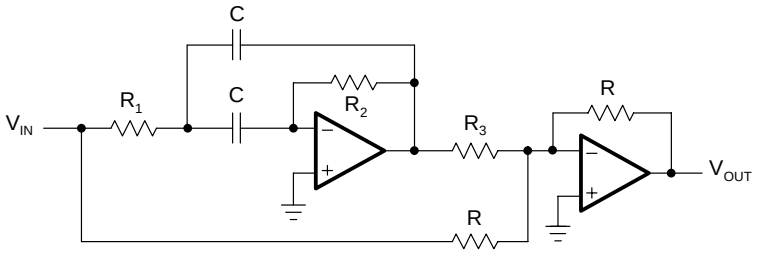


Figure 16–44. Second-Order All-Pass Filter

The transfer function of the circuit in Figure 16–44 is:

$$A(s) = \frac{1 + \left(2R_1 - \alpha R_2\right)C\omega_c \cdot s + R_1R_2C^2{\omega_c}^2 \cdot s^2}{1 + 2R_1C\omega_c \cdot s + R_1R_2C^2{\omega_c}^2 \cdot s^2}$$

The coefficient comparison with Equation 16–23 yields:

$$a_1 = 4\pi f_c R_1 C \tag{16–34}$$

$$b_1 = a_1 \pi f_c R_2 C \tag{16–35}$$

$$\alpha = \frac{{a_1}^2}{b_1} = \frac{R}{R_3} \tag{16–36}$$

To design the circuit, specify f_C, C, and R, and then solve for the resistor values:

$$R_1 = \frac{a_1}{4\pi f_c C} \tag{16–37}$$

$$R_2 = \frac{b_1}{a_1 \pi f_c C} \tag{16–38}$$

$$R_3 = \frac{R}{\alpha} \tag{16–39}$$

Inserting Equation 16–34 into Equation16–30 and substituting ω_C with Equation 16–27 gives the maximum group delay of a second-order all-pass filter:

$$t_{gr0} = 4R_1C \tag{16–40}$$

16.7.3 Higher-Order All-Pass Filter

Higher-order all-pass filters consist of cascaded first-order and second-order filter stages.

Example 16–7. 2-ms Delay All-Pass Filter

A signal with the frequency spectrum, 0 < f < 1 kHz, needs to be delayed by 2 ms. To keep the phase distortions at a minimum, the corner frequency of the all-pass filter must be $f_C \geq 1$ kHz.

Equation 16–27 determines the normalized group delay for frequencies below 1 kHz:

$$T_{gro} = \frac{t_{gr0}}{T_C} = 2\ \text{ms} \cdot 1\ \text{kHz} = 2.0$$

Figure 16–42 confirms that a seventh-order all-pass is needed to accomplish the desired delay. The exact value, however, is $T_{gr0} = 2.1737$. To set the group delay to precisely 2 ms, solve Equation 16–27 for f_C and obtain the corner frequency:

$$f_C = \frac{T_{gr0}}{t_{gr0}} = 1.087 \text{ kHz}$$

To complete the design, look up the filter coefficients for a seventh-order all-pass filter, specify C, and calculate the resistor values for each partial filter.

Cascading the first-order all-pass with the three second-order stages results in the desired seventh-order all-pass filter.

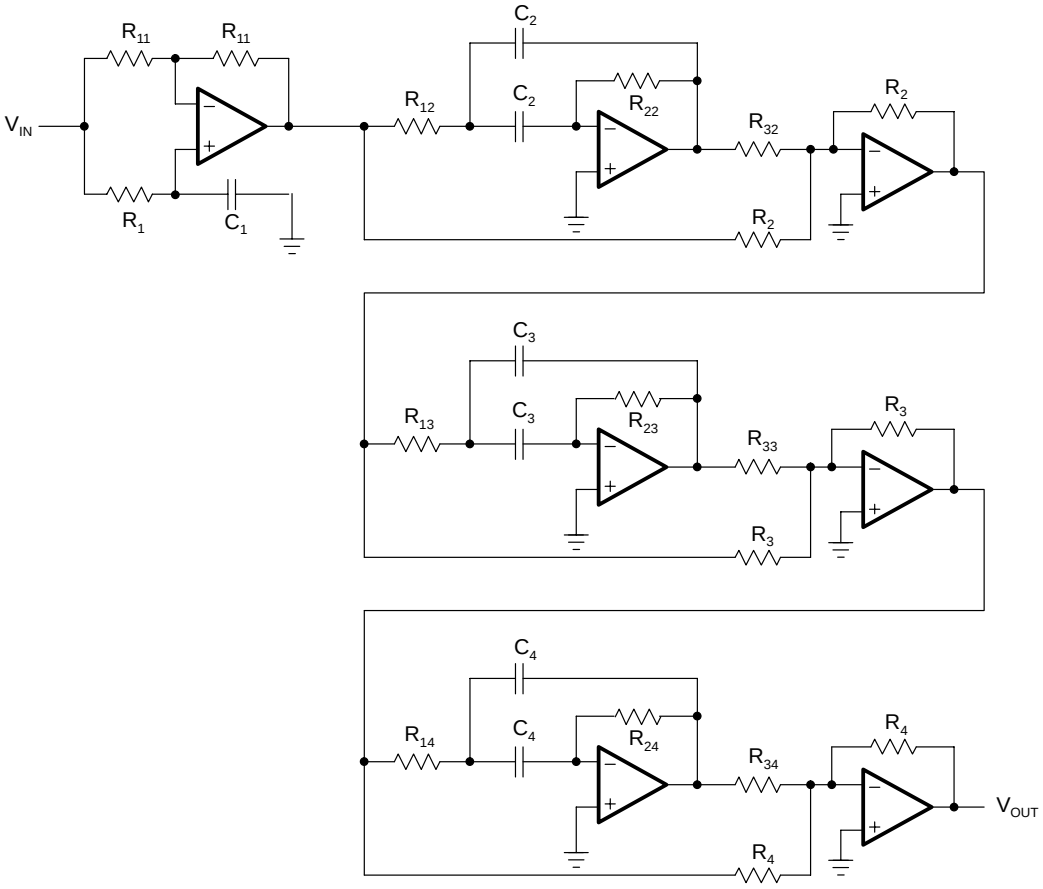


Figure 16–45. Seventh-Order All-Pass Filter

16.8 Practical Design Hints

This section introduces dc-biasing techniques for filter designs in single-supply applications, which are usually not required when operating with dual supplies. It also provides recommendations on selecting the type and value range of capacitors and resistors as well as the decision criteria for choosing the correct op amp.

16.8.1 Filter Circuit Biasing

The filter diagrams in this chapter are drawn for dual supply applications. The op amp operates from a positive and a negative supply, while the input and the output voltage are referenced to ground (Figure 16–46).

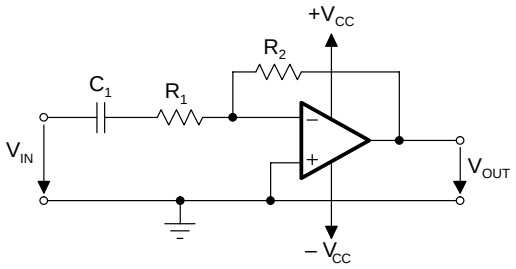


Figure 16–46. Dual-Supply Filter Circuit

For the single supply circuit in Figure 16–47, the lowest supply voltage is ground. For a symmetrical output signal, the potential of the noninverting input is level-shifted to midrail.

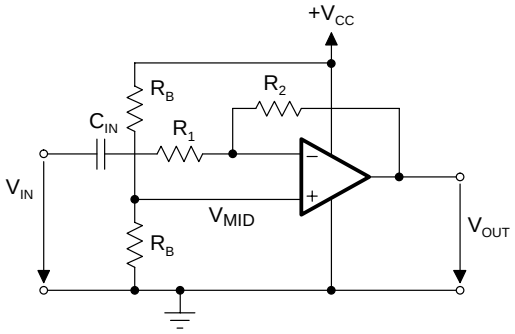


Figure 16–47. Single-Supply Filter Circuit

The coupling capacitor, C_{IN} in Figure 16–47, ac-couples the filter, blocking any unknown dc level in the signal source. The voltage divider, consisting of the two equal-bias resistors R_B, divides the supply voltage to V_{MID} and applies it to the inverting op amp input.

For simple filter input structures, passive RC networks often provide a low-cost biasing solution. In the case of more complex input structures, such as the input of a second-order

low-pass filter, the RC network can affect the filter characteristic. Then it is necessary to either include the biasing network into the filter calculations, or to insert an input buffer between biasing network and the actual filter circuit, as shown in Figure 16–48.

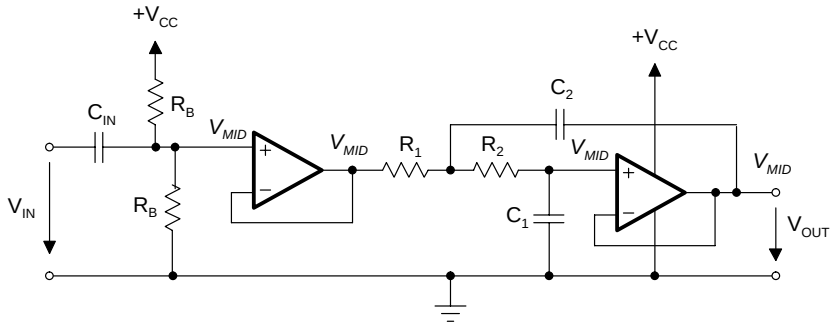


Figure 16–48. Biasing a Sallen-Key Low-Pass

C_{IN} ac-couples the filter, blocking any dc level in the signal source. V_{MID} is derived from V_{CC} via the voltage divider. The op amp operates as a voltage follower and as an impedance converter. V_{MID} is applied via the dc path, R_1 and R_2, to the noninverting input of the filter amplifier.

Note that the parallel circuit of the resistors, R_B , together with C_{IN} create a high-pass filter. To avoid any effect on the low-pass characteristic, the corner frequency of the input high-pass must be low versus the corner frequency of the actual low-pass.

The use of an input buffer causes no loading effects on the low-pass filter, thus keeping the filter calculation simple.

In the case of a higher-order filter, all following filter stages receive their bias level from the preceding filter amplifier.

Figure 16–49 shows the biasing of an multiple feedback (MFB) low-pass filter.

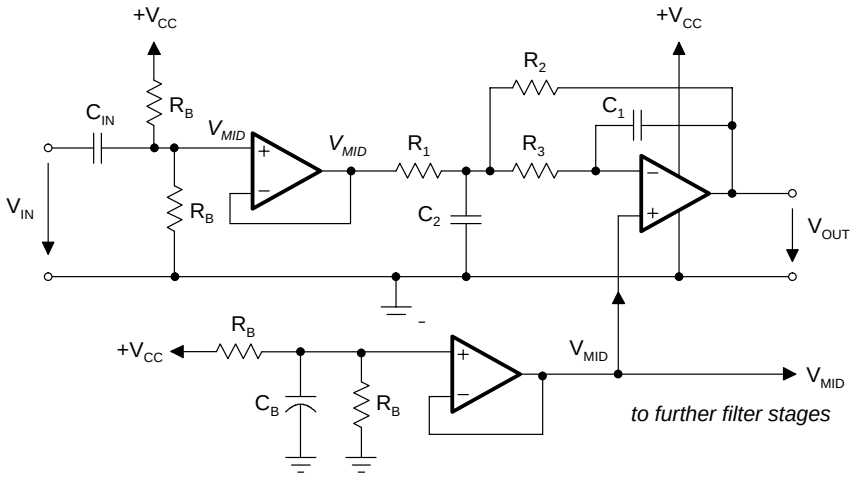


Figure 16–49. Biasing a Second-Order MFB Low-Pass Filter

The input buffer decouples the filter from the signal source. The filter itself is biased via the noninverting amplifier input. For that purpose, the bias voltage is taken from the output of a V_{MID} generator with low output impedance. The op amp operates as a difference amplifier and subtracts the bias voltage of the input buffer from the bias voltage of the V_{MID} generator, thus yielding a dc potential of V_{MID} at zero input signal.

A low-cost alternative is to remove the op amp and to use a passive biasing network instead. However, to keep loading effects at a minimum, the values for R_B must be significantly higher than without the op amp.

The biasing of a Sallen-Key and an MFB high-pass filter is shown in Figure 16–50.

The input capacitors of high-pass filters already provide the ac-coupling between filter and signal source. Both circuits use the V_{MID} generator from Figure 16–50 for biasing. While the MFB circuit is biased at the noninverting amplifier input, the Sallen-Key high-pass is biased via the only dc path available, which is R_1. In the ac circuit, the input signals travel via the low output impedance of the op amp to ground.

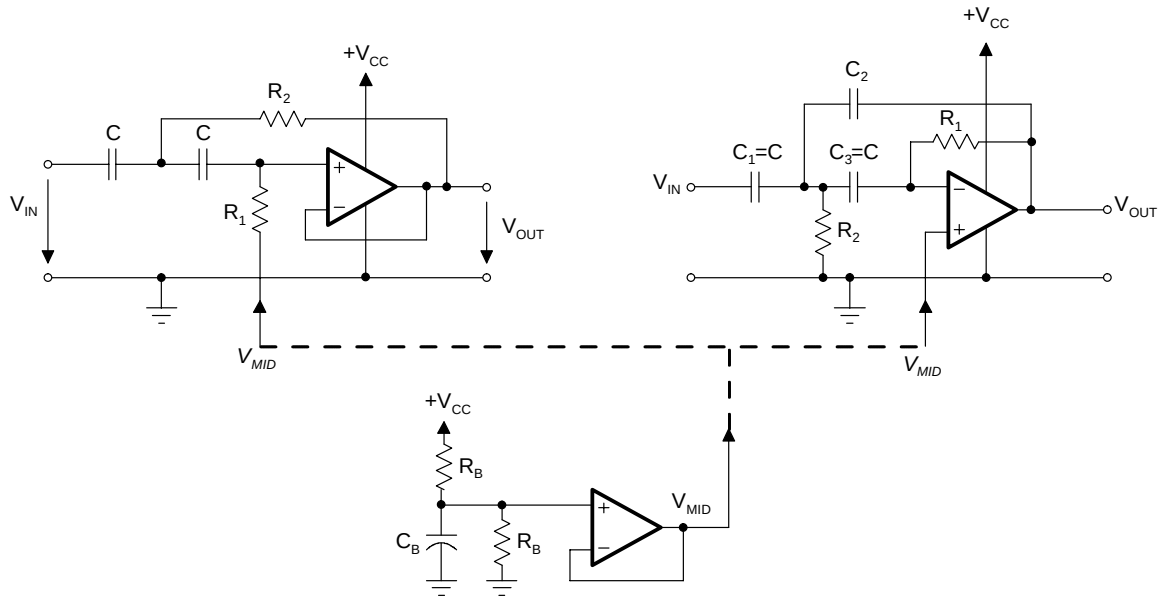


Figure 16–50. Biasing a Sallen-Key and an MFB High-Pass Filter

16.8.2 Capacitor Selection

The tolerance of the selected capacitors and resistors depends on the filter sensitivity and on the filter performance.

Sensitivity is the measure of the vulnerability of a filter's performance to changes in component values. The important filter parameters to consider are the corner frequency, f_C, and Q.

For example, when Q changes by $\pm$ 2% due to a $\pm$ 5% change in the capacitance value, then the sensitivity of Q to capacity changes is expressed as: $s\frac{Q}{C} = \frac{2\%}{5\%} = 0.4\frac{\%}{\%}$. The following sensitivity approximations apply to second-order Sallen-Key and MFB filters:

$$s\frac{Q}{C} \approx s\frac{Q}{R} \approx s\frac{f_c}{C} \approx s\frac{f_c}{R} \approx \pm\ 0.5\frac{\%}{\%}$$

Although 0.5 %/% is a small difference from the ideal parameter, in the case of higher-order filters, the combination of small Q and f_C differences in each partial filter can significantly modify the overall filter response from its intended characteristic.

Figures 16.51 and 16.52 show how an intended eighth-order Butterworth low-pass can turn into a low-pass with Tschebyscheff characteristic mainly due to capacitance changes from the partial filters.

Figure 16–51 shows the differences between the ideal and the actual frequency responses of the four partial filters. The overall filter responses are shown in Figure 16–52.

The difference between ideal and real response peaks with 0.35 dB at approximately 30 kHz, which is equivalent to an enormous 4.1% gain error can be seen.

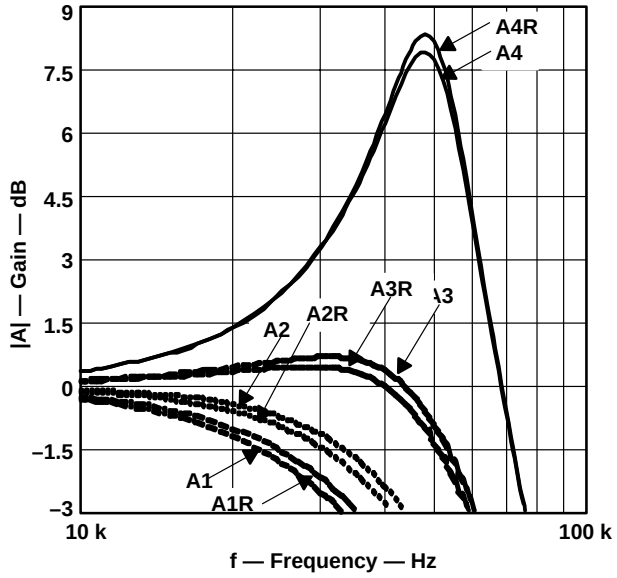


Figure 16–51. Differences in Q and f_C in the Partial Filters of an Eighth-Order Butterworth Low-Pass Filter

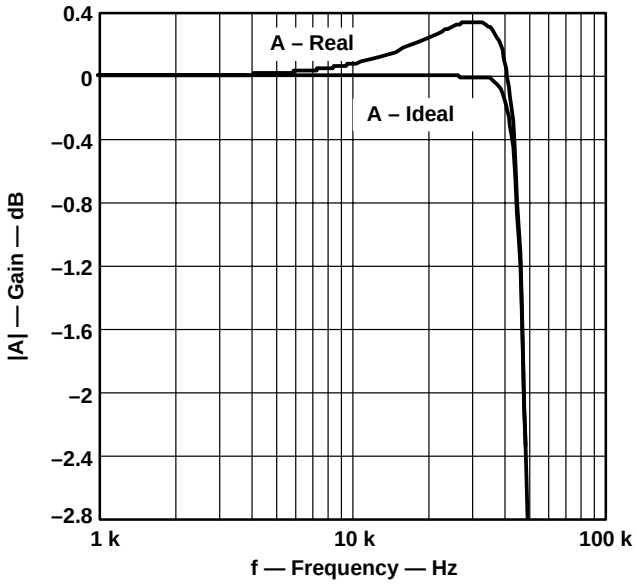


Figure 16–52. Modification of the Intended Butterworth Response to a Tschebyscheff-Type Characteristic

If this filter is intended for a data acquisition application, it could be used at best in a 4-bit system. In comparison, if the maximum full-scale error of a 12-bit system is given with ½ LSB, then maximum pass-band deviation would be – 0.001 dB, or 0.012%.

To minimize the variations of f_C and Q, NPO (COG) ceramic capacitors are recommended for high-performance filters. These capacitors hold their nominal value over a wide temperature and voltage range. The various temperature characteristics of ceramic capacitors are identified by a three-symbol code such as: COG, X7R, Z5U, and Y5V.

COG-type ceramic capacitors are the most precise. Their nominal values range from 0.5 pF to approximately 47 nF with initial tolerances from ± 0.25 pF for smaller values and up to ±1% for higher values. Their capacitance drift over temperature is typically 30ppm/°C.

X7R-type ceramic capacitors range from 100 pF to 2.2 µF with an initial tolerance of +1% and a capacitance drift over temperature of ±15%.

For higher values, tantalum electrolytic capacitors should be used.

Other precision capacitors are silver mica, metallized polycarbonate, and for high temperatures, polypropylene or polystyrene.

Since capacitor values are not as finely subdivided as resistor values, the capacitor values should be defined prior to selecting resistors. If precision capacitors are not available to provide an accurate filter response, then it is necessary to measure the individual capacitor values, and to calculate the resistors accordingly.

For high performance filters, 0.1% resistors are recommended.

16.8.3 Component Values

Resistor values should stay within the range of 1 kΩ to 100 kΩ. The lower limit avoids excessive current draw from the op amp output, which is particularly important for single-supply op amps in power-sensitive applications. Those amplifiers have typical output currents of between 1 mA and 5 mA. At a supply voltage of 5 V, this current translates to a minimum of 1 kΩ.

The upper limit of 100 kΩ is to avoid excessive resistor noise.

Capacitor values can range from 1 nF to several µF. The lower limit avoids coming too close to parasitic capacitances. If the common-mode input capacitance of the op amp, used in a Sallen-Key filter section, is close to 0.25% of C1, (C1 / 400), it must be considered for accurate filter response. The MFB topology, in comparison, does not require input-capacitance compensation.

16.8.4 Op Amp Selection

The most important op amp parameter for proper filter functionality is the unity-gain bandwidth. In general, the open-loop gain (A_{OL}) should be 100 times (40 dB above) the peak gain (Q) of a filter section to allow a maximum gain error of 1%.

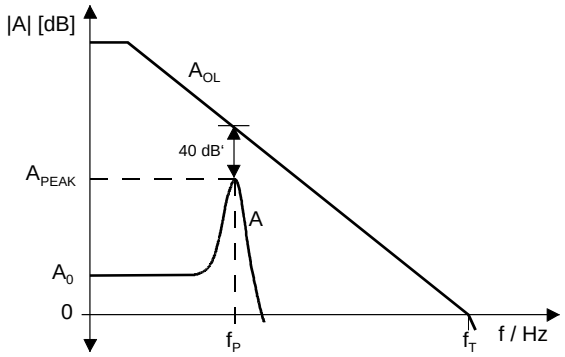


Figure 16–53. Open-Loop Gain (A_{OL}) and Filter Response (A)

The following equations are good rules of thumb to determine the necessary unity-gain bandwidth of an op amp for an individual filter section.

1) First-order filter:

$$f_T = 100 \cdot Gain \cdot f_c$$

2) Second-order filter (Q < 1):

$$f_T = 100 \cdot Gain \cdot f_c \cdot k_i \qquad \text{with} \qquad k_i = \frac{f_{ci}}{f_c}$$

3) Second-order filter (Q > 1):

$$f_T = 100 \cdot Gain \cdot \frac{f_c}{a_i} \sqrt{\frac{Q_i^2 - 0.5}{Q_i^2 - 0.25}}$$

For example, a fifth-order, 10-kHz, Tschebyscheff low-pass filter with 3-dB passband ripple and a dc gain of $A_0 = 2$ has its worst case Q in the third filter section. With $Q_3 = 8.82$ and $a_3 = 0.1172$, the op amp needs to have a unity-gain bandwidth of:

$$f_T = 100 \cdot 2 \cdot \frac{10\ kHz}{0.1172} \sqrt{\frac{8.82^2 - 0.5}{8.82^2 - 0.25}} = 17\ MHz$$

In comparison, a fifth-order unity-gain, 10-kHz, Butterworth low-pass filter has a worst case Q of $Q_3 = 1.62$; $a_3 = 0.618$. Due to the lower Q value, f_T is also lower and calculates to only:

$$f_T = 100 \cdot \frac{10\ \text{kHz}}{0.618} \sqrt{\frac{1.62^2 - 0.5}{1.62^2 - 0.25}} = 1.5\ \text{MHz}$$

Besides good dc performance, low noise, and low signal distortion, another important parameter that determines the speed of an op amp is the slew rate (SR). For adequate full-power response, the slew rate must be greater than:

$$SR = \pi \cdot V_{PP} \cdot f_C$$

For example, a single-supply, 100-kHz filter with 5 V_{PP} output requires a slew rate of at least:

$$SR = \pi \cdot 5V \cdot 100\ \text{kHz} = 1.57 \frac{V}{\mu s}$$

Texas Instruments offers a wide range of op amps for high-performance filters in single supply applications. Table 16–3 provides a selection of single-supply amplifiers sorted in order of rising slew rate.

Table 16–3. Single-Supply Op Amp Selection Guide (T_A = 25°C, V_{CC} = 5 V)

OP AMP	BW (MHz)	FPR (kHz)	SR (V/μs)	V_{IO} (mV)	Noise (nV/√Hz)
TLV2721	0.51	11	0.18	0.6	20
TLC2201A	1.8	159	2.5	0.6	8
TLV2771A	4.8	572	9	1.9	21
TLC071	10	1000	16	1.5	7
TLE2141	5.9	2800	45	0.5	10.5
THS4001	270	127 MHz ($1V_{PP}$)	400	6	7.5

16.9 Filter Coefficient Tables

The following tables contain the coefficients for the three filter types, Bessel, Butterworth and Tschebyscheff. The Tschebyscheff tables (Table 16–9) are split into categories for the following passband ripples: 0.5 dB, 1 dB, 2 dB, and 3 dB.

The table headers consist of the following quantities:

- n is the filter order
- i is the number of the partial filter.
- a_i, b_i are the filter coefficients.
- k_i is the ratio of the corner frequency of a partial filter, f_{Ci} , to the corner frequency of the overall filter, f_C. This ratio is used to determine the unity-gain bandwidth of the op amp, as well as to simplify the test of a filter design by measuring f_{Ci} and comparing it to f_C.
- Q_i is the quality factor of the partial filter.
- f_i / f_C this ratio is used for test purposes of the allpass filters, where f_i is the frequency, at which the phase is 180° for a second-order filter, respectively 90° for a first-order all-pass.
- T_{gr0} is the normalized group delay of the overall all-pass filter.

Table 16–4. Bessel Coefficients

n	i	a_i	b_i	$k_i = f_{Ci} / f_C$	Q_i
1	1	1.0000	0.0000	1.000	—
2	1	1.3617	0.6180	1.000	0.58
3	1	0.7560	0.0000	1.323	—
	2	0.9996	0.4772	1.414	0.69
4	1	1.3397	0.4889	0.978	0.52
	2	0.7743	0.3890	1.797	0.81
5	1	0.6656	0.0000	1.502	—
	2	1.1402	0.4128	1.184	0.56
	3	0.6216	0.3245	2.138	0.92
6	1	1.2217	0.3887	1.063	0.51
	2	0.9686	0.3505	1.431	0.61
	3	0.5131	0.2756	2.447	1.02
7	1	0.5937	0.0000	1.648	—
	2	1.0944	0.3395	1.207	0.53
	3	0.8304	0.3011	1.695	0.66
	4	0.4332	0.2381	2.731	1.13
8	1	1.1112	0.3162	1.164	0.51
	2	0.9754	0.2979	1.381	0.56
	3	0.7202	0.2621	1.963	0.71
	4	0.3728	0.2087	2.992	1.23
9	1	0.5386	0.0000	1.857	—
	2	1.0244	0.2834	1.277	0.52
	3	0.8710	0.2636	1.574	0.59
	4	0.6320	0.2311	2.226	0.76
	5	0.3257	0.1854	3.237	1.32
10	1	1.0215	0.2650	1.264	0.50
	2	0.9393	0.2549	1.412	0.54
	3	0.7815	0.2351	1.780	0.62
	4	0.5604	0.2059	2.479	0.81
	5	0.2883	0.1665	3.466	1.42

Table 16–5. Butterworth Coefficients

n	i	a_i	b_i	$k_i = f_{Ci} / f_C$	Q_i
1	1	1.0000	0.0000	1.000	—
2	1	1.4142	1.0000	1.000	0.71
3	1	1.0000	0.0000	1.000	—
	2	1.0000	1.0000	1.272	1.00
4	1	1.8478	1.0000	0.719	0.54
	2	0.7654	1.0000	1.390	1.31
5	1	1.0000	0.0000	1.000	—
	2	1.6180	1.0000	0.859	0.62
	3	0.6180	1.0000	1.448	1.62
6	1	1.9319	1.0000	0.676	0.52
	2	1.4142	1.0000	1.000	0.71
	3	0.5176	1.0000	1.479	1.93
7	1	1.0000	0.0000	1.000	—
	2	1.8019	1.0000	0.745	0.55
	3	1.2470	1.0000	1.117	0.80
	4	0.4450	1.0000	1.499	2.25
8	1	1.9616	1.0000	0.661	0.51
	2	1.6629	1.0000	0.829	0.60
	3	1.1111	1.0000	1.206	0.90
	4	0.3902	1.0000	1.512	2.56
9	1	1.0000	0.0000	1.000	—
	2	1.8794	1.0000	0.703	0.53
	3	1.5321	1.0000	0.917	0.65
	4	1.0000	1.0000	1.272	1.00
	5	0.3473	1.0000	1.521	2.88
10	1	1.9754	1.0000	0.655	0.51
	2	1.7820	1.0000	0.756	0.56
	3	1.4142	1.0000	1.000	0.71
	4	0.9080	1.0000	1.322	1.10
	5	0.3129	1.0000	1.527	3.20

Table 16–6. Tschebyscheff Coefficients for 0.5-dB Passband Ripple

n	i	a_i	b_i	$k_i = f_{Ci} / f_C$	Q_i
1	1	1.0000	0.0000	1.000	—
2	1	1.3614	1.3827	1.000	0.86
3	1	1.8636	0.0000	0.537	—
	2	0.0640	1.1931	1.335	1.71
4	1	2.6282	3.4341	0.538	0.71
	2	0.3648	1.1509	1.419	2.94
5	1	2.9235	0.0000	0.342	—
	2	1.3025	2.3534	0.881	1.18
	3	0.2290	1.0833	1.480	4.54
6	1	3.8645	6.9797	0.366	0.68
	2	0.7528	1.8573	1.078	1.81
	3	0.1589	1.0711	1.495	6.51
7	1	4.0211	0.0000	0.249	—
	2	1.8729	4.1795	0.645	1.09
	3	0.4861	1.5676	1.208	2.58
	4	0.1156	1.0443	1.517	8.84
8	1	5.1117	11.9607	0.276	0.68
	2	1.0639	2.9365	0.844	1.61
	3	0.3439	1.4206	1.284	3.47
	4	0.0885	1.0407	1.521	11.53
9	1	5.1318	0.0000	0.195	—
	2	2.4283	6.6307	0.506	1.06
	3	0.6839	2.2908	0.989	2.21
	4	0.2559	1.3133	1.344	4.48
	5	0.0695	1.0272	1.532	14.58
10	1	6.3648	18.3695	0.222	0.67
	2	1.3582	4.3453	0.689	1.53
	3	0.4822	1.9440	1.091	2.89
	4	0.1994	1.2520	1.381	5.61
	5	0.0563	1.0263	1.533	17.99

Table 16–7. Tschebyscheff Coefficients for 1-dB Passband Ripple

n	i	a_i	b_i	$k_i = f_{Ci} / f_C$	Q_i
1	1	1.0000	0.0000	1.000	—
2	1	1.3022	1.5515	1.000	0.96
3	1	2.2156	0.0000	0.451	—
	2	0.5442	1.2057	1.353	2.02
4	1	2.5904	4.1301	0.540	0.78
	2	0.3039	1.1697	1.417	3.56
5	1	3.5711	0.0000	0.280	—
	2	1.1280	2.4896	0.894	1.40
	3	0.1872	1.0814	1.486	5.56
6	1	3.8437	8.5529	0.366	0.76
	2	0.6292	1.9124	1.082	2.20
	3	0.1296	1.0766	1.493	8.00
7	1	4.9520	0.0000	0.202	—
	2	1.6338	4.4899	0.655	1.30
	3	0.3987	1.5834	1.213	3.16
	4	0.0937	1.0432	1.520	10.90
8	1	5.1019	14.7608	0.276	0.75
	2	0.8916	3.0426	0.849	1.96
	3	0.2806	1.4334	1.285	4.27
	4	0.0717	1.0432	1.520	14.24
9	1	6.3415	0.0000	0.158	—
	2	2.1252	7.1711	0.514	1.26
	3	0.5624	2.3278	0.994	2.71
	4	0.2076	1.3166	1.346	5.53
	5	0.0562	1.0258	1.533	18.03
10	1	6.3634	22.7468	0.221	0.75
	2	1.1399	4.5167	0.694	1.86
	3	0.3939	1.9665	1.093	3.56
	4	0.1616	1.2569	1.381	6.94
	5	0.0455	1.0277	1.532	22.26

Table 16–8. Tschebyscheff Coefficients for 2-dB Passband Ripple

n	i	a_i	b_i	$k_i = f_{Ci} / f_C$	Q_i
1	1	1.0000	0.0000	1.000	—
2	1	1.1813	1.7775	1.000	1.13
3	1	2.7994	0.0000	0.357	—
	2	0.4300	1.2036	1.378	2.55
4	1	2.4025	4.9862	0.550	0.93
	2	0.2374	1.1896	1.413	4.59
5	1	4.6345	0.0000	0.216	—
	2	0.9090	2.6036	0.908	1.78
	3	0.1434	1.0750	1.493	7.23
6	1	3.5880	10.4648	0.373	0.90
	2	0.4925	1.9622	1.085	2.84
	3	0.0995	1.0826	1.491	10.46
7	1	6.4760	0.0000	0.154	—
	2	1.3258	4.7649	0.665	1.65
	3	0.3067	1.5927	1.218	4.12
	4	0.0714	1.0384	1.523	14.28
8	1	4.7743	18.1510	0.282	0.89
	2	0.6991	3.1353	0.853	2.53
	3	0.2153	1.4449	1.285	5.58
	4	0.0547	1.0461	1.518	18.39
9	1	8.3198	0.0000	0.120	—
	2	1.7299	7.6580	0.522	1.60
	3	0.4337	2.3549	0.998	3.54
	4	0.1583	1.3174	1.349	7.25
	5	0.0427	1.0232	1.536	23.68
10	1	5.9618	28.0376	0.226	0.89
	2	0.8947	4.6644	0.697	2.41
	3	0.3023	1.9858	1.094	4.66
	4	0.1233	1.2614	1.380	9.11
	5	0.0347	1.0294	1.531	29.27

Table 16–9. Tschebyscheff Coefficients for 3-dB Passband Ripple

n	i	a_i	b_i	$k_i = f_{Ci} / f_C$	Q_i
1	1	1.0000	0.0000	1.000	—
2	1	1.0650	1.9305	1.000	1.30
3	1	3.3496	0.0000	0.299	—
	2	0.3559	1.1923	1.396	3.07
4	1	2.1853	5.5339	0.557	1.08
	2	0.1964	1.2009	1.410	5.58
5	1	5.6334	0.0000	0.178	—
	2	0.7620	2.6530	0.917	2.14
	3	0.1172	1.0686	1.500	8.82
6	1	3.2721	11.6773	0.379	1.04
	2	0.4077	1.9873	1.086	3.46
	3	0.0815	1.0861	1.489	12.78
7	1	7.9064	0.0000	0.126	—
	2	1.1159	4.8963	0.670	1.98
	3	0.2515	1.5944	1.222	5.02
	4	0.0582	1.0348	1.527	17.46
8	1	4.3583	20.2948	0.286	1.03
	2	0.5791	3.1808	0.855	3.08
	3	0.1765	1.4507	1.285	6.83
	4	0.0448	1.0478	1.517	22.87
9	1	10.1759	0.0000	0.098	—
	2	1.4585	7.8971	0.526	1.93
	3	0.3561	2.3651	1.001	4.32
	4	0.1294	1.3165	1.351	8.87
	5	0.0348	1.0210	1.537	29.00
10	1	5.4449	31.3788	0.230	1.03
	2	0.7414	4.7363	0.699	2.94
	3	0.2479	1.9952	1.094	5.70
	4	0.1008	1.2638	1.380	11.15
	5	0.0283	1.0304	1.530	35.85

Table 16–10. All-Pass Coefficients

n	i	a_i	b_i	f_i / f_C	Q_i	T_{gr0}
1	1	0.6436	0.0000	1.554	—	0.2049
2	1	1.6278	0.8832	1.064	0.58	0.5181
3	1	1.1415	0.0000	0.876	—	0.8437
	2	1.5092	1.0877	0.959	0.69	
4	1	2.3370	1.4878	0.820	0.52	1.1738
	2	1.3506	1.1837	0.919	0.81	
5	1	1.2974	0.0000	0.771	—	1.5060
	2	2.2224	1.5685	0.798	0.56	
	3	1.2116	1.2330	0.901	0.92	
6	1	2.6117	1.7763	0.750	0.51	1.8395
	2	2.0706	1.6015	0.790	0.61	
	3	1.0967	1.2596	0.891	1.02	
7	1	1.3735	0.0000	0.728	—	2.1737
	2	2.5320	1.8169	0.742	0.53	
	3	1.9211	1.6116	0.788	0.66	
	4	1.0023	1.2743	0.886	1.13	
8	1	2.7541	1.9420	0.718	0.51	2.5084
	2	2.4174	1.8300	0.739	0.56	
	3	1.7850	1.6101	0.788	0.71	
	4	0.9239	1.2822	0.883	1.23	
9	1	1.4186	0.0000	0.705	—	2.8434
	2	2.6979	1.9659	0.713	0.52	
	3	2.2940	1.8282	0.740	0.59	
	4	1.6644	1.6027	0.790	0.76	
	5	0.8579	1.2862	0.882	1.32	
10	1	2.8406	2.0490	0.699	0.50	3.1786
	2	2.6120	1.9714	0.712	0.54	
	3	2.1733	1.8184	0.742	0.62	
	4	1.5583	1.5923	0.792	0.81	
	5	0.8018	1.2877	0.881	1.42	

16.10 References

D.Johnson and J.Hilburn, *Rapid Practical Designs of Active Filters*, John Wiley & Sons, 1975.

U.Tietze and Ch.Schenk, *Halbleiterschaltungstechnik*, Springer–Verlag, 1980.

H.Berlin, *Design of Active Filters with Experiments*, Howard W.Sams & Co., 1979.

M.Van Falkenburg, *Analog Filter Design*, Oxford University Press, 1982.

S.Franko, *Design with Operational Amplifiers and Analog Integrated Circuits*, McGraw–Hill, 1988

Chapter 17

Circuit Board Layout Techniques

Bruce Carter

17.1 General Considerations

Prior discussions have focused on how to design op amp circuitry, how to use ICs, and the usage of associated passive components. There is one additional circuit component that must be considered for the design to be a success — the printed circuit board on which the circuit is to be located.

17.1.1 The PCB is a Component of the Op Amp Design

Op amp circuitry is analog circuitry, and is very different from digital circuitry. It must be partitioned in its own section of the board, using special layout techniques.

Printed circuit board effects become most apparent in high-speed analog circuits, but common mistakes described in this chapter can even affect the performance of audio circuits. The purpose of this chapter is to discuss some of the more common mistakes made by designers and how they degrade performance, and provide simple fixes to avoid the problems.

In all but very rare cases, the PCB layout for analog circuitry must be designed such that the effect of the PCB is transparent to the circuit. Any effect caused by the PCB itself should be minimized, so that the operation of the analog circuitry in production will be the same as the performance of the design and prototype.

17.1.2 Prototype, Prototype, PROTOTYPE!

Normal design cycles, particularly of large digital boards, dictate layout of the PCB as soon as possible. The digital circuitry has been simulated, but in most cases, the production PCB itself is the prototype, and may even be sold to a customer. Digital designers can correct small mistakes by implementing *cuts and jumpers*, reprogramming gate arrays or flash memories, and go on to the next project. This is not the case with analog circuitry. Some of the common design mistakes discussed in this chapter cannot be corrected by the cut and jumper method. They can and will render the entire PCB unusable. It is very

important for the digital designer, who is used to cuts and jumpers, to read and understand this chapter prior to releasing a board to a layout service.

A little care, taken up front, can save a board worth thousands of dollars from becoming scrap because of blunders in a tiny section of analog circuitry. This author has been the unfortunate recipient of a simple analog circuit designed by another engineer, who was accustomed to the cut and jumper method of correcting his mistakes. This resulted in a design that was full of mistakes. Not only was the op amp hooked up with inverting and noninverting inputs reversed, but an RC time constant had to be added to prevent a race condition. Repercussions from these mistakes, and associated rework problems, caused literally hundreds of hours to be lost from a tight production schedule. Prototyping this circuit would have taken less than a day. PROTOTYPE ALL ANALOG CIRCUITRY!

17.1.3 Noise Sources

Noise is the primary limitation on analog circuitry performance. Internal op amp noise is covered in Chapter 10. Other types of noise include:

- Conducted Emissions — noise that the analog circuitry generates through its connections to other circuits. This is usually negligible in analog circuitry, unless it is high power (such as an audio amplifier that draws heavy currents from its power supply).
- Radiated Emissions — noise that the analog circuitry generates, or transmits, through the air. This is also usually negligible in analog circuitry, unless it is high frequency such as video.
- Conducted Susceptibility — noise from external circuitry that is conducted into the analog circuit through its connections to other circuits. Analog circuitry must be connected to the “outside world” by at least a ground connection, a power connection, an input, and an output. Noise can be conducted into the circuit through all of these paths, as well as any others that are present.
- Radiated Susceptibility — noise that is received through the air (or transmitted into the analog circuitry) from external sources. Analog circuitry, in many cases, resides on a PCB that may have high-speed digital logic including DSP chips. High-speed clocks and switching digital signals create considerable radio frequency interference (RFI). Other sources of radiated noise are endless: the switching power supply in a digital system, cellular telephones, broadcast radio and TV, fluorescent lighting, nearby PCs, lightning in thunderstorms, and so on. Even if the analog circuitry is primarily audio in frequency, RFI may produce noticeable noise in the output.

17.2 PCB Mechanical Construction

It is important to choose a PCB with the right mechanical characteristics for the application.

17.2.1 Materials — Choosing the Right One for the Application

PC Board materials are available in various grades, as defined by the National Electrical Manufacturers Association (NEMA). It would be very convenient for designers if this organization was closely allied with the electronics industry — controlling parameters such as resistivity and dielectric constant of the material. Unfortunately, that is not the case. NEMA is an electrical safety organization, and the different PCB grades primarily describe the flammability, high temperature stability, and moisture absorption of the board. Therefore, specifying a given NEMA grade does not guarantee electrical parameters of the material. If this becomes critical for an application, consult the manufacturer of the raw board stock.

Laminated materials are designated with *FR* (flame resistant) and *G* grades. FR–1 is the least flame resistant, and FR–5 is the most. G10 and G11 have special characteristics as described in Table 17–1.

Table 17–1. PCB Materials

GRADE DESIGNATION	MATERIAL/COMMENTS
FR–1	Paper/phenolic: room temperature punchable, poor moisture resistance.
FR–2	Paper/phenolic: suitable for single-sided PCB consumer equipment, good moisture resistance.
FR–3	Paper/epoxy: designed for balance of good mechanical and electrical characteristics.
FR–4	Glass cloth/epoxy: excellent mechanical and electrical properties.
FR–5	Glass cloth/epoxy: high strength at elevated temperatures, self-extinguishing.
G10	Glass cloth/epoxy: high insulation resistance, highest bond strength of glass laminates, high humidity resistance.
G11	Glass cloth/epoxy: high flexural strength retention at high temperature, extreme resistance to solvents.

Do not use FR–1. There are many examples of boards with burned spots, where high wattage components have heated a section of the board for a period of time. This grade of PCB material has more in common with cardboard than anything else.

FR–4 is commonly used in industrial-quality equipment, while FR–2 is used in high-volume consumer applications. These two board materials appear to be industry standards. Deviating from these standards can limit the number of raw board material suppliers and PCB houses that can fabricate the board because their tooling is already set up for these materials. Nevertheless, there are applications in which one of the other grades may make sense. For very high frequency applications, it may even be necessary to consider

Teflon or even ceramic board substrate. One thing can be counted on, however: the more exotic the board substrate, the more expensive it will be.

In selecting a board material, pay careful attention to the moisture absorption. Just about every desirable performance characteristic of the board will be negatively impacted by moisture. This includes surface resistance of the board, dielectric leakage, high-voltage breakdown and arcing, and mechanical stability. Also, pay attention to the operating temperature. High operating temperatures can occur in unexpected places, such as in proximity to large digital ICs that are switching at high speeds. Be aware that heat rises, so if one of those 500-pin monster ICs is located directly under a sensitive analog circuit, both the PCB and circuit characteristics may vary with the temperature.

After the board substrate material has been selected, the next decision is how thick to make the copper foil laminated to it. For most applications, 1-ounce copper is sufficient. If the circuit consumes a lot of power, 2-ounce may be better. Avoid ½-ounce copper, because it tends to break between the trace and the pad.

17.2.2 How Many Layers are Best?

Depending on the complexity of the overall circuitry being designed, a designer must decide how many layers the PCB should be.

17.2.2.1 Single-Sided

Very simple consumer electronics are sometimes fabricated on single-sided PCBs, keeping the raw board material inexpensive (FR–1 or FR–2) with thin copper cladding. These designs frequently include many jumper wires, simulating the circuit routing on a double-sided board. This technique is only recommended for low-frequency circuitry. For reasons described below, this type of design is extremely susceptible to radiated noise. It is harder to design a board of this type, because of the many things can go wrong. Many complex designs have been successfully implemented with this technique, but they require a lot of forethought. An example is a television set that puts all of the analog circuitry on a single-sided board at the bottom of the case, and uses the metallized CRT itself to shield the board from a separate digital tuning board near the top of the set. Be prepared to get creative if the design demands high volume, low cost PCBs.

If a single-sided PCB is a requirement, remember the trace resistance! The op amp is not a 3-terminal device (inverting input, noninverting input, and output). It is a 7-terminal device as shown in Figure 17–1.

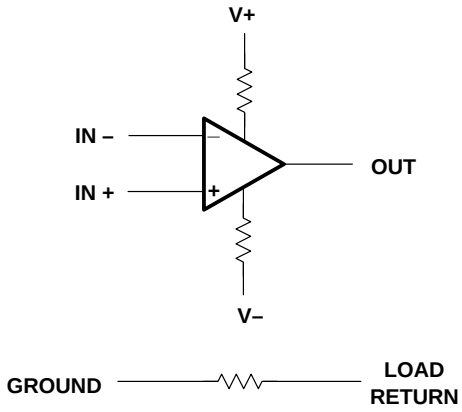


Figure 17–1. Op Amp Terminal Model

Resistance in the power leads of the device must be taken into account, as well as resistance between the return for the inputs and the load. Both the input and load must return somewhere. If these are at different potentials due to trace resistance, there can be problems.

17.2.2.2 Double-Sided

The next level of complexity is double-sided. Although there are some double-sided FR–2 boards, they are more commonly fabricated with FR–4 material. The increased strength of FR–4 material supports vias better. Doubled-sided boards are easier to route because there are two layers of foil, and it is possible to route signals by crossing traces on different layers. Crossing traces, however, is not recommended for analog circuitry. Wherever possible, the bottom layer should be devoted to a ground plane, and all other signals routed on the top layer. A ground plane provides several benefits:

- Ground is frequently the most common connection in the circuit. Having it continuous on the bottom layer usually makes the most sense for circuit routing.
- It increases mechanical strength of the board.
- It lowers the impedance of all ground connections in the circuit, which reduces undesirable conducted noise.
- It adds a distributed capacitance to every net in the circuit — helping to suppress radiated noise.
- It acts a shield to radiated noise coming from underneath the board.

17.2.2.3 Multi-Layer

Double-sided boards, in spite of their benefits, are not the best method of construction, especially for sensitive or high-speed designs. The most common board thickness is 1.5 mm. This separation is too great for full realization of some of the benefits listed above. Distributed capacitance, for example, is very low due to the separation.

Critical designs call for multi-layer boards. Some of the reasons are obvious:

- Better routing for power as well as ground connections. If the power is also on a plane, it is available to all points in the circuit simply by adding vias.
- Other layers are available for signal routing, making routing easier.
- There will be distributed capacitance between the power and ground planes, reducing high frequency noise.

There are other reasons for multi-layer boards, however, that may not be obvious or intuitive.

- Better EMI/RFI rejection. There is due to the *image plane effect*, which has been known since the time of Marconi. When a conductor is placed close to a parallel conductive surface, most of the high frequency currents will return directly under the conductor, flowing in the opposite direction. This mirror image of the conductor within the plane creates a transmission line. Since currents are equal and opposite in the transmission line, it is relatively immune to radiated noise. It also couples the signal very efficiently. The image plane effect works equally well with ground and power planes, but they must be continuous. Any gap or discontinuity causes the beneficial effects to quickly vanish. There is more on this in the following paragraphs.
- Reduced overall project cost for small production runs. Although multi-layer boards are more expensive to manufacture, EMI/RFI requirements from the FCC or other agencies may require expensive testing of the design. If there are problems, it can force a complete redesign of the PCB, leading to additional rounds of testing. A multi-layer PCB can have as much as 20-dB better EMI/RFI performance over a 2-layer PCB. If production volumes are going to be small, it makes sense to make a better PCB to begin with, than try to cut costs and take the risk of failing $25,000 to $50,000 tests.

17.2.3 Board Stack-Up — The Order of Layers

There has been a lot of confusion in the past over what is the optimum order for PCB layers. Take, for example, a 4-layer board consisting of two signal layers, a power plane, and a ground plane. Is it better to route the signal traces between the layers, thus providing shielding for the signal traces – or is it better to make the ground and power planes the two inner planes?

In considering this question, it is important to remember that no matter what is decided, there will still be signals exposed on one or both of the top and bottom planes. The leads of the op amp PCB package, and the traces on the board leading to nearby passive com-

ponents and feed-throughs will be exposed. Therefore, any shielding effects are compromised. It is far better to take advantage of the distributed capacitance between the power and ground plane by making them internal.

Another advantage of placing the planes internally is that the signal traces are available for probing and modification on the top and bottom layers. Anyone who has had to change connections on buried traces will appreciate this feature.

For more than four layers, it is a general rule to shield higher speed signals between the ground and power planes, and route slower signals on the outer layers.

17.3 Grounding

Good grounding is a system-level design consideration. It should be planned into the product from the first conceptual design reviews.

17.3.1 The Most Important Rule: Keep Grounds Separate

Separate grounding for analog and digital portions of circuitry is one of the simplest and most effective methods of noise suppression. One or more layers on multi-layer PCBs are usually devoted to ground planes. If the designer is not careful, the analog circuitry will be connected directly to these ground planes. The analog circuitry return, after all, is the same net in the netlist as digital return. Autorouters respond accordingly and connect all of the grounds together, creating a disaster.

After the fact separation of grounds on a mixed digital and analog board is almost impossible. Every ground connection in the analog circuitry must be lifted from the board and connected together. For surface mount boards, this results in a colossal mess of “tombstoned” passive components and floating IC leads.

17.3.2 Other Ground Rules

- Ground and power planes are at the same ac potential, due to decoupling capacitors and distributed capacitance. Therefore, it is important to isolate the power planes as well.
- Do not overlap digital and analog planes (Figure 17–2). Place analog power coincident with analog ground, and digital power coincident with digital ground. If any portion of analog and digital planes overlap, the distributed capacitance between the overlapping portions will couple high-speed digital noise into the analog circuitry. This defeats the purpose of isolated planes.

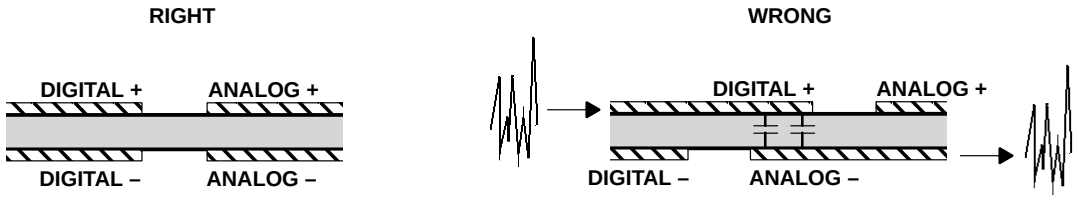


Figure 17–2. Digital and Analog Plane Placement

- *Separate grounds* does not mean that the grounds are electrically separate in the system. They have to be common at some point, preferably a single, low-impedance point. System-wise, there is only one ground, and it is the electrical safety ground in an ac-powered system or battery ground in a dc-powered system. Everything else "returns" to that ground. It would be a good idea to develop the discipline to refer to everything that is not a *ground* as a *return*. All returns should be connected together at a single point, which is system ground. At some point, this will be the chassis. It is important to avoid ground loops by multiple connections to the chassis. Ensuring only one chassis ground point is one of the most difficult aspects of system design.

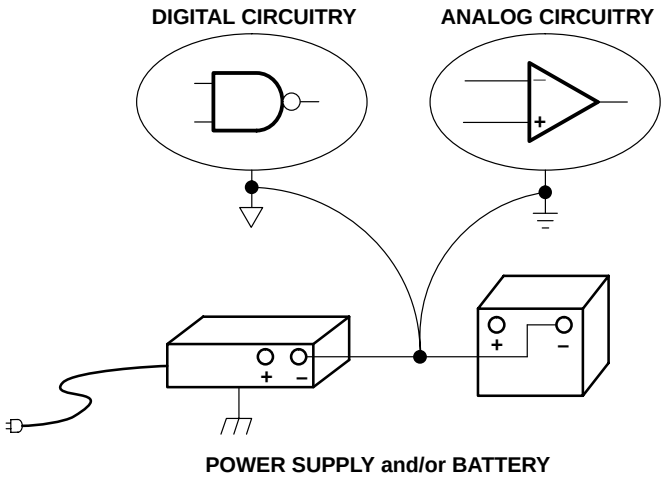


Figure 17–3. Separate Grounds

- If at all possible, dedicate separate connector pins to separate returns, and combine the returns only at system ground. Aging and repeated mating causes connector pins to increase in contact resistance, so several pins are needed. Many digital boards consist of many layers and hundreds or thousands of nets. The addition of one more net is seldom an issue, but the addition of several connector pins almost always is. If this cannot be done, then it will be necessary to make the two returns a single net on the PCB — with very special routing precautions.
- It is important to keep digital signals away from analog portions of the circuit. It makes little sense to isolate planes, keep analog traces short, and place passive

components carefully if there are high-speed digital traces running right next to the sensitive analog traces. Digital signals must be routed around analog circuitry, and not overlap analog ground and power planes. If not, the design will include a new schematic symbol shown in Figure 17–4 — the broadcasting antenna!

Most digital clocks are high enough in frequency that even small capacitances between traces and planes can couple significant noise. Remember that it is not only the fundamental frequency of the clock that can cause a potential problem, but also the higher frequency harmonics.

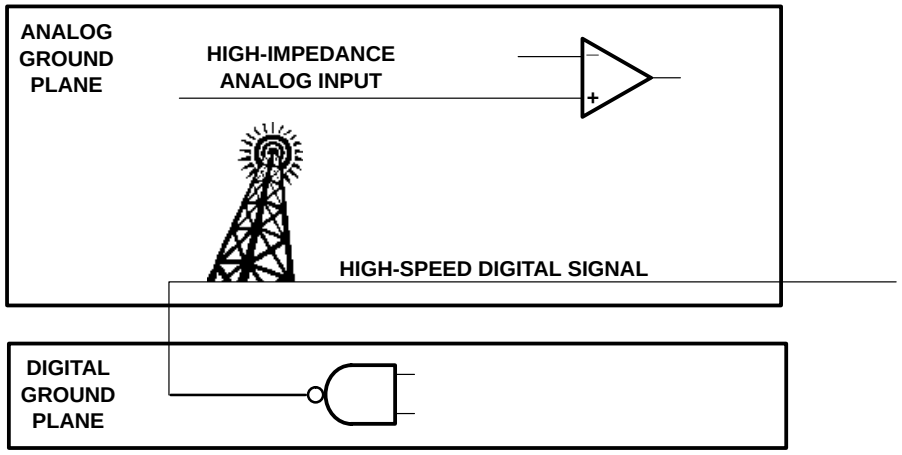


Figure 17–4. Broadcasting From PCB Traces

- It is a good idea to locate analog circuitry as close as possible to the I/O connections of the board. Digital designers, used to high current ICs, will be tempted to make a 50-mil trace run several inches to the analog circuitry thinking that reducing the resistance in the trace will help get rid of noise. What they have actually done is create a long, skinny capacitor that couples noise from digital ground and power planes into the op amp, making the problem worse!

17.3.3 A Good Example

Figure 17–5 shows one possible board layout. In this system, all electronics, including the power supply, reside on one PCB. Three separate and isolated ground/power planes are employed: one for power, one for digital, and one for analog. Power and ground connections from digital and analog sections of the board are combined only in the supply section, and are combined in close proximity. High-frequency conducted noise on the power lines is limited by inductors (chokes). In this case, the designer has even located low frequency analog circuitry close to low speed digital, keeping high frequency digital and analog physically apart on the board. This is a good, careful design that has a high likelihood of success — providing that good layout and decoupling rules are also followed.

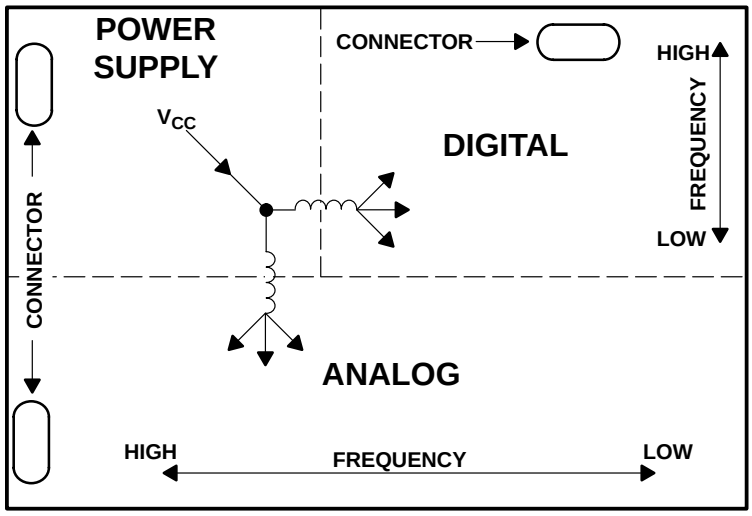


Figure 17–5. A Careful Board Layout

17.3.4 A Notable Exception

There is one case where it is necessary to combine analog and digital signals on the analog ground plane. Analog to digital (A/D) and digital to analog (D/A) converters are packaged as ICs with analog and digital ground pins coming out of the package. One might assume, based on the previous discussion, that the digital ground pin should be connected to digital ground and the analog ground pin to analog ground. That, however, is not correct.

The pin names *analog ground* and *digital ground* refer to internal connections in the IC, not the plane to which they should be connected. Both should connect to the analog ground plane. The connection would have been made inside the IC, but it is impossible to get low enough impedance at the typical geometries inside ICs. The IC designer actually counts on the end user to supply a low impedance connection outside the IC. Otherwise, the performance of the converter will be worse than specified.

One might suspect that the digital portions of the converter would make circuit performance worse by coupling digital switching noise onto the analog ground and power plane. Converter designers realize this, and design digital portions without a lot of output power to minimize switching transients. If the converter does not drive large fanouts, this should not be a problem. Be sure to properly decouple the logic supply for the converter to analog ground (see the following paragraphs).

17.4 The Frequency Characteristics of Passive Components

Choosing the right passive components for an analog design is important. In most cases, a *right* passive component will fit on the same pads as a *wrong* passive component, but not always. Start the design process by carefully considering the high frequency characteristics of passive components, and putting the correct part outline on the board from the start.

Most designers are totally ignorant of the frequency limitations of the passive components they use in analog circuitry. Passive components have limited frequency ranges, and operation of the part outside of that range can have some very unexpected results. One might think that this discussion only applies to high-speed analog circuits. But high frequencies that are radiated or conducted into a low-speed circuit will affect passive components as well. For example: a simple op amp low-pass filter may well turn into a high-pass filter at RF frequencies.

17.4.1 Resistors

High-frequency performance of resistors is approximated by the schematic shown in Figure 17–6.

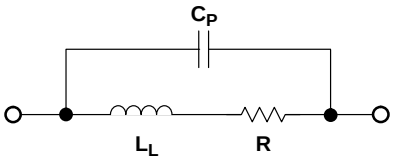


Figure 17–6. Resistor High-Frequency Performance

Resistors are typically one of three types: wire-wound, carbon composition, and film. It does not take a lot of imagination to understand how wire-wound resistors can become inductive because they are coils of resistive wire. Most designers are not aware of the internal construction of film resistors, which are also coils of thin metallic film. Therefore, film resistors are also inductive at high frequencies. The inductance of film resistors is lower, however, and values under 2 kΩ are usually suitable for high frequency work.

The end caps of resistors are parallel, and there will be an associated capacitance. Usually, the resistance will make the parasitic capacitor so “leaky” that the capacitance does not matter. For very high resistances, the capacitance will appear in parallel with the resistance, lowering its impedance at high frequencies.

17.4.2 Capacitors

High-frequency performance of capacitors is approximated by the schematic shown in FIgure 17–7.

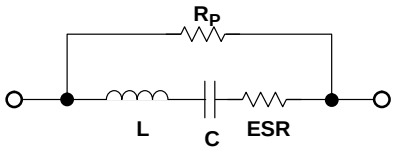


Figure 17–7. Capacitor High-Frequency Performance

Capacitors are used in analog circuitry for power supply decoupling and as filter components. For an ideal capacitor, reactance decreases by the formula:

$$X_C = \frac{1}{(2\pi fC)} \qquad (17–1)$$

Where:

X_C	capacitive reactance in Ohms
F	frequency in Hz
C	capacitance in microfarads

Therefore, a 10-µF electrolytic capacitor has a reactance of 1.6 Ω at 10 kHz, and 160-µΩ at 100 MHz. Right?

In reality, one will never see the 160 µΩ with the electrolytic capacitor. Film and electrolytic capacitors have layers of material wound around each other, which creates a parasitic inductance. Self-inductance effects of ceramic capacitors are much smaller, giving them a higher operating frequency. There is also some leakage current from plate to plate, which appears as a resistance in parallel with the capacitor, as well as resistance within the plates themselves, which add a parasitic series resistance. The electrolyte itself in electrolytic capacitors is not perfectly conductive (to reduce leakage current). These resistances combine to create the equivalent series resistance (ESR). The capacitors used for decoupling should be low ESR types, as any series resistance limits the effectiveness of the capacitor for ripple and noise rejection. Elevated temperatures also severely increase ESR, and can be permanently destructive to capacitors. Therefore, if an aluminum electrolytic will be subjected to high temperatures, use the high-temperature grade (105°C), not the low temperature grade (85°C).

For leaded parts, the leads themselves also add a parasitic inductance. For small values of capacitance, it is important to keep the lead lengths short. The combination of parasitic inductance and capacitance can produce resonant circuits! Assuming a lead self-inductance of 8 nH/cm (see the following paragraphs), a 0.01-µF capacitor with two 1-cm leads will resonate at 12.5 MHz. This effect was well known to engineers many decades ago,

who designed vacuum tube-based products with leaded components. Woe be to any hobbyist restoring antique radios that is unaware of this effect!

If electrolytic capacitors are used in a design, make sure that the polarity is correctly observed. The positive terminal of the capacitor must be connected to the more positive of two dc potentials. If there is any doubt whatsoever which polarity is correct, design calculations must continue until it is known, or a prototype must be built. Incorrect polarity of electrolytic capacitors will cause them to conduct dc current, in most cases destroying the part — and probably the rest of the circuit as well. If there is a rare case in which there will be both polarities present, use a nonpolarized electrolytic (which is constructed by connecting two polarized electrolytic capacitors in series). Of course, one can always connect two capacitors in series on the PCB, keeping in mind that the effective capacitance will be cut in half for equal values of capacitor.

17.4.3 Inductors

High frequency performance of inductors is approximated by the schematic shown in Figure 17–8.

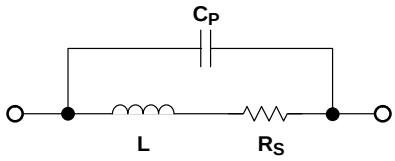


Figure 17–8. Inductor High-Frequency Performance

Inductive reactance is described by the formula:

$$X_L = 2\pi fL \qquad (17–2)$$

Where:

X_L	inductive reactance in Ohms
F	frequency in Hz
L	inductance in Henrys

Therefore, a 10-mH inductor has a reactance of 628 Ω at 10 kHz, which increases to 6.28 MΩ at 100 MHz. Right?

In reality, one will never see the 6.28 MΩ with this inductor. Parasitic resistances are easy to understand — the inductor is constructed of wire, which has a given resistance per unit length. Parasitic capacitance is harder to visualize, unless one considers the fact that each turn of wire in the inductor is located next to adjacent turns, forming a capacitor. This parasitic capacitance limits the upper frequency of this inductor to under 1 MHz. Even small wire-wound inductors start to become ineffective in the 10 MHz to 100 MHz range.

17.4.4 Unexpected PCB Passive Components

In addition to the obvious passive components above, the PCB itself has characteristics that form components every bit as real as those discussed previously — just not as obvious.

17.4.4.1 PCB Trace Characteristics

The layout pattern on a PCB can make it susceptible to radiated noise. A good layout is one that minimizes the susceptibility of analog circuitry to as many radiated noise sources as possible. Unfortunately, there is always a level of RF energy that will be able to upset the normal operation of the circuit. If good design techniques are followed, that level will be one that the circuit never encounters in normal operation.

17.4.4.1.1 *Trace Antennas*

A board is susceptible because the pattern of traces and component leads form antennas. Antenna theory is a complex subject, well beyond the scope of this book. Nevertheless, a few basics are presented here.

One basic type of antenna is the whip, or straight conductor. This antenna works because a straight conductor has parasitic inductance, and therefore can concentrate flux from external sources. The impedance of any straight conductor has a resistive and an inductive component:

$$Z = R + j\omega L \tag{17–3}$$

For dc and low frequencies, resistance is the major factor. As the frequency increases, however, the inductance becomes more important. Somewhere in the range of 1 kHz to 10 kHz, the inductive reactance exceeds the resistance, so the conductor is no longer a low resistance connection, but rather an inductor.

The formula for the inductance of a PCB trace is:

$$L(\mu H) = 0.0002X \cdot \left[\ln\left(\frac{2X}{W + H}\right) + 0.2235\left(\frac{W + H}{X}\right) + 0.5 \right] \tag{17–4}$$

Where:

X	length of the trace
W	width of the trace
H	thickness of the trace

The inductance is relatively unaffected by the diameter, since it varies as the logarithm of the circumference. Common wires and PCB traces vary between 6 nH and 12 nH per centimeter.

For example, a 10 cm PCB trace has a resistance of 57 mΩ and an inductance of 8 nH per cm. At 100 kHz, the inductive reactance reaches 50 mΩ. At frequencies above 100 kHz, the trace is inductive — not resistive.

A rule of thumb for whip antennas is that they begin to couple significant energy at about 1/20 of the wavelength of the received signal, peaking at ¼ the wavelength. Therefore, the 10-cm conductor of the previous paragraph will begin to be a fairly good antenna at frequencies above 150 MHz. Remember that although the clock generator on a digital PCB may not be operating at a frequency as high as 150 MHz, it approximates a square wave. Square waves will have harmonics throughout the frequency range where PCB conductors become efficient antennas. If through-hole components are mounted in a way that leaves significant lead length, those component leads also become antennas, particularly if they are bent.

Another major type of antenna is the loop. The inductance of a straight conductor is dramatically increased by bending it into partial or complete loops. Increased inductance lowers the frequency at which the conductor couples radiated signals into the circuit.

Without realizing it, most digital designers are well versed in loop antenna theory. They know not to make loops in critical signal pathways. Some designers, however, that would never think of making a loop with a high speed clock or reset signal will turn right around and create a loop by the technique they use for layout of the analog section of the board. Loop antennas constructed as loops of wire are easy to visualize. What is not as obvious is that slot antennas are just as efficient. Consider the three cases shown in Figure 17–9.

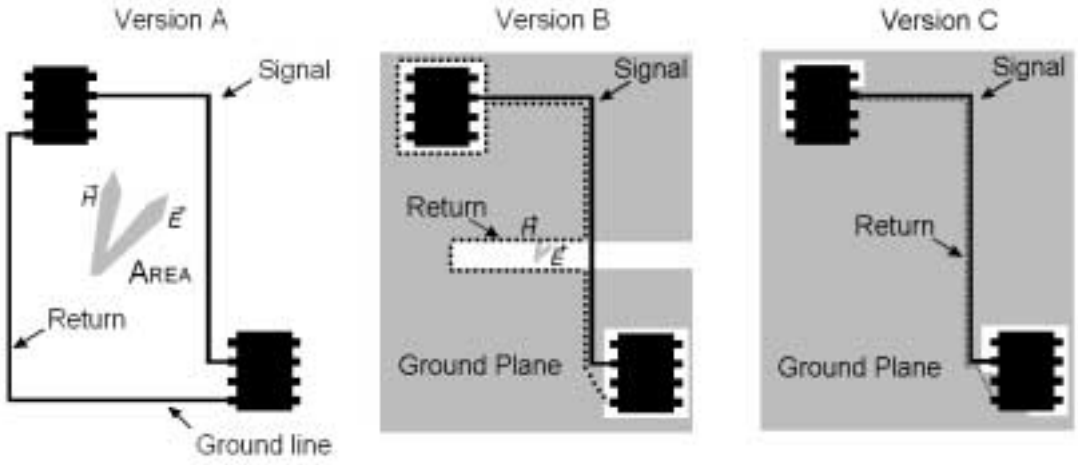


Figure 17–9. Loop and Slot Antenna Board Trace Layouts

Version **A** is a poor design. It does not use an analog ground plane at all. A loop is formed by the ground and signal traces. An electric field *E* and perpendicular magnetic field *H* are created, and form the basis of a loop antenna. A rule of thumb for loop antennas is that the length of each leg is equal to half the most efficiently received wavelength. Remember, however, that even at 1/20 of the wavelength, the loop will still be a fairly efficient antenna.

Version **B** is a better design, but there is intrusion into the ground plane, presumably to make room for a signal trace. A much smaller slot antenna is formed by the difference in pathways between signal and return. A second loop is created by the cutout for the IC.

Version **C** is the best design. Signal and return are coincident with each other, eliminating loop antenna effects completely. Note that there is still a cutout for the IC, but it is located away from the return path for the signal.

17.4.4.1.2 *Trace Reflections*

Reflections and matching are closely related to loop antenna theory, but different enough to warrant their own discussion.

When a PCB trace turns a corner at a 90° angle, a reflection can occur. This is primarily due to the change of width of the trace. At the apex of the turn, the trace width is increased to 1.414 times its width. This upsets the transmission line characteristics, especially the distributed capacitance and self–inductance of the trace — resulting in the reflection. It is a given that not all PCB traces can be straight, and so they will have to turn corners. Most CAD systems give some rounding effect on the trace. Sharp 90° corners in traces are a relic of the *tape up* days of PCB layout. The rounding effects of CAD programs, however, still do not necessarily maintain constant width as the trace rounds the corner. Figure 17–10 shows progressively better techniques of rounding corners. Only the last example maintains constant trace width and minimizes reflections. Most CAD programs now support these methods, but they can entail a little more work to master.

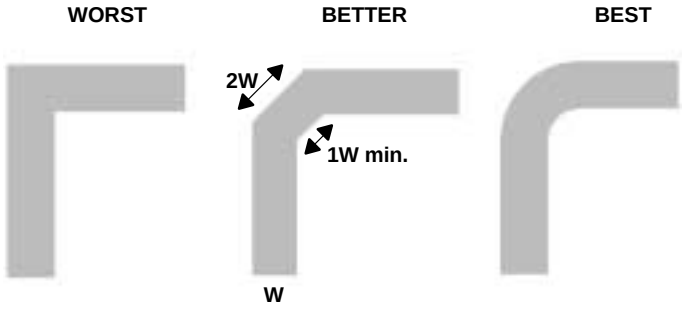
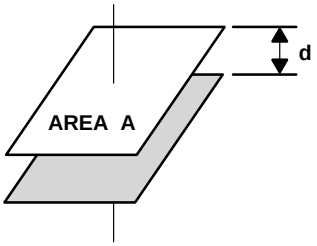


Figure 17–10. PCB Trace Corners

A suggestion for the advanced PCB layout engineer: leave rounding to the last step before tear-dropping and flood-filling. Otherwise, the CAD program will slow down doing numerical calculations as the traces are moved around during routing.

17.4.4.2 Trace-to-Plane Capacitors

PCB traces, being composed of foil, form capacitance with other traces that they cross on other layers. For two traces crossing each other on adjacent planes, this is seldom a problem. Coincident traces (those that occupy the same routing on different layers), form a long, skinny capacitor. The formula for capacitance is shown in Figure 17–11.



$$C = 0.0085 \times \epsilon_R \times \frac{A}{d}$$

where:

C = capacitance (pF)

ϵ_R = dielectric constant

A = area of plate (mm^2)

d = separation of plates (mm)

Figure 17–11. PCB Trace-to-Plane Capacitance Formula

For example, if the capacitance formula is applied to the following trace:

- 4 Layer board — signal routing next to ground plane
- Board layer thickness: 0.188 mm
- Trace Width: 0.75 mm
- Trace Length: 7.5 mm

A typical value for E_R of FR–4 PCB material is 4.5. Due to the variations of material from which an FR–4 board can be fabricated, this value is not guaranteed, but should be in the range of 4 to 5.

The capacitance between these traces would be 1.1 pF. Of course, the antenna effect on a 7.5-mm trace would be devastating, so this example is a bit extreme. Ignoring the antenna effects for now, there are cases in which even a very small parasitic capacitance like 1 pF is unacceptable. Figure 17–12 dramatically illustrates the effect of 1 pF capacitance occurring at the inverting input of the op amp. It causes a doubling of the output amplitude near the bandwidth limit of the op amp. This is an invitation to oscillation, especially since the trace is an efficient antenna above 180 MHz.

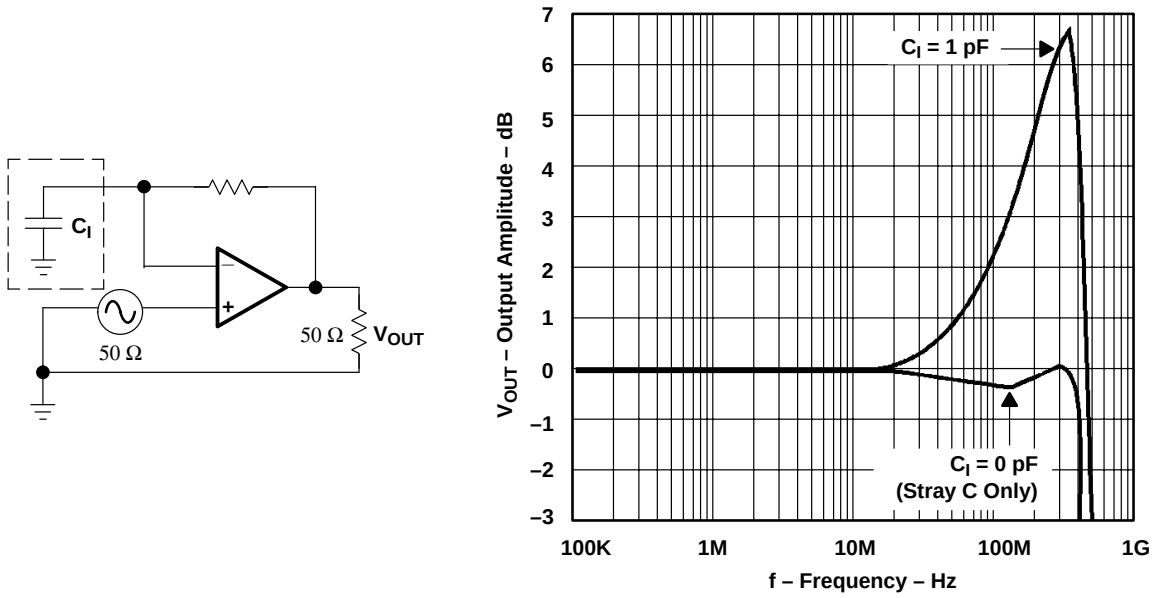


Figure 17–12. Effect of 1-pF Capacitance on Op Amp Inverting Input

There are numerous fixes to the problem above. The most obvious would be to shorten the length of the traces. Another not-so-obvious fix would be to use a different trace width. There is no reason why an inverting op amp trace has to be 0.75 mm wide — it carries almost no current. If the trace length is reduced to 2.5 mm (one third as long) and the trace width is changed to 0.188 mm (1/4 as wide), the capacitance becomes 0.1 pF — much less significant in the example above. Another fix is to remove the ground plane under the inverting input and the traces leading to it.

The inverting input of op amps, particularly high-speed op amps, is especially prone to oscillation in high gain circuits. This is due to unwanted capacitance on the input stage. It is important to minimize capacitance on this input by reducing trace width and placing components as close as possible to this input. If this input still oscillates, it may be necessary to scale the input and feedback resistors lower by a decade or two to change the resonance of the circuit. Scaling the resistors up will seldom help, as the problem is also related to the impedance of the circuit. If filter components are involved, they will also have to be scaled to avoid changing the filter characteristics of the circuit. The power consumption of the circuit will also increase if resistors are lowered.

17.4.4.3 Trace-to-Trace Capacitors and Inductors

PCB traces are not infinitely thin. They have some finite thickness, as defined by the *ounce* parameter of the copper clad foil. The higher the number of ounces, the thicker the copper. If two traces run side-by-side, then there will be capacitive and inductive coupling

between them (Figure 17–13). The formulas for these parasitic effects can be found in transmission line and/or microstrip references, but are too complex for inclusion here.

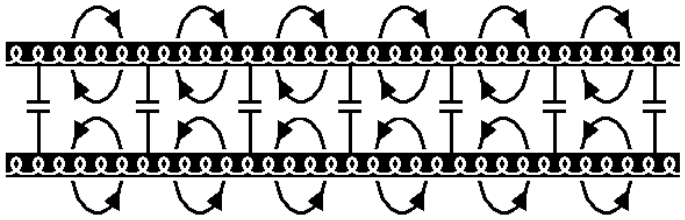


Figure 17–13. Coupling Between Parallel Signal Traces

Signal lines should not be routed parallel to each other, unless transmission line or microstrip effects are desired. Otherwise, a gap of at least three times the signal trace width should be maintained.

Capacitance between traces in an analog design can become a problem if fixed resistors in the design are large (several MΩ). Capacitance between the inverting and noninverting inputs of an op amp could easily cause oscillation.

17.4.4.4 Inductive Vias

Whenever routing constraints force a *via* (connection between layers of a PCB, FIgure 17–14), a parasitic inductor is also formed. At a given diameter (d) the approximate inductance (L) of an via at a height of (h) may be calculated as follows:

$$L = \approx \frac{h}{5} \times \left(1 + \ln\left(\frac{4h}{d}\right)\right) nH \qquad (17–5)$$

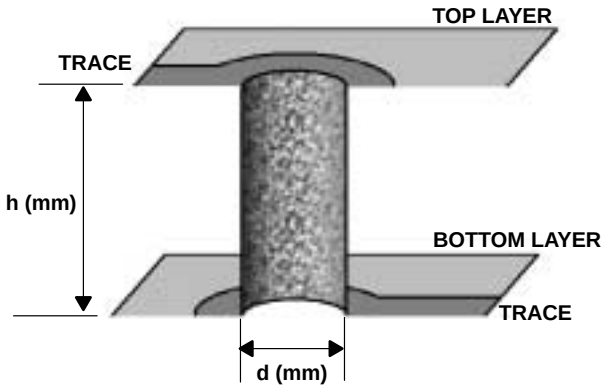


Figure 17–14. Via Inductance Measurements

For example, a 0.4-mm diameter via through a 1.5-mm thick PCB has an inductance of 1.1 nH.

Keep in mind that inductive vias combined with parasitic capacitance can form resonant circuits. The self-inductance of a via is small enough that these resonances are in the GHz range, but inductors add in series, lowering the resonant frequency. Do not put several vias on a critical trace of a high-speed analog circuit! Another concern is that the vias put holes in ground planes, potentially creating ground loops. They should be avoided — the best analog layout is one that routes all signal traces on the top layer of the PCB.

17.4.4.5 Flux Residue Resistance

Yes, even an unclean board can affect analog circuit performance.

Be aware that if the circuit has very high resistances — even in the low MΩ — special attention may need to be paid to cleaning. A finished assembly may be adversely affected by flux or cleansing residue. The electronics industry in the past few years has joined the rest of the world in becoming environmentally responsible. Hazardous chemicals are being removed from the manufacturing process — including flux that has to be cleaned with organic solvents. Water-soluble fluxes are becoming more common, but water itself can become contaminated easily with impurities. These impurities will lower the insulation characteristics of the PCB substrate. It is vitally important to clean with freshly distilled water every time a high-impedance circuit is cleaned. There are applications that may call for the older organic fluxes and solvents, such as very low power battery powered equipment with resistors in the 10s of MΩ range. Nothing can beat a good vapor defluxing machine for ensuring that the board is clean.

17.5 Decoupling

Noise, of course, can propagate into analog circuitry through the power pins of the circuit as a whole and op amp itself. Bypass capacitors are used to reduce the coupled noise by providing low impedance power sources local to the analog circuitry.

17.5.1 Digital Circuitry — A Major Problem for Analog Circuitry

If analog circuitry is located on the same board with digital circuitry, it is important to understand a little about the electrical characteristics of digital gates.

A typical digital output consists of two transistors connected in series between power and ground (Figure 17–15). One transistor is turned on and the other turned off to produce logic high and vice versa for logic low. Because one transistor is turned off for either logic state, the power consumption for either logic state is low, while the gate is static at that level.

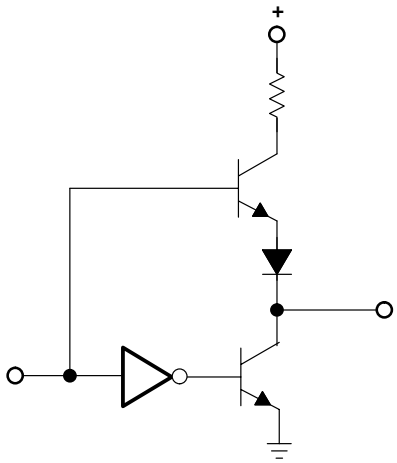


Figure 17–15. Logic Gate Output Structure

The situation changes dramatically whenever the output switches from one logic state to the other. There may be a brief period of time when both transistors conduct simultaneously. During this period of time, current drawn from the power supply increases dramatically — since there is now a low impedance path through the two transistors from power to ground. Power consumption rises dramatically and then falls, creating a droop on the power supply voltage, and a corresponding current spike. The current spike will radiate radio frequency (RF) energy. There may be dozens, or even hundreds of such outputs on a digital IC, so the aggregate effect may be quite dramatic.

It is impossible to predict the frequencies of these spikes, because the frequencies are affected by the propagation delays of the transistors in the gate. Propagation delay is affected by random factors that occur during manufacture. Digital switching noise will be broadband, with harmonics throughout the spectrum. A general rejection technique is required, rather than one that rejects a specific frequency.

17.5.2 Choosing the Right Capacitor

Table 17–2 is a rough guideline describing the maximum useful frequencies of common capacitor types.

Table 17–2. Recommended Maximum Frequencies for Capacitors

TYPE	MAX FREQUENCY
Aluminum Electrolytic	100 kHz
Tantalum Electrolytic	1 MHz
Mica	500 MHz
Ceramic	1 GHz

Obviously from the table, tantalum electrolytic capacitors are useless for frequencies above 1 MHz. Effective high frequency decoupling at higher frequencies demands a ceramic capacitor. Self-resonances of the capacitor must be known and avoided or the capacitor may not help, or even make the problem worse. Figure 17–16 illustrates the typical self-resonance of two capacitors commonly used for bypassing — 10 μF tantalum electrolytic and 0.01 μF ceramic.

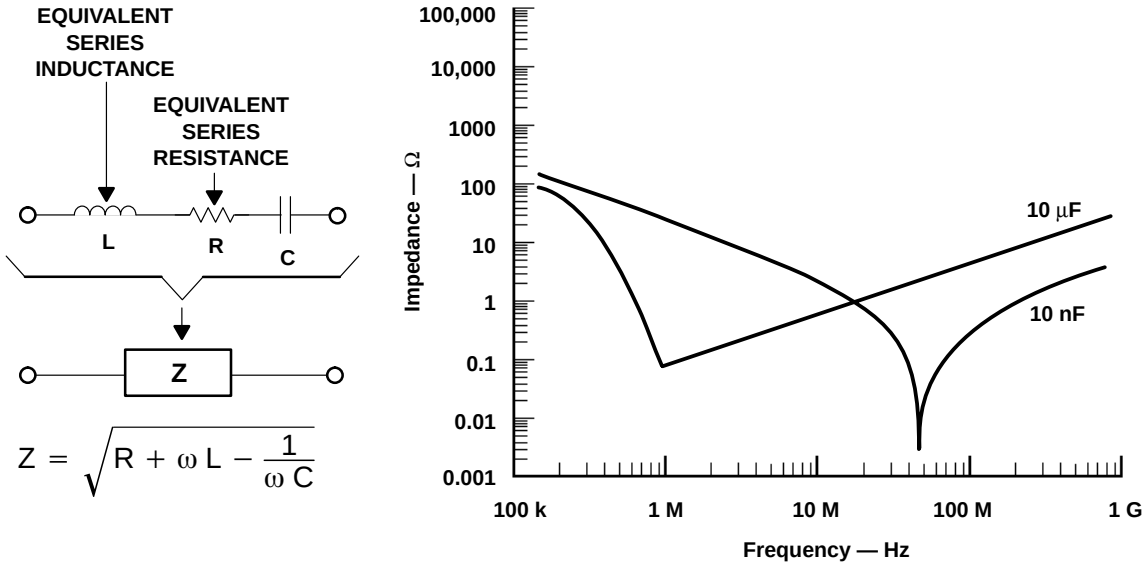


Figure 17–16. Capacitor Self Resonance

Consider these resonances to be typical values, the characteristics of actual capacitors can vary from manufacturer to manufacturer and grade of part to grade of part. The important thing is to make sure that the self-resonance of the capacitor occurs at a frequency above the range of the noise that must be rejected. Otherwise, the capacitor will enter a region where it is inductive.

Do not assume that a single 0.1-μF capacitor will decouple all frequencies. Smaller capacitors may work better at higher frequencies than larger ones. When poor decoupling at higher frequencies is suspected, try a smaller capacitor rather than a larger one.

17.5.3 Decoupling at the IC Level

The method most often used to decouple the high-frequency noise is to include a capacitor, or multiple capacitors connected from the op amp power pin to the op amp ground pin. It is important to keep the traces on this decoupling capacitor short. If not, the traces on the PCB will have significant self-inductance, defeating the purpose of the capacitor.

A decoupling capacitor must be included on every op amp package — whether it contains 1, 2 or 4 devices per package. The value of capacitor must be picked carefully to reject the type of noise present in the circuit.

In particularly troublesome cases, it may be necessary to add a series inductor into the power supply line connecting to the op amp. This inductor is in addition to the decoupling capacitors, which are the first line of defense. The inductor should be located before, not after the capacitors.

Another technique that is lower in cost is to replace the series inductor with a small resistor in the 10-Ω to 100-Ω range. The resistor forms a low-pass filter with the decoupling capacitors. There is a penalty to pay for this technique — depending on the power consumption of the op amp, it will reduce the rail-to-rail voltage range. The resistor forms a voltage divider with the op amp as a resistive active component in the lower leg of the divider. Depending on the application, this may or may not be acceptable.

17.5.4 Decoupling at the Board Level

There is usually enough low frequency ripple on the power supply at the board input to warrant a bulk decoupling capacitor at the power input. This capacitor is used primarily to reject low frequency signals, so an aluminum or tantalum capacitor is acceptable. An additional ceramic cap at the power input will decouple any stray high frequency switching noise that may be coupled off of the other boards.

17.6 Input and Output Isolation

Many noise problems are the result of noise being conducted into the circuit through its input and output pins. Due to the high-frequency limitations of passive components, the response of the circuit to high frequency noise may be quite unpredictable.

In situations in which conducted noise is substantially different in frequency from the normal operating range of the circuit, the solution may be as simple as a passive RC low-pass filter that rejects RF frequencies while having negligible effect at audio frequencies. A good example is RF noise being conducted into an audio op amp circuit. Be careful, though. A low pass filter loses its characteristics at 100 to 1000 times f_{3db}. More stages may be required to cover different frequency ranges. If this is the case, make the highest frequency filter the one nearest to the source of noise. Inductors or ferrite beads can also be used in a noise rejection filter network to eliminate conducted noise. Ferrite beads are inductive up to a specified frequency, and then become resistive.

The effect of radiated energy coupling into an analog circuit can be so bad that the only solution to the problem may be to completely shield the circuit from radiated energy. This shield is called a *Faraday Cage*, and must be carefully designed so that frequencies that

are causing the problem are not allowed to enter the circuit. This means that the shield must have no holes or slots larger that 1/20 the wavelength of the offending frequency. This can be a demanding requirement. It is a good idea to design a PCB from the beginning to have enough room to add a metal shield if it becomes necessary. If a shield is used, frequently the problem will be severe enough that ferrite beads will also be required on all connections to the circuit.

17.7 Packages

Op amps are commonly supplied one, two, or four per package. Single op amps often contain additional inputs for features such as offset nulling. Op amps supplied two and four per package only offer inverting and noninverting inputs, and the output. If the additional features are important, the only package choice is single. Be aware, though, that the offset-nulling pins on a single op amp package can act as secondary inputs, and must be treated carefully. Consult the data sheet on the particular device being used.

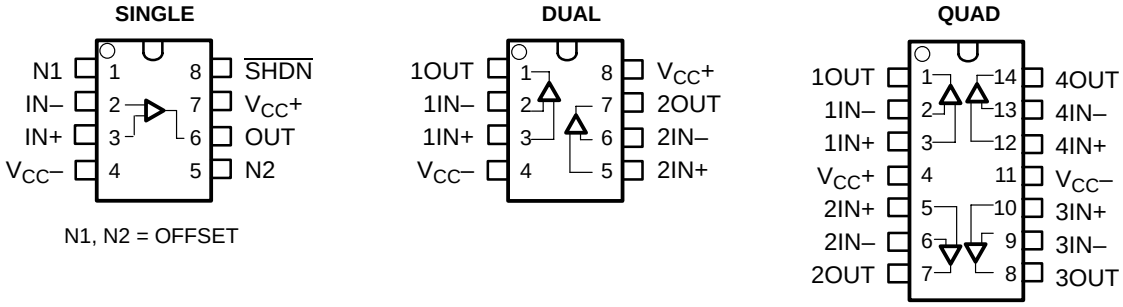


Figure 17–17. Common Op Amp Pinouts

The single op amp package places the output on the opposite side from the inputs. This can be a disadvantage at high speeds, because it forces longer PCB traces. One way to overcome this difficulty is to use a surface-mount package, and mount the feedback components on the rear of the board. This, however, adds at least two vias to the design, and places a hole in the ground plane for a 2-layer design. It may make more sense to use a dual op amp package, even if the second op amp is unused (remember to terminate it properly — see paragraph 17.7.3). Figure 17–18 is an illustration of trace length for an inverting op amp stage.

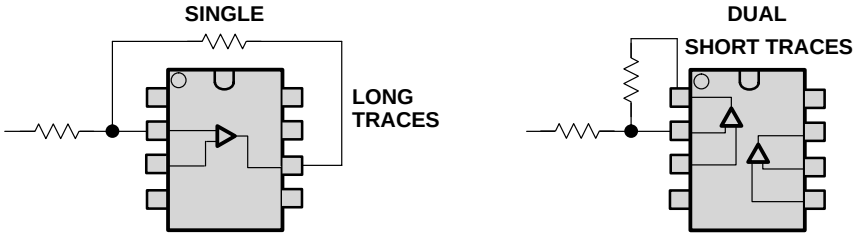


Figure 17–18. Trace Length for an Inverting Op Amp Stage

It is popular to use dual op amps for stereophonic circuits, and quad op amps for filter stages with many sections. There is a penalty for doing so, however. Although modern processing techniques provide high levels of isolation between amplifiers on the same piece of silicon, there will be some crosstalk. If isolation between amplifiers is important, then single packages should be considered. Crosstalk problems are not limited to the IC – the dual and quad packages place a high density of passive components in close proximity to each other. This proximity will lead to some crosstalk.

Dual and quad op amp packages offer some additional benefits beyond density. The amplifier stages tend to be mirror images of each other. If similar stages are to be laid out on the PCB, the layout only needs to be done once, then it can be mirror-imaged to form the other stage. Figure 17–19 illustrates this effect for four inverting op amp stages implemented in a quad package:

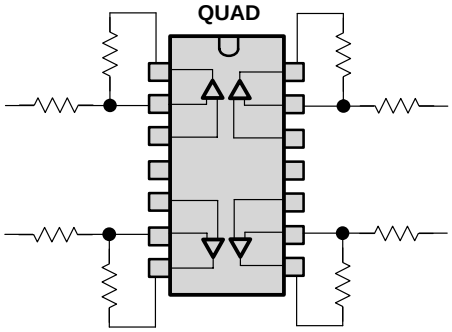


Figure 17–19. Mirror-Image Layout for Quad Op Amp Package

These illustrations, however, do not show all connections required for operation, in particular, the half-supply generator for single-supply operation. Modifying the diagram of Figure 17–19 to use the fourth op amp as a half supply generator is shown in Figure 17–20.

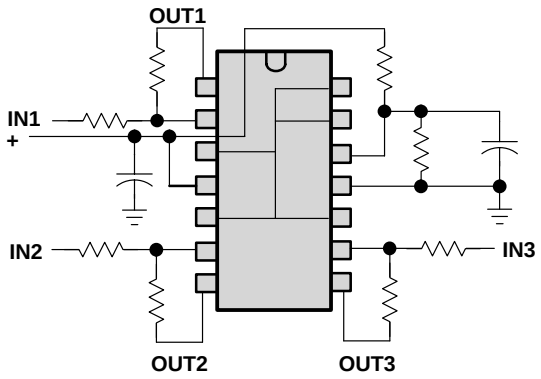


Figure 17–20. Quad Op Amp Package Layout with Half-Supply Generator

This example shows all of the connections actually required to produce three independent inverting stages. Note that the half-supply connection can be made entirely under the IC, keeping the trace length short. This example is not meant to be used as a suggested PCB layout, it merely illustrates what can be done. The half-supply op amp, for example, could be any one of the four. The passive components can be selected so that they span the lead pitch. Surface-mount 0402 packages, for example, span a width equal to or less than a standard SO package. This can keep traces lengths extremely short for high frequency applications.

Package styles for op amps include the dual-in-line (DIP) and small-outline (SO). Lead pitches for op amps have been continually decreasing, as has been the case for all ICs in general. Decreasing lead pitches have been accompanied by a decrease in the size of passive components as well. Decreasing the overall circuit dimensions reduces parasitic inductance, which should allow higher frequency circuits, but it also increases the potential for crosstalk by placing conductors closer to each other where capacitive effects increase.

17.7.1 Through-Hole Considerations

The older technology for op amps and other components is through-hole. Components are constructed with leads that insert through holes in the board — hence the name.

Through-hole components, due to their size, are more suited to applications where space is not an issue. The components themselves are frequently lower in cost, but the PCB is more expensive due to the fact that the PCB fabrication house has to drill holes for component leads. PCBs are primarily a mechanical fabrication — the number of holes and number of different drills have a big impact on the price.

The leads on a through-hole op amp are arranged on a 0.1-inch grid. Many PCB layout people like to maintain the 0.1-inch grid for the rest of the components as well. Resistors

and other passive components can even be purchased with leads pre-bent to land on a 0.1 inch grid. Some electrolytic capacitors have leads that are on a 0.025-inch grid.

These component sizes may force a lot of wasted area on the PCB. Components that ideally should be placed close to the op amp itself may be forced several tenths of an inch away, due to intervening components. Therefore, through hole circuitry is not recommended for high speed analog circuitry, or for analog circuitry in proximity to high speed digital.

Some designers attempt to overcome the long trace length caused by resistors by placing the resistors on the board vertically, one lead of the resistor bent close to the body of the part. This is common in older consumer electronics. This allows for denser placement of parts, and may help some with trace length — but each resistor exposes almost 1 cm of one component lead to radiated signals, and lead self-inductance.

An advantage of the through-hole approach of PCB layout is that the through-holes themselves can serve as feedthroughs, reducing the number of vias in complex circuits.

17.7.2 Surface Mount

Surface-mount circuitry does not require a hole for each component lead. Automated testing, however, may require vias on every node. The holes were never an issue with through-hole circuitry, because every component lead made a hole in the board. The PCB layout designer, who is used to designing a board with a minimum number of vias, now has to put a via on EVERY node of the circuit. This can make a Swiss cheese out of a nice continuous ground plane — negating many of the advantages it provides.

Fortunately, there is a close variation of the “via on every node” requirement. This requirement can often be met by putting a test pad on every node. The automated test station can then access the analog circuitry from the top of the board. A *clamshell* test fixture is significantly more expensive than one that accesses only one side of the board. The extra cost can be justified if there is documentation that circuit performance will be unacceptable with vias.

Signal connections to ground or the power supply may have to be made through a small fixed resistor instead, so the automated equipment can access that pin of the IC and test its function.

17.7.3 Unused Sections

In many op amp designs, one or more op amps may be unused. If this is the case, the unused section must be terminated properly. Improper termination can result in greater power consumption, more heat, and more noise in op amps on the same physical IC. If the unused section of the op amp is connected as shown in the **better** side of Figure 17–21, it will be easier to use it for design changes.

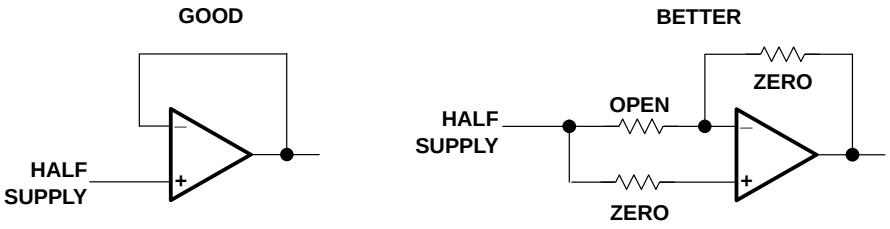


Figure 17–21. Proper Termination of Unused Op Amp Sections

17.8 Summary

Keep the following points in mind when designing a PCB for analog circuitry:

17.8.1 General

- Think of the PCB as a component of the design.
- Know and understand the types of noise the circuit will be subjected to.
- Prototype the circuit.

17.8.2 Board Structure

- Use a high-quality board material such as FR–4.
- Multi-layer boards are as much as 20-dB better than double-sided boards.
- Use separate, nonoverlapping ground and power planes.
- Place power and ground planes to the interior of the board instead of exterior.

17.8.3 Components

- Be aware of frequency limitations of traces and other passive components.
- Avoid vertical mounting of through-hole passive components in high-speed circuits.
- Use surface mount for high-speed analog circuitry.
- Keep traces as short as possible.
- Use narrow traces if long traces are required.
- Terminate unused op amp sections properly.

17.8.4 Routing

- Keep analog circuitry close to the power connector of the board.
- Never route digital traces through analog sections of the board, or vice versa.
- Make sure that traces to the inverting input of the op amp are short.
- Make sure that traces to the inverting and noninverting inputs of the op amp do not parallel each other for any significant length.
- It is better to avoid vias, but the self-inductance of vias is small enough that a few should cause few problems.
- Do not use right angle traces, use curves if at all possible.

17.8.5 Bypass

- Use the correct type of capacitor to reject the conducted frequency range.
- Use tantalum capacitors at power input connectors for filtering power supply ripple.
- Use ceramic capacitors at power input connectors for high frequency conducted noise.
- Use ceramic capacitors at the power connections of every op amp IC package. More than one capacitor may be necessary to cover different frequency ranges.
- Change the capacitor to a smaller value — not larger — if oscillation occurs.
- Add series inductors or resistors for stubborn cases.
- Bypass analog power only to analog return, never to digital return.

17.9 References

1. *Analog Dialog 30th Anniversary Issue*, Analog Devices, Ask the Applications Engineer –10, James Bryant

2. *High-Speed Operational Amplifier Layout Made Easy*, Texas Instruments, Andy Hahn, SLOA046

Chapter 18

Designing Low-Voltage Op Amp Circuits

Ron Mancini

18.1 Introduction

In one respect, voltage is like water: you don't appreciate its value until your supply runs low. Low-voltage systems, defined here as a single power supply less than 5 V, teach us to appreciate voltage. We aren't the first electronic types to learn how valuable voltage is; about 15 years ago the audio console design engineers appreciated the relationship between voltage and dynamic range. They needed more dynamic range to satisfy their customers; thus, they ran op amps at the full rated voltage, not the recommended operating voltage, so they could squeeze a few more dB of dynamic range from the op amp. These engineers were willing to take a considerable risk running op amps at the full rated voltage; but their customers demanded more dynamic range. The moral of this story is that dynamic range is an important parameter, and supply voltage is tied directly to dynamic range.

Knowing how to obtain and use the maximum dynamic range and input/output voltage range is critical to achieving success in low voltage design. We will investigate these subjects in detail later, but for now it is useful to review the history of op amps. Knowing how op amps evolved into the today's marvels is interesting, and it gives designers an insight into system problems that they encounter as they design in the low voltage world.

When power supplies were ±15 V, the output voltage swing of an op amp didn't seem important. When the power supply was 30 V the typical circuit designer could afford to sacrifice 3 V from each end of the output voltage swing (this was because of transistor saturation or cutoff). The transistors in the op amp need enough voltage across them to operate correctly, so why worry about 6 V out of 30 V. Also, the input transistors required base bias, so an op amp with 30-V supplies often offered a common-mode input voltage range of 24 V or less. These numbers come from the μA741 data sheet; the μA741 (about 1969) is the first internally compensated op amp to achieve wide popularity.

A later generation op amp, the LM324, had better dynamic range characteristics than the μA741. The LM324's output voltage swing is 26 V when operated from a 30-V power supply, and the common-mode input voltage range is 28.5 V. The LM324 was big news because it was specified to operate with a 5-V power supply. The LM324's output voltage

swing at V_{CC} = 5 V is 3.48 V, and this presented problems for the early low voltage circuit designers because the output voltage swing was smaller than most analog-to-digital converter (ADC) input voltage ranges. You must fill the ADC input voltage range to obtain its full dynamic range. The LM324 had an input voltage common-mode range of (V_{CC}–1.5 V) to 0 V; at least this op amp could work with transducers connected to the lower power supply rail if the transducer did not have an ac output voltage swing.

The next incremental improvement in op amps was the LM10 because it operated on 1.1-V power supplies. It was introduced almost as an afterthought because there was no pressing demand for it. Its brilliant designer, Robert J. Widlar, wrote “IC op amps have reached a certain maturity in that there no longer seems to be a pressing demand for better performance.” There were no pressing demands for a low voltage op amp in 1978 because portable (portable means battery applications that are almost always single supply) did not become popular until the late 1980s or early 1990s.

Cell phones, calculators, and portable instruments — not new battery technology — opened the market for low-voltage op amps, and when the portable concept caught on, the demand for low-voltage op amps increased. The increasing demand did not breed new companies committed to low-voltage IC design; rather, the established IC manufacturers threw a few low voltage op amps into their portfolio. These op amps, like the LM324, could operate on a low voltage, but they were severely lacking in input common-mode voltage range and output voltage swing. Circuit designers had to be satisfied with this generation of op amps until something better came along. Well, something better is here now!

The next generation of low-voltage op amps has much better specifications. The TLV278X operates off a power supply ranging from 1.8 V to 3.6 V, and it has an output voltage swing of 1.63 V (when the power supply is 1.8 V) coupled with an input common-mode voltage range of –0.2 V to 2 V. The TLV240X operates off a power supply ranging from 2.5 V to 16 V, and it has an output voltage swing of 2.53 V when the power supply voltage is 2.7 V. Also, when it is operated off a 2.7-V power supply, it has an input common-mode voltage range of –0.1 V to 7.7 V. These new op amps are far superior to their predecessors when evaluated on the their merits, which are extended output voltage swing and input common voltage range.

The latest op amps make it possible to design more accurate and cost effective electronic equipment, but there is one problem that they don’t solve. Low-voltage applications are defined here as single-supply applications, and in single-supply design, the op amp input voltage and output voltage is referenced to the midpoint of the power supply (V_{CC}/2). Unfortunately, most transducers are not connected to the midpoint of the power supply because, in the majority of cases, this requires a third wire beyond V_{CC} and ground. It doesn’t help to create V_{CC}/2 at the transducer location (to save a wire) because it is not identical to the midpoint of the power supply (unresolved errors enter because of the reference voltage difference). When the transducer in a single supply design is referenced to any voltage other than the midpoint of the power supply, the reference voltage is amplified with the transducer voltage.

The trick to designing single supply op amp circuits is using external biasing to strip off or null out the reference voltage. Designing op amp circuits with biasing normally involves an iterative cut and try approach where the designer assumes a circuit configuration, solves equations, changes the configuration, and repeats the process until a solution is found. A technique that solves the problem the first time is presented later.

18.2 Dynamic Range

It is extremely hard to define dynamic range (DR) for an op amp, so lets start with a digital-to-analog converter (DAC) where DR is defined as the ratio of the maximum output voltage to the smallest output voltage the DAC can produce (least significant bit or LSB). Dynamic range is usually expressed in dB using the formula given in Equation 18–1.

$$DR = 20\ \mathrm{Log}_{10}\left(\frac{V_{OUT(MAX)}}{V_{OUT(MIN)}}\right) \tag{18–1}$$

The same definition of DR can be used for an op amp, and the maximum output voltage swing equals V_{OUTMAX}. This output voltage swing is defined as the maximum output voltage the op amp can achieve (V_{OH}) minus the minimum output voltage the op amp can achieve (V_{OL}). V_{OH} and V_{OL} are easily obtainable from an op amp IC data sheet. Normally, V_{OH} and V_{OL} are guaranteed minimum and maximum parameters respectively. This yields Equation 18–2.

$$V_{OUT(MAX)} = V_{OH(MIN)} - V_{OL(MAX)} \tag{18–2}$$

Equation 18–2 can be used to illustrate the role that power supply voltage plays in limiting the DR. $V_{OH(MIN)}$ is the most positive power supply voltage minus the voltage drop across the upper output transistor, thus $V_{OH(MIN)}$ is directly proportional to the most positive power supply voltage. For any op amp, the output voltage swing is directly proportional to the power supply voltage, thus, in the same op amp, the DR is directly proportional to the power supply voltage.

At first thought, one might think that the smallest output voltage that an op amp can have is zero, and the natural conclusion based on this assumption is that the DR is equal to infinity. This is never the case because op amp and external circuit imperfections ensure that the smallest op amp output voltage is greater than zero. It turns out that $V_{OUT(MIN)}$ is actually determined by a series of error terms. These error terms are the op amp's internal noise (V_n and I_n), external resistor noise (V_{nR}), power supply rejection ratio (k_{SVR}), voltage offset (V_{IO}), current offset (I_{IO}), common-mode rejection ratio (CMRR), and closed loop gain (G). Each of these error terms is referred to the input of the op amp, so they must be multiplied by the closed loop gain to be referred to the output (see Figure 18–1).

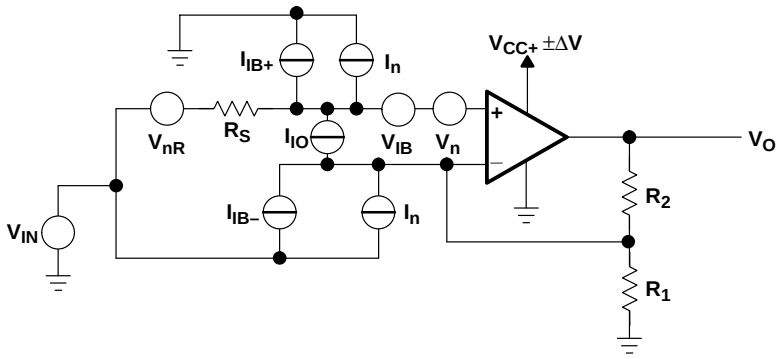


Figure 18–1. Op Amp Error Sources

The error sources are taken into account in Equation 18–3, and this equation refers them to the op amp output by multiplying them by the op amp's closed loop gain.

$$V_{OUT(MIN)} = \\ = G_{CL}\left(V_{IO} + I_{IO}R_S + \alpha V_{IO}\,\Delta T + \frac{V_{IN} + I_{IB}R_P}{CMRR} + \frac{\Delta V}{k_{SVR}} + V_n + I_nR_{EQ} + I_{nR}R\right) \quad (18–3)$$

The maximum DR that can be achieved by an op amp is given in Equation 18–4.

$$DR = 20\,Log_{10}\left(\frac{V_{OH(MIN)} - V_{OL(MAX)}}{V_{OUT(MIN)}}\right) = 20\,Log_{10}\frac{V_{OUT(MAX)}}{V_{OUT(MIN)}} \quad (18–4)$$

The DR is reduced by the sum of the error terms, so it is proper to conclude that the maximum power supply voltage and the op amp choice (this defines the error magnitude) both establish the DR of an op amp. The first two terms in Equation 18–3 are dc error terms, thus, they can be adjusted to zero by one of several methods not mentioned here. The input offset current and input bias current error terms were big factors with older generation ICs, but today's technology render them much less significant (see Table 18–1).

Table 18–1. Comparison of Op Amp Error Terms

PARAMETER	LM324	TLV278X	TLV240X	UNITS
V_{IO}	9	1.5	4.5	mV
αV_{IO}	No spec	8	3	μV
I_{IB}	500,000	100	350	pA
I_{IO}	50,000	100	300	pA
CMMR	50	50	65	dB
k_{SVR}	65	70	100	dB
Vn	No spec	18 (1 kHz)	500 (100 Hz)	$\frac{nV}{(Hz)^{1/2}}$
In	No spec	0.9 (1 kHz)	8 (100 Hz)	$\frac{fA}{(Hz)^{1/2}}$

The data in Table 18–1 indicates that older low voltage op amps are not capable of yielding the DR that that later technology op amps are.

18.3 Signal-to-Noise Ratio

Noise sets a limit on the information and signals that can be handled by a system. The ability of an amplifier, receiver, or other device to discern a signal is degraded by noise. Noise mixed with the incoming signal, noise generated by the op amp, resistor noise, and power supply noise ultimately determine the size of the signal that can be recovered and measured.

Noise fluctuates randomly over a period of time, so instantaneous signal or noise levels don't describe the situation adequately. Averages over a long period of time (root mean squared or RMS) are used to describe both the signal and the noise. Signal-to-noise ratio (SNR) was initially established as a measure of the quality of the signal that exists in the presence of noise. This SNR was a power ratio, and it was established at the output of a circuit. The SNR that we are interested in is a voltage ratio because the impedance is constant, and it is established at the input to the op amp. This means that all noise voltages, including resistor noise voltage, must be calculated in RMS volts at the op amp input. The SNR is given in Equation 18–5.

$$SNR = 20 \, Log_{10}\left(\frac{V_{SIGNAL}}{V_{NOISE}}\right) \tag{18–5}$$

The signal is established by a transducer; a device that senses a change in a variable and converts that change into a voltage change. Transducers also convert some of their physical surroundings into a noise voltage that is combined with the signal. Noise from the physical surroundings of the transducer, unless its nature is well known, is almost impos-

sible to separate from the transducer signal. When transducers are connected to the electronics, cabling picks up noise, and some transducers like thermocouples can pick up noise from the connecting junctions. Thus, the signal is never clean as it enters the electronics. The noise generated by the op amp was defined in the previous section as V_n, I_nR_{EQ}, $I_{nR}R$, and $\Delta V/k_{SVR}$, and this noise is added to the signal.

The transducer often has a very small output voltage swing, so when the transducer output voltage swing is converted to least significant bits (LSB) the noise voltage should be very small compared to an LSB. Consider a temperature transducer that has a 10-mV swing over its range. When the transducer output voltage swing is considered to be the full-scale voltage (FSV) of an ADC, the LSB is very small as is shown in Equation 18–6 for a 12-bit (N) ADC.

$$\text{LSB} = \frac{\text{FSV}}{2^N} = \frac{10\text{ mV}}{2^{12}} = \frac{10\text{ mV}}{4096} = 2.44\ \mu\text{V} \qquad (18\text{–}6)$$

The op amp for this application must be a very low noise op amp because an op amp with a 20-nV/(Hz)$^{1/2}$ equivalent input noise voltage and a bandwidth of 4 MHz contributes 40 μV of noise. This high noise contribution is why extensive filtering and “optimally” low bandwidth is found desirable in the input stages of some electronic systems. If there is power supply noise, some of that noise passes through the op amp to its input. The power supply noise is divided by the power supply rejection ratio, but there is always a residual noise component of the power supply on the op amp input as shown in Equation 18–7 where k_{SVR} is 60 dB.

$$V_{PS(INPUT)} = \frac{V_{PS}}{k_{SVR}} = \frac{10\text{ mV}}{1000} = 10\ \mu\text{V} \qquad (18\text{–}7)$$

18.4 Input Common-Mode Range

Years ago the op amp’s input common-mode voltage range (V_{ICR}) did not include the power supply rails. The best V_{ICR} that was available was (V_{CC} +|V_{EE}|–6 V), and when the input voltage approached V_{ICR}, distortion occurred. If the input voltage exceeded the power supply rails, the output stage might invert phase (it sometimes latched in the inverted position causing control problems) or the IC might self destruct. The vast majority of transducers were connected to ground (0 V) because it was easy to make a ground connection and because a split supply op amp has inputs referenced to ground. In a split supply application with the transducer connected to ground, latch-up or self destruction is unlikely.

In special cases, transducers are connected to a power supply rail (usually V_{CC} when power supply current sensing) or some other voltage, and in this special case, additional bias circuitry was added to split power supply designs to keep the input voltage swing within V_{ICR}. Bias circuitry in conjunction with external components removed the effects of the power supply rail connection.

Many low voltage op amp input signals come from transducers connected to a power supply rail like the circuit shown in Figure 18–2.

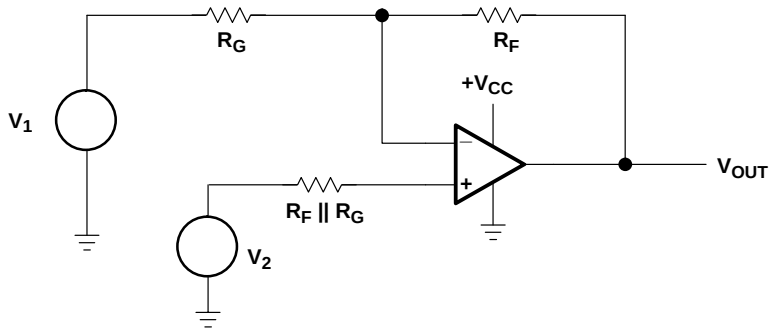


Figure 18–2. Noninverting Op Amp

When $V_1 = 0$ and V_2 is the transducer input, the op amp must be capable of handling input voltages that go to 0 V. Furthermore, the transducer voltage may be ac, so it swings above and below ground, thus the transducer voltage drops below the low power supply rail. This situation requires that the op amp's V_{ICR} exceed the power supply voltage. Rail-to-rail input (RRI) voltage capability is a necessary requirement for a low-voltage op amp that handles transducers connected to a power supply rail.

When the input voltage is connected to ground and the input voltage swing is very small, a standard op amp like the LM324 suffices. Referring to Figure 18–3 it can be seen that the PNP input transistors are biased by the emitter current source. If the positive input is connected to ground bias current still flows and the transistor stays active. If the input transistors are selected very carefully for operation with low collector-base junction reverse bias, the input voltage can go slightly below ground (–200 mV for the TLV278X) and the op amp will still operate correctly. The circuit operation is one sided though because when the input voltage approaches the positive supply rail, the emitter current source and input transistors turn off. This type of circuit does not offer rail-to-rail operation, but it does offer from rail to (V_{CC} –1.5V) operation.

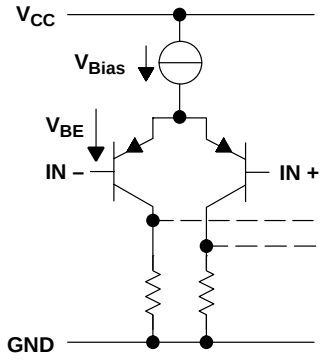


Figure 18–3. Input Circuit of a NonRRI Op Amp

An op amp with a NPN input stage works in a similar way around the positive supply rail. It can sense voltages close to V_{CC} and maybe slightly above V_{CC}, but it won't work when it is within 1.5 V of ground. The solution for this problem is to include parallel input circuits as shown in Figure 18–4.

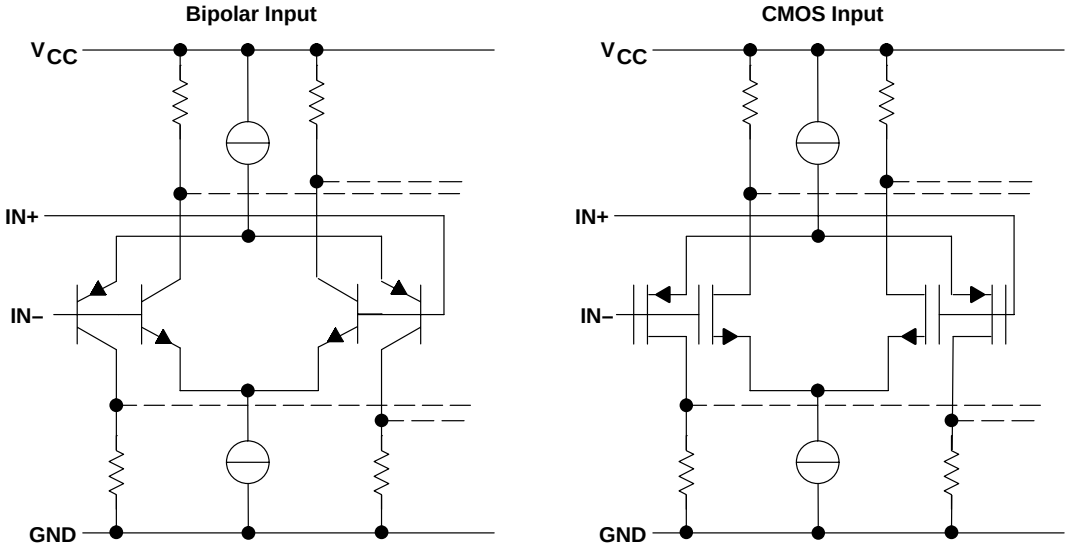


Figure 18–4. Input Circuit of an RRI Op Amp

The RRI op amps have parallel input stages. There are both PNP and NPN differential amplifiers used in the input stages of the RRI op amp, thus the RRI op amp can operate above and below the power supply voltage. As Figure 18–4 shows, the parallel input stages can be made in bipolar or MOS technology.

The input stages operate in three different ranges. When the input voltage ranges from about –0.2 V to 1 V, the PNP differential amplifier is active and the NPN differential amplifier is cutoff. When the input voltage ranges from about 1 V to (V_{CC} –1 V), both the NPN and PNP differential amplifiers are active. When the input voltage ranges from about (V_{CC} –1 V) to (V_{CC} + 0.2 V), the NPN differential amplifier is active and the PNP differential amplifier is cut off. Inclusion of complementary differential input amplifiers achieves V_{ICR} exceeding the power supply limits, but there is a penalty to pay in input bias current, input offset voltage, and distortion. Figures 18–5 and 18–6 show the input bias current and input offset voltage as a function of the input common-mode voltage.

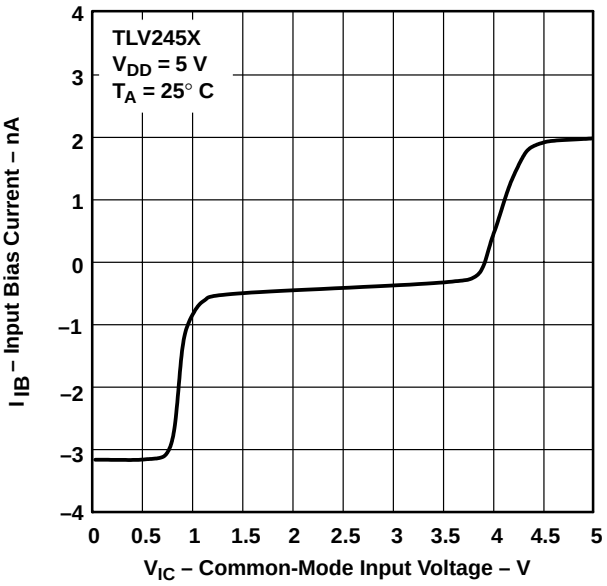


Figure 18–5. Input Bias Current Changes with Input Common-Mode Voltage

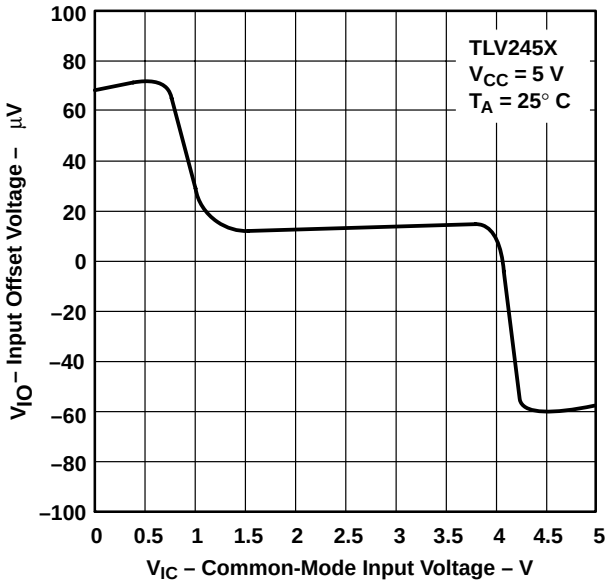


Figure 18–6. Input Offset Voltage Changes with Input Common-Mode Voltage

When both transistors are conducting current the input bias currents have a tendency to cancel, so in the range of ±1 V, the bias current is extremely low even when bipolar transistors are used to make the op amp. Above this range, the PNP differential amplifier cuts off so the full bias current requirement of the NPN transistor becomes apparent. The same action happens below this range when the NPN differential amplifier cuts off. Notice that the PNP bias current is significantly larger than the NPN bias current; this is expected because NPN transistors have better gain characteristics than PNP transistors. The base-emitter voltage of the NPN and PNP transistors is well matched because the magnitude of the input offset voltage at the extremes is almost equal.

The bias current and offset voltage variation with input signal amplitude cause errors and distortion of the input signal. Inserting a resistance equal to the parallel combination of R_F and R_G into the positive op amp lead minimizes the effect of input bias current. The resistor, R_P, has the same voltage drop across it that the parallel combination of R_F and R_G has, hence the bias current is converted to a common-mode voltage. The common-mode voltage is normally in the µV-range because I_{IB} is in the fractional nA range and R_P is in the tens of KΩ. The CMMR is approximately 60 dB, so the input bias current effect is reduced to the nV range where it is insignificant compared to the offset voltage. The input offset current is multiplied by R_P, and it shows up as an input error. If the design can't tolerate these errors it is wise to switch to a CMOS op amp because its input currents are in the pA range.

Another type of error creeps in when complementary differential amplifiers are used to obtain DR, and this error is results from the different gain of the PNP and NPN transistors.

Op amps always suffer to a limited extent from distortion introduced by different gains when operating in different quadrants The positive quadrant is above $V_{CC}/2$ where NPN transistors operate, and the negative quadrant is below $V_{CC}/2$ where the PNP transistors operate. Normally, this is a very minor effect because only the gain of the output stage changes with quadrant, but with complementary input stages the input and output gains change with quadrant. These errors are small, and they are accepted as the sacrifice required for obtaining RRI operation.

18.5 Output Voltage Swing

Rail-to-rail output voltage swing (RRO) is desirable for at least two reasons. First, the DR can achieve the maximum obtainable value if the op amp is RRO. Second RRO op amps can drive any converter connected to the same power supply if the impedance is compatible. The schematic of a RRO op amp output stage, part of the TLC227X, is shown in Figure 18–7.

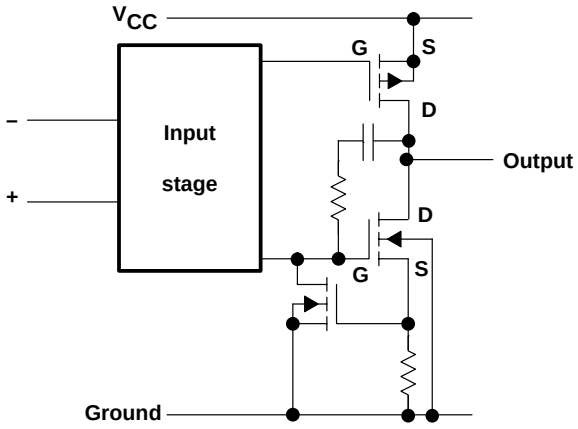


Figure 18–7. RRO Output Stage

The RRO characteristic is achieved in the construction of the op amp output stage. A totem pole design that has upper and lower output transistors is used, and the output transistors are a complimentary pair. Each transistor in the pair is a “self-locking” type of transistor operating in the common-source mode. Consider the p-channel output transistor; as long as this transistor has a drain-source resistance it forms a voltage divider with the load resistance. When the load is a very large resistor or if the output current flow is very small, the voltage drop across the output transistor can be neglected. Output current flows through the output transistor, and because current drops a voltage (V_{DS}) across the drain-source resistor, the output voltage swing is reduced. The voltage drop subtracts from the power supply voltage, reducing the output voltage to less than RRO.

RRO op amps can’t drive heavy loads and maintain their RRO capability because of the voltage dropped across the output transistors. Load resistance or output current is a test

condition when the measurement of an op amp's output voltage swing is made. The size of the load resistor or output current is a measure of the op amp's ability to retain its RRO capability while sourcing or sinking an output current. When selecting a RRO op amp, the designer must consider the load resistance or output current required because these conditions control the output voltage swing.

When an op amp is made that has RRI and RRO capability, it is called a rail-to-rail input/output op amp. This long name is shortened to RRIO.

18.6 Shutdown and Low Current Drain

Low voltage design often is accompanied by a requirement that the power supply current drain be low. The power supply current drain is kept low to decrease battery size and prolong battery charge so recharging can be put off as long as possible. Many methods are employed to keep the current drain low including using high-value resistors, low bias current regulators/references, slow speed logic, keeping logic transitions to a minimum, low voltage power supplies, selecting op amps for low current drain, and shutting off unused ICs.

High-value resistors have less current flowing through them than low value resistors do, and they can be used effectively in ratio applications, but there are some downsides to using high-value resistors. When resistor values exceed 2 MΩ to 10 MΩ, depending on the type of resistor, the temperature drift, vibration, and time-induced drift increases rapidly compared to that of lower value resistors. The input resistor to an op amp, R_G, works with the stray capacitance from the input node to ground to form a pole in the loop gain. As the resistance increases, the pole moves towards the zero frequency intercept and the circuit overshoots, rings, or becomes unstable. The feedback resistor, R_F, works with stray capacitance in parallel with R_F to form a low-pass filter. Sometimes this filter action is desirable, but the filter often distorts the signal.

Very often, low bias current regulators and references are just standard ICs specified at a lower current. These devices generally do not have the same small tolerances at low bias currents that they had at high bias currents. Although they are more often costly, redesigned low bias current regulators and references are becoming available. Ensure that the reference or regulator bias current used in the application is the same as that used to specify the device, because sometimes the error curves for references are nonlinear. Also, investigate the reference noise voltage to ensure that low bias current has not moved the device to a noisy portion of its operating curve.

Saturated logic is the choice for low current drain applications because nonsaturated logic stays in the active region and has a higher current drain. Always pick the slowest logic gates that you can get away with. Speed in saturated logic requires enough current to drive low impedance loads, and that means high power supply currents coupled with logic-generated noise. High-speed logic has a low impedance totem pole output stage, and

every time the output is switched, both totem pole transistors are on causing a current spike through the power supply. Large decoupling capacitors are required to localize the current spike at the logic IC, thus preventing noise propagation. CMOS logic draws the least quiescent current, and if the logic transitions are kept at a minimum, the current drain stays small. One method of minimizing logic transitions is to use asynchronous logic.

The op amp should be selected with current drain in mind. Three rail-to-rail op amps have widely differing current drains because they are designed for different applications. The TLV240X is designed for micropower applications, and its current drain is 1.29 μA. The TLV411X is designed high output drive, and its current drain is 800 μA. The TLV287X is designed for high speed, and its current drain is 820 μA. These three op amps are low-voltage op amps, but they each serve a different application.

The best method of conserving current is to shut the op amp down if you are not using it. Most op amps designed for low voltage applications have shutdown pins. A typical op amp that draws 820 μA when operating, draws 1.7 μA when it is shut down. The problem with shutdown is the time that it takes to wake the op amp up and knowing when to wake the op amp up. A typical low-voltage op amp turns on in less that 1 μs, but the system designer usually has to choose the variable that eventually wakes the op amp up.

18.7 Single-Supply Circuit Design

The op amp is a linear device, so it follows the equation of a straight line. The equation of a straight line has four forms as shown in Equation 18–8.

$$Y = \pm mx \pm b \tag{18–8}$$

These four forms can be implemented with four single supply circuits. When the designer discovers the form of Equation 18–8 that yields the transform function required, it is a small task to find the corresponding circuit. Once the circuit and transfer function are established, the task reduces to matching coefficients between the transfer function and the circuit equation, and then calculating the resistor values. The key required to unlock the puzzle is to determine the form of Equation 18–8 that yields the required transfer function. This key is found in simultaneous equations because they define the equation of a straight line. Several examples of using simultaneous equations to determine the required form of the op amp transfer function are given in the next two sections.

18.8 Transducer to ADC Analog Interface

An example is a transducer that needs to be interfaced to an ADC. The transducer specifications are V_{MIN} = 0.2 V, V_{MAX} = 0.5 V, and R_{OUT} = 600 Ω. The ADC specifications are $V_{IN(LOW)}$ = 1.5 V, $V_{IN(HIGH)}$ = 4.5 V, and R_{IN} = 20 kΩ. The system specifies a 5-V power supply and 5% tolerance resistors. The transducer is connected to input of the amplifier

(see Figure 18–8), so its output voltage swing is renamed V_{IN}, and the ADC is connected to the output of the amplifier, so its input voltage range is renamed V_{OUT}. Now, two data points are constructed as V_{IN1} = 0.2 V @ V_{OUT1} = 1.5 V and V_{IN2} = 0.5 V @ V_{OUT2} = 4.5 V. The data points are substituted into the equation Y = mX + b; m is named the slope and b is named the X axis intercept or just the intercept for short. Don't worry about the sign of m or b because it is determined by the math, and it is substituted into the equation that determines the transfer equation. The simultaneous equations are given below.

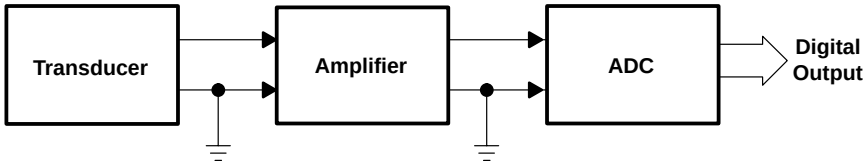


Figure 18–8. Data Acquisition System

$$1.5 = 0.2m + b \tag{18–9}$$

$$4.5 = 0.5m + b \tag{18–10}$$

From these equations we find that b = –0.5 and m = 10. The slope and intercept values are substituted into Equation 18–8 to get Equation 8–11.

$$Y = 10X - 0.5 \tag{18–11}$$

The mathematical terminology in Equation 18–11 is replaced by electronics terminology in Equation 18–12, and this is the transfer function required for the amplifier. The next step is to select the op amp, and this isn't a hard task because there are many candidates that could do the job with these undemanding specifications, so let us not dwell on the selection process. Assume that the selected op amp operates on a 5-V power supply, can drive the ADC input resistance of 20 kΩ with no voltage divider action, and that the op amp input impedance is so big that it doesn't load the transducer.

$$V_{OUT} = 10V_{IN} - 0.5 \tag{18–12}$$

The circuit that produces the desired transfer function is given in Figure 18–9.

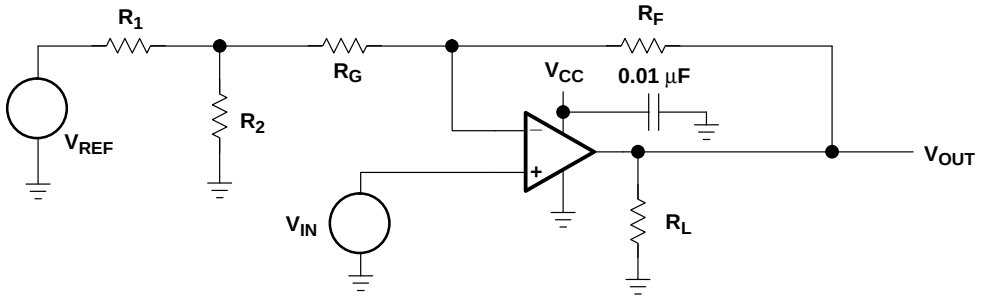


Figure 18–9. Schematic for the Transducer to ADC Interface Circuit

The circuit equation is obtained with the aid of superposition.

$$V_{OUT} = V_{IN}\left(\frac{R_F + R_G + R_1 \| R_2}{R_G + R_1 \| R_2}\right) - V_{REF}\left(\frac{R_2}{R_1 + R_2}\right)\left(\frac{R_F}{R_G + R_1 \| R_2}\right) \quad (18–13)$$

Comparing terms between Equations 18–12 and 18–13 enables the extraction of m and b.

$$m = \left(\frac{R_F + R_G + R_1 \| R_2}{R_G + R_1 \| R_2}\right) \quad (18–14)$$

$$|b| = V_{REF}\left(\frac{R_2}{R_1 + R_2}\right)\left(\frac{R_F}{R_G + R_1 \| R_2}\right) \quad (18–15)$$

Making the assumption that $R_1||R_2 << R_G$ simplifies the calculations.

$$m = 10 = \frac{R_F + R_G}{R_F} \quad (18–16)$$

$$R_F = 9R_G \quad (18–17)$$

Let R_G = 20 kΩ, and then R_F = 180 kΩ, add, and let $V_{REF} = V_{CC}$.

$$|b| = V_{CC}\left(\frac{R_F}{R_G}\right)\left(\frac{R_2}{R_1 + R_2}\right) \quad (18–18)$$

$$R_1 = \frac{1 - 0.01111}{0.01111}R_2 = 89R_2 \quad (18–19)$$

Select R_2 = 0.82 kΩ and R_1 = 72.98 kΩ. Since 72.98 kΩ is not a standard 5% resistor value, R_1 is selected as 75 kΩ. The difference between the selected and calculated value

of R_1 introduces about 3% error in the b coefficient, and this error shows up in the transfer function as an intercept rather than a slope error. The parallel resistance of R_1 and R_2 is approximately 0.82 kΩ and this is much less than R_G which is 20 kΩ, thus the earlier assumption that $R_1||R_2<<R_G$ is justified. R_2 could have been selected as a smaller value, but the smaller values yielded poor standard 5% values for R_1. The final circuit is shown in Figure 18–10.

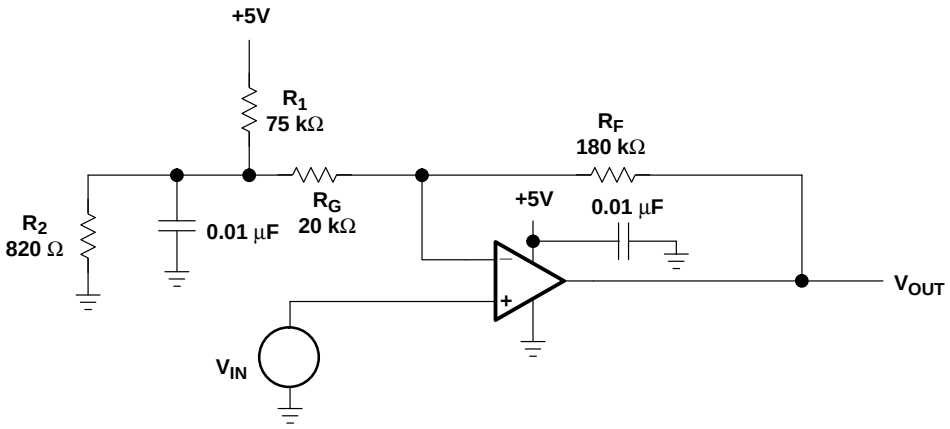


Figure 18–10. Final Schematic for the Transducer to ADC Interface Circuit

18.9 DAC to Actuator Analog Interface

An amplifier is also used to interface a DAC with an actuator (see Figure 18–11).

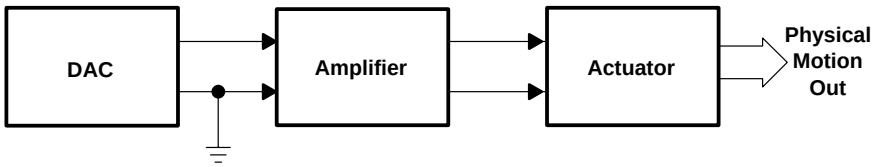


Figure 18–11. Digital Control System

The interface is different from the ADC interface because the DAC output signal is usually current rather than voltage. The first inclination is to stuff the DAC output into a current-to-voltage converter like we have always done with split power supplies. This doesn't always work because the DAC current can be sinked or sourced from ground or the power supply. If a current sourced from the positive power supply is put into a standard current-to-voltage circuit, it wants to drive the op amp output negative, and you need a negative resistor to counter this. The alternate solution for a sourced current from the positive power supply

is to terminate the DAC output in a resistor that converts current into voltage, and then level shift and amplify the terminated voltage. The circuit that performs this function is shown in Figure 18–12.

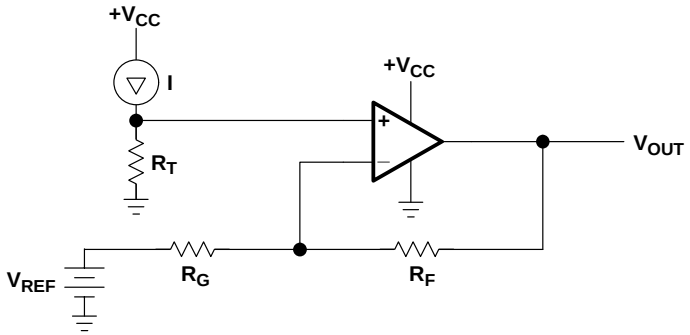


Figure 18–12. DAC Current Source to Actuator Interface Circuit

The DAC output current sources from $I_{OUT(ZEROS)}$ = 1 mA to $I_{OUT(ONES)}$ = 2 mA at an output compliance of 4.33 V. The actuator requires an input voltage swing of V_{IN1} = 1 V to V_{IN2} = 4 V to drive it, and its input resistance is 100 kΩ. The system specifications include one 5-V power supply and 5% resistors. The DAC is connected to input of the amplifier (see Figure 18–11), so its output current swing is renamed I_{IN}, and the actuator is connected to the output of the amplifier, so its input voltage range is renamed V_{OUT}. Now, two data points are constructed as I_{IN1} = 1 mA @ V_{OUT1} = 1 V and I_{IN2} = 2 mA @ V_{OUT2} = 4 V. The data points are substituted into the Equation 18–20. Don't worry about the sign of m or b because it is determined by the math, and it is substituted into the equation that determines the transfer equation. The simultaneous equations are given below.

$$V_{OUT} = mI_{IN} + b \tag{18–20}$$

$$1 = m + b \tag{18–21}$$

$$4 = 2m + b \tag{18–22}$$

From these equations we find that b = –2 and m = 3. The slope and intercept values are substituted into Equation 18–20 to get Equation 18–23.

$$V_{OUT} = 3I_{IN} - 2 \tag{18–23}$$

The equation for the circuit shown in Figure 18–12 is derived with the aid of superposition, and it is given below in Equation 18–24.

$$V_{OUT} = I_{IN}R_T\left(1 + \frac{R_F}{R_G}\right) - V_{REF}\frac{R_F}{R_G} \tag{18–24}$$

Comparing terms between Equations 18–20 and 18–24 enables the extraction of m and b.

$$|b| = V_{REF}\frac{R_F}{R_G} \tag{18–25}$$

$$m = R_T\left(1 + \frac{R_F}{R_G}\right) \tag{18–26}$$

$$\frac{R_F}{R_G} = \frac{2}{5} \tag{18–27}$$

$$R_T = \frac{5\frac{R_F}{R_G} + 1}{\frac{R_F}{R_G} + 1} = 2.14 \tag{18–28}$$

These equations are written in terms of mA and kΩ, so R_T = 2.14 kΩ. There is no 2.14-kΩ resistor in the 5% standard values; thus, R_T is split into 1.8-kΩ and 0.33-kΩ resistors. R_G is selected as 51 kΩ, so R_F = 20 kΩ. When I_{IN} = 2 mA V_{RT} = 4.28 V, so the compliance of the DAC is not violated. You might find that standard DACs are not so generous with their compliance specifications.

When the current is sinked from the power supply by the DAC its sign reverses and the previous circuit is not usable. Consider these specifications: the DAC output sinks current from the power supply $I_{OUT(ZEROS)}$ = –1 mA to $I_{OUT(ONES)}$ = –2 mA at an output compliance of 4.33 V. The actuator requires an input voltage swing of V_{IN1} = 1 V to V_{IN2} = 4 V to drive it, and its input resistance is 100 kΩ. The system specifications include one 5-V power supply and 5% resistors. The DAC is connected to input of the amplifier (see Figure 18–11), so its output current swing is renamed I_{IN}, and the actuator is connected to the output of the amplifier, so its input voltage range is renamed V_{OUT}. Now, two data points are constructed as I_{IN1} = –1 mA @ V_{OUT1} = 1 V and I_{IN2} = –2 mA @ V_{OUT2} = 4 V. The data points are substituted into the Equation 18–20. Don't worry about the sign of m or b because it is determined by the math, and it is substituted into the equation that determines the transfer equation. The transfer function for the current sink DAC is given in Equation 18–29.

$$V_{OUT} = I_{IN}m + b \tag{18–29}$$

The simultaneous equations are given below.

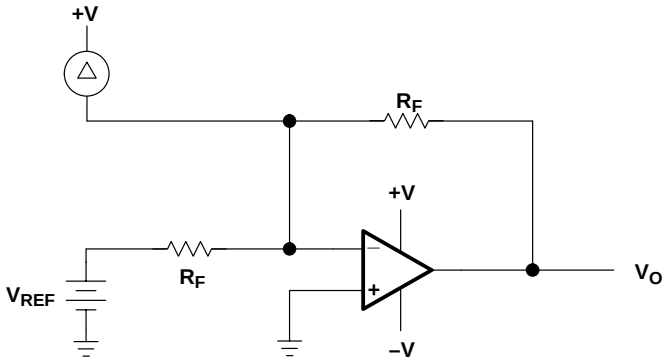


Figure 18–13. DAC Current Sink to Actuator Interface Circuit

$$1 = -m + b \tag{18–30}$$

$$4 = -2 + b \tag{18–31}$$

From these equations we find that b = –2 and m = –3. The slope and intercept values are substituted into Equation 18–28 to get Equation 18–32.

$$V_{OUT} = -I_{IN}(-m) - b = mI_{IN} - b \tag{18–32}$$

The current equation for the circuit shown in Figure 18–13 is given below as Equation 18–33, and after algebraic manipulation it becomes Equation 18–34.

$$\frac{V_{OUT}}{R_F} = I_{IN} - \frac{V_{REF}}{R_G} \tag{18–33}$$

$$V_{OUT} = I_{IN}R_F - V_{REF}\frac{R_F}{R_G} \tag{18–34}$$

Comparing terms between Equations 18–29 and 18–34 enables the extraction of m and b.

$$R_F = |m| = 3 \tag{18–35}$$

$$|b| = V_{REF}\frac{R_F}{R_G} \tag{18–36}$$

$$\frac{R_F}{R_G} = \frac{2}{5} \tag{18–37}$$

These equations are written in terms of mA and kΩ. R_G is selected as 51 kΩ, so R_F = 20 kΩ. When I_{IN} = 2 mA, the compliance of the DAC is 0.0 V.

When DAC interface circuits are designed, two parameters that have not been considered in detail can control the design. The DAC has a compliance voltage requirement, and that requirement must be met regardless of the circuit demands. If the DAC compliance requirements are not met, the DAC saturates or is starved for current, and either of these situations introduces considerable error. The actuators driven in these examples are quite benign because most actuators require considerably more current or voltage than is available from an op amp. This fact does not negate the analysis given here. Regardless of the actuator current or voltage requirements, the design procedure is similar. Very often the low voltage device is plugged into a booster that supplies power to the actuator.

One last item to consider is the output capacitance of DACs. The DAC output can have large amounts of stray capacitance that shows up as a capacitor across the op amp inverting input node when the DAC is interfaced into circuits as shown in Figure 18–13. The DAC capacitance from the inverting node to ground acts with R_G to form a pole in the op amp loop gain. Adding a pole to the op amp loop gain leads to overshoot, then ringing, and finally oscillation. The effect that the DAC capacitance has on stability must be investigated. Also, the DAC output capacitance is a function of the digital number addressing the DAC. This capacitance can range from near zero to hundreds of pF, thus the op amp must be compensated for the worst case which is the largest capacitance.

Compensation schemes include connecting a capacitor across the feedback resistor. This compensation scheme is called a compensated attenuator, and if the RC time constants are equal, there will excellent performance at that DAC output capacitance. Alas, the circuit can only be ideally compensated at one point, and this point is normally chosen as the highest DAC output capacitance. The remainder of the DAC range suffers from poor bandwidth because of overcompensation.

18.10 Comparison of Op Amps

Since the author is only familiar with Texas Instruments op amps, a comparison involving actual op amp parameters would be unfair to other op amp manufacturers. Also, any comparison using today's production op amps becomes invalid in a short period of time. I write about Texas Instruments op amps, so I get plenty of samples, and the new product introductions come so fast that I have a hard time keeping up with them. The other point to consider is that teaching how to make the op amp comparison is a much more powerful tool, thus a table containing the op amp parameters is established, and each of the parameters is discussed only in terms of low-voltage design.

Table 18–2. Op Amp Parameters

PARAMETER	RANGE OF VALUES	UNITS	DISCUSSION
V_{IO}	25 to 8000	mV	DC parameter that can be adjusted out.
αV_{IO}	3 to 1000	μV/°C	Drift parameter that ends up as an error.
I_{IB}	0.1 to 9999	pA	Input bias current that can be cancelled out.
I_{IO}	0.1 to 9999	pA	Input offset current that can be adjusted out.
R_{IN}	0.0002 to 1	GΩ	Acts as a voltage divider with driving circuit.
CMRR	0 to 90	dB	Common-mode voltage is a nonlinear error.
A_{VD}	20 to 140	dB	Determines high-frequency errors.
V_{ICR}	V_{CC} –1.5 V to V_{CC} +0.5 V	V	Input voltage range over which op amp works correctly with specified error.
V_{OH}	V_{CC} –1.5 V to V_{CC} +0.5 V	V	Maximum high output voltage swing. Load resistance or current is important. Limits dynamic range.
V_{OL}	0 + 1.5 V to 0 – 1.5 V	v	Minimum low output voltage swing. Load resistance or current is important. Limits dynamic range.
I_O	1 to 100	mA	Current available to drive loads.
I_{CC}	0.1 to 10000	μA	Power supply current.
k_{SVR}	20 to 120	dB	Power supply noise rejection.
$I_{CC(SHDN)}$	0.01 to 100	μA	Power supply current in shutdown mode.
V_n	1 to 1000	$\frac{nV}{(Hz)^{1/2}}$	Noise voltage limits dynamic range.
I_n	0.01 to 100	$\frac{fA}{(Hz)^{1/2}}$	Noise current limits dynamic range.
$t_{(ON)}$	0.5 to 10	μs	Op amp wake up time.
$t_{(OFF)}$	0.1 to 5	μs	Time until power supply current reaches $I_{CC(SHDN)}$.

As the table shows, there are a lot of parameters to be considered in the selection of a low voltage op amp. The application weeds out some of these parameters thus easing the selection process. If the application is measuring the output from a low-bandwidth transducer such as a thermocouple, bandwidth is not a parameter of interest, so speed can immediately be sacrificed for power supply current.

When the maximum supply voltage available is 3 V, rafts of op amps requiring more than 3 V operating voltage are eliminated. Taking this concept further, if dynamic range is an important specification in the 3-V design, those op amps that operate on 3 V without RRIO specifications can be eliminated. Because dynamic range is important, the noise voltage and current that detract from dynamic range are important parameters. If this 3-V application requires very low power supply current, the choice is narrowed down to a few candidates. The application selects the op amp, and if this application puts a few more requirements on the 3-V op amp, we may have to design a new IC.

Always start the selection process with the parameter that absolutely can't be wavered. The next parameter considered in the selection process should be the next most important parameter, and this process is continued parameter by parameter until all requirements are exhausted. Sometimes the supply of op amp candidates runs out before the parameter requirements do. When you reach this point, it is time to renegotiate the design specifications, find a new op amp, negotiate with op amp manufacturers, or announce that you won't meet specifications. These designs are hard to work because the low power supply voltage requirement and the specifications usually leave very little room to work. As time marches on, more low power supply voltage op amps will come on the market, and these designs will get easier to work.

18.11 Summary

It is extremely hard to achieve large dynamic range when the application is limited to a low power supply voltage. In an attempt to approach the dynamic range obtained by ± 30-V power supply designs, the new op amp designs put increased emphasis on the output voltage swing. The ratio of output voltage swing to power supply voltage was 0.8 for ± 30-V powered op amps, 0.7 for the first 5-V powered op amps, and it has risen to 0.9 for the newest family of 1.8-V powered op amps. The ratio output voltage swing to power supply voltage has increased with each new generation of low power supply voltage op amps, but this improvement has reached the point of diminishing returns.

The op amp's dc offset diminishes the output voltage swing, but in most cases the offsets are adjusted out, so they have less importance in the design. New op amp technology is not being pushed hard to improve in this area because the passive components continue to require the adjustments. Drift and noise continue to decrease the dynamic range. Op amp noise has decreased in new-generation op amps, and another decrease should put noise in the category of "don't care" parameters.

The signal-to-noise ratio has several components that have to be analyzed. The signal comes to the op amp with a noise burden caused by the transducer, cabling, and connections. Making the op amp a filter/amplifier combination eliminates some of this noise. The biggest drawback to making the op amp a filter is the time required to charge the ADC input capacitance. ADC charging has not been investigated here because of scope limitations, but suffice it to say that filters slow down op amps. The internally-generated op amp noise is multiplied by the closed loop gain, and the SNR should be established in the front end, so the closed loop gain hurts one way and helps the other way. There is always system noise, and a portion of this noise propagates through the op amp into the signal. The system noise is minimized by extensive use of decoupling capacitors and a high power supply rejection ratio. In higher voltage systems, a resistor is placed in series with the power supply, thus making the decoupling capacitors more effective; low power supply voltage designs usually can't afford that trick.

RRI op amps are able to work with transducers connected to the power supply rails. As long as the ac component of the transducer output voltage does not exceed the input com-

mon-mode range of the op amp, the design is reliable. RRI op amps are troubled by distortion introduced by the change in bias current, input offset voltage, and gain, but their contribution to the system's signal handling capability is invaluable. RRO op amps yield the highest output voltage swing of any series of op amps. Beware: RRO op amps are specified at a load resistance or current, and the output voltage swing decreases dramatically when the load resistance or current is increased. RRIO op amps contain the input and output features of RRI and RRO op amps. They also contain the drawbacks of both features.

Shutdown is a current-saving feature that is becoming standard on RRIO op amps. When you are working off a battery, there is no reason to waste power when the electronics is not busy, and the shutdown feature accomplishes the power savings by turning the op amp off when it is not needed. The shutdown feature has a disadvantage in that it has a finite wakeup time that the designer must allow for. Don't just depend on shutdown to reduce current drain because there are other ways to reduce current drain, and some of these ways are using high value resistors, low speed logic, fewer logic transitions, and using low bias current regulators/references.

The final thing to be considered is that low power supply voltage invariably means single-supply design, and single-supply design is tougher than split-supply design. Remember to get the two sets of data points, put them in simultaneous equations, solve for the slope and intercept, select the circuit configuration, and calculate the component values. DACs are a little different because you have to account for the polarity of the current, but their design generally follows the same procedure. A good reference for single supply design is the Texas Instruments application report *Single Supply Op Amp Design Techniques* (TI literature number SLOA030).

Appendix A

Single-Supply Circuit Collection

Ron Mancini and Richard Palmer

A.1 Introduction

Portable and single-supply electronic equipment is becoming more popular each day. The demand for single-supply op amp circuits increases with the demand for portable electronic equipment because most portable systems have one battery. Split- or dual-supply op amp circuit design is straightforward because op amp inputs and outputs are referenced to the normally grounded center tap of the supplies. In the majority of split-supply applications, signal sources driving the op amp inputs are referenced to ground, thus with one input of the op amp referenced to ground, as shown in Figure A–1, there is no need to consider input common-mode voltage problems.

$$V_{OUT} = -V_{IN}\frac{R_F}{R_G} \qquad \text{A–1A}$$

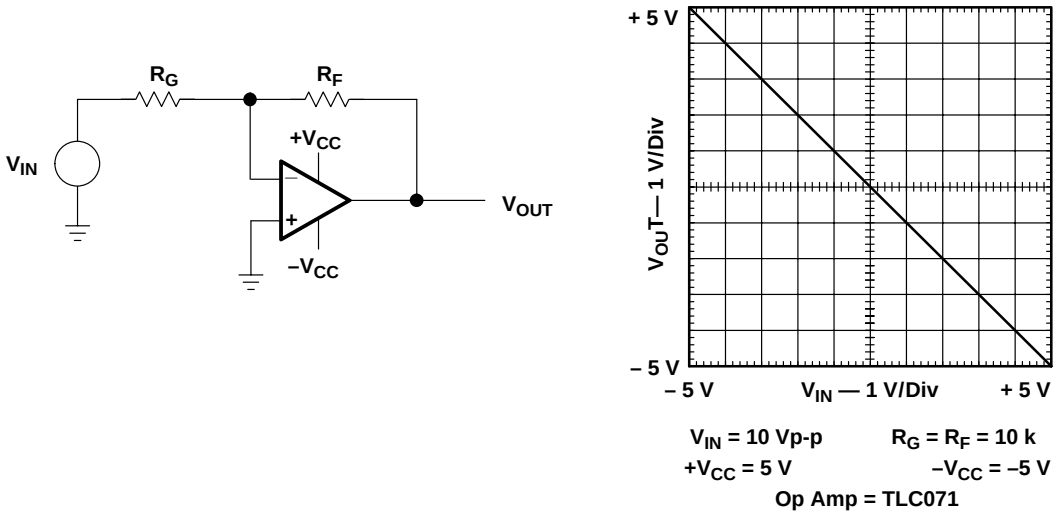


Figure A–1. Split-Supply Op Amp Circuit

When the signal source is not referenced to ground (see Figure A–2 and Equation A–1B), the voltage difference between ground and the reference voltage shows up amplified in the output voltage. Sometimes this situation is OK, but other times the difference voltage must be stripped out of the output voltage.

$$V_{OUT} = -\left(V_{IN} + V_{REF}\right)\frac{R_F}{R_G} \qquad \text{A–1B}$$

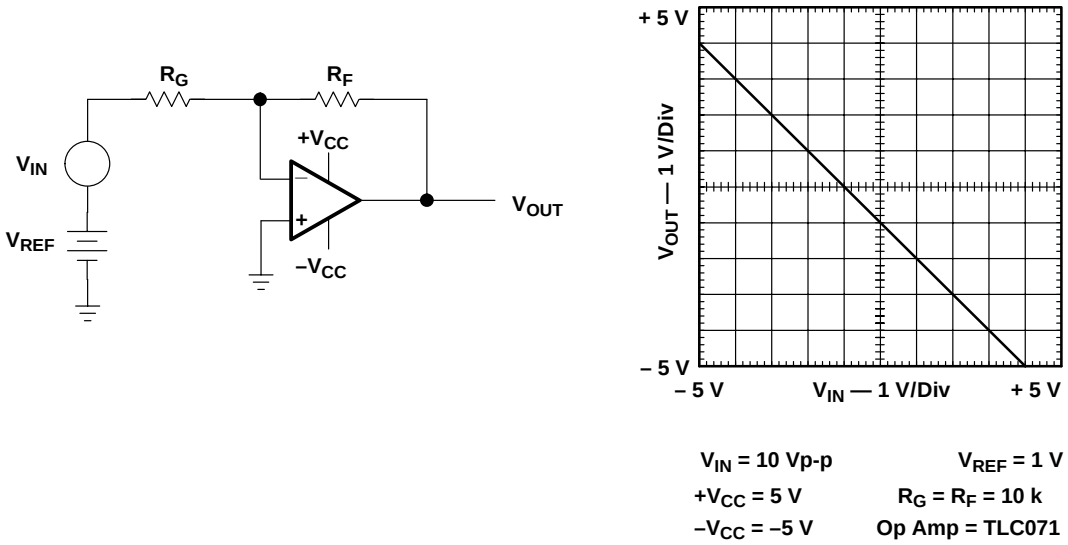


Figure A–2. Split-Supply Op Amp Circuit With Reference Voltage Input

An input bias voltage is used to eliminate the difference voltage when it must not appear in the output voltage (see Figure A–3 and Equation A–1C). The voltage, V_{REF}, is in both input circuits, hence it is named a common-mode voltage. Voltage-feedback op amps, like those used in this document, reject common-mode voltages because their input circuit is constructed with a differential amplifier (chosen because it has natural common-mode voltage rejection capabilities).

$$V_{OUT} = -\left(V_{IN} + V_{REF}\right)\frac{R_F}{R_G} + V_{REF}\left(\frac{R_F}{R_F + R_G}\right)\left(\frac{R_F + R_G}{R_G}\right) = -V_{IN}\frac{R_F}{R_G} \qquad \text{A–1C}$$

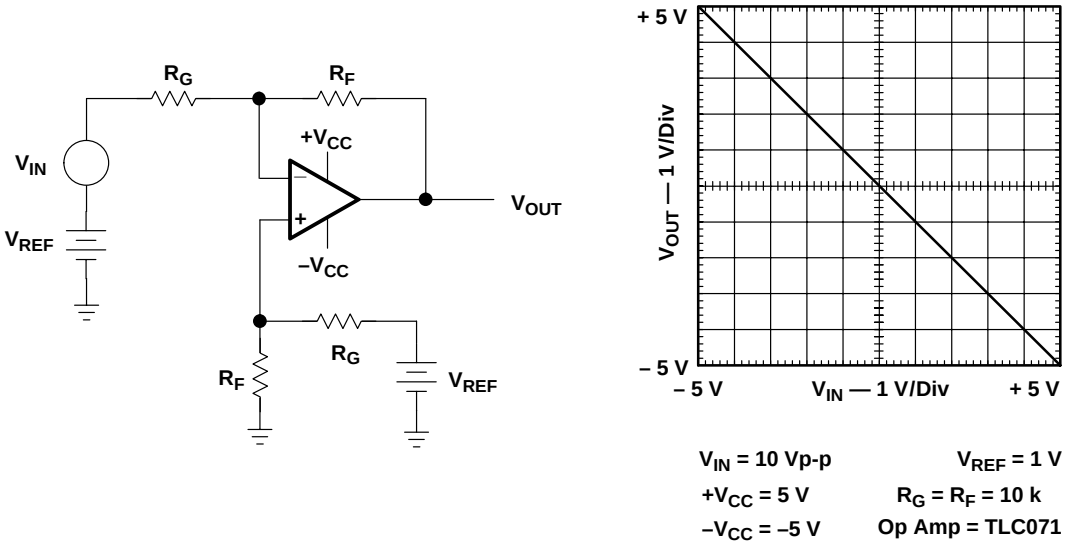


Figure A–3. Split-Supply Op Amp Circuit With Common-Mode Voltage

When signal sources are referenced to ground, single-supply op amp circuits always have a large input common-mode voltage. Figure A–4 shows a single-supply op amp circuit that has its input voltage referenced to ground. The input voltage is not referenced to the midpoint of the supplies like it would be in a split-supply application, rather it is referenced to the lower power supply rail. This circuit malfunctions when the input voltage is positive because the output voltage would have to go negative — hard to do with a positive supply. It operates marginally with small negative input voltages because most op amps do not function well when the inputs are connected to the supply rails.

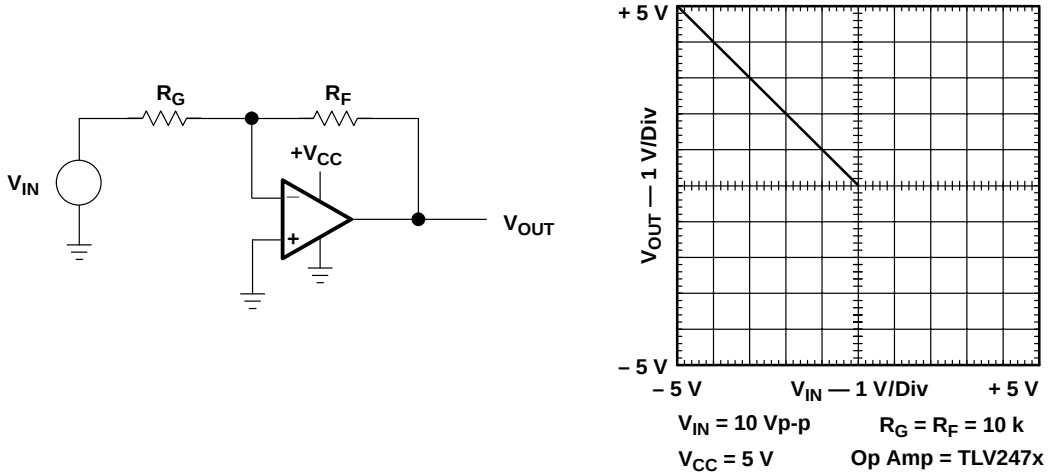


Figure A–4. Single-Supply Op Amp Circuit

The constant requirement to account for inputs connected to ground or other reference voltages makes it difficult to design single-supply op amp circuits. This appendix presents a collection of single-supply op amp circuits, including their description and transfer equation. Those without a good working knowledge of op amp equations should reference the *Understanding Basic Analog* series of application notes available from Texas Instruments. Application note SLAA068, *Understanding Basic Analog — Ideal Op Amps* develops the ideal op amp equations. Circuit equations in this appendix are written with the ideal op amp assumptions as specified in *Understanding Basic Analog — Ideal Op Amps.* The assumptions appear in Table A–1 for easy reference.

Table A–1. Ideal Op Amp Assumptions

PARAMETER NAME	PARAMETER SYMBOL	VALUE
Input current	I_{IN}	0
Input offset voltage	V_{OS}	0
Input impedance	Z_{IN}	∞
Output impedance	Z_{OUT}	0
Op amp gain	a	∞

Detailed information about designing single-supply op amp circuits appears in application note SLOA030, *Single-Supply Op Amp Design Techniques*. Unless otherwise specified, all op amp circuits shown here are single-supply circuits. The single supply may be wired with the negative or positive lead connected to ground, but as long as the supply polarity is correct, the wiring does not affect circuit operation.

A.2 Boundary Conditions

All op amps are constrained to output voltage swings less than or equal to their power supply. Use of a single supply limits the output voltage to the range of the supply voltage. For example, when the supply voltage V_{CC} equals +10 V, the output voltage is limited to the range $0 \le V_{OUT} \le 10$. This limitation precludes negative output voltages when the circuit has a positive supply voltage, but it does not preclude negative input voltages. As long as the voltage on the op amp input leads does not become negative, the circuit can handle negative voltages applied to the input resistors.

Beware of working with negative (positive) input voltages when the op amp is powered from a positive (negative) supply because op amp inputs are highly susceptible to reverse-voltage breakdown. Also, ensure that no *start-up* condition reverse biases the op amp inputs when the input and supply voltage are opposite polarity. It may be advisable to protect the op amp inputs with a diode (Schottky or germanium) connected anode to ground and cathode to the op amp input.

A.3 Amplifiers

Many types of amplifiers can be created using op amps. This section consists of a selection of some basic, single-supply op amp circuits that are available to the designer during the concept stage of a design. The circuit configuration and correct single-supply dc biasing techniques are presented for the following cases: inverting, noninverting, differential, T-network, buffer and ac-coupled amplifiers.

A.3.1 Inverting Op Amp with Noninverting Positive Reference

The ideal transfer equation is given in Equation A–1.

$$V_{OUT} = -V_{IN}\frac{R_F}{R_G} + V_{REF}\left(\frac{R_F + R_G}{R_G}\right) \qquad \text{A–1}$$

The transfer equation for this circuit (Figure A–5) takes the form of Y = –mX + b. The transfer function slope is negative, and the dc intercept is positive. R_F and R_G are contained in both halves of the equation, thus it is hard to obtain the desired slope and dc intercept without modifying V_{REF}. This is the minimum component count configuration for this transfer function. When the reference voltage is 0, the input voltage is constrained to negative voltages because positive input voltages would cause the output voltage to saturate at ground.

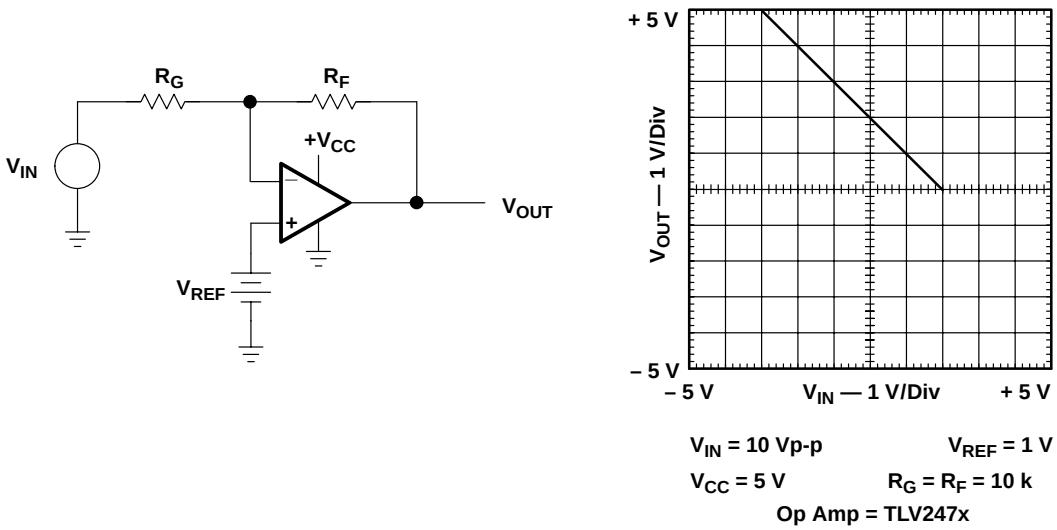


Figure A–5. Inverting Op Amp with Noninverting Positive Reference

A.3.2 Inverting Op Amp with Inverting Negative Reference

The transfer equation takes the form of Y = –mX + b and is given in Equation A–2.

$$V_{OUT} = -V_{IN}\frac{R_F}{R_{G1}} + V_{REF}\frac{R_F}{R_{G2}} \qquad \text{A–2}$$

The transfer function slope is negative, and the dc intercept is positive (Figure A–6). R_{G1} and R_{G2} are contained in the equation, thus it is easy to obtain the desired slope and dc intercept by adjusting the value of both resistors. Because of the virtual ground at the inverting input, R_{G2} is the terminating impedance for V_{REF}. When the reference voltage is 0, the input voltage is constrained to negative voltages because positive input voltages would cause the output voltage to saturate at ground.

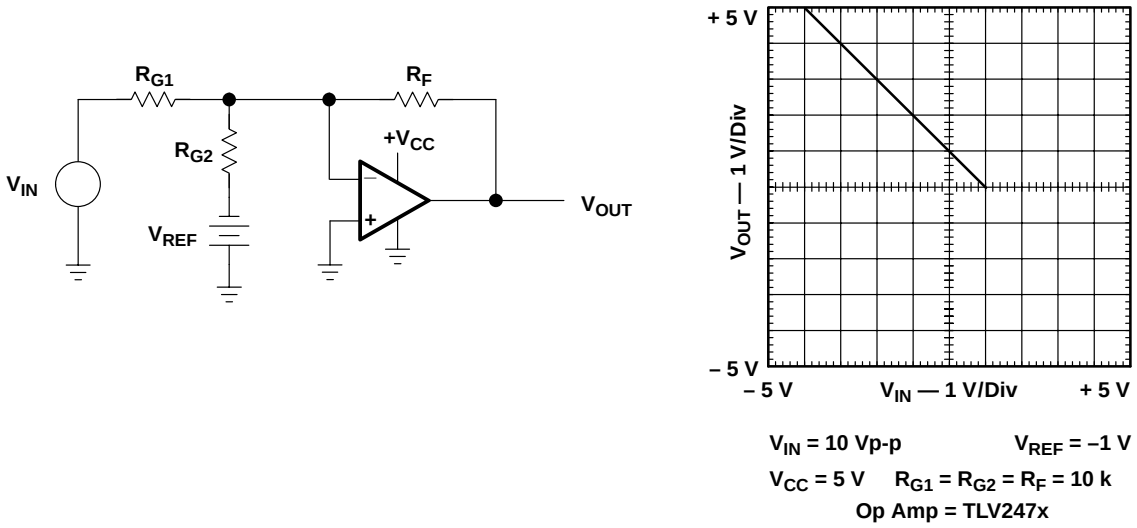


Figure A–6. Inverting Op Amp with Inverting Negative Reference

A.3.3 Inverting Op Amp with Noninverting Negative Reference

The transfer equation takes the form of Y = –mX – b and is given in Equation A–3.

$$V_{OUT} = -V_{IN}\frac{R_F}{R_G} - V_{REF}\left(\frac{R_F + R_G}{R_G}\right) \qquad \text{A–3}$$

The transfer function slope is negative, and the dc intercept is negative. R_F and R_G are contained in both halves of the equation, thus it is hard to obtain the desired slope and dc intercept without modifying V_{REF}. This is the minimum component count configuration for this transfer function. The slope and dc intercept terms in Equation A–3 are both negative, hence, unless the correct input voltage range is selected, the output voltage will saturate at ground. The negative input voltage must be limited to less than –400 mV because op amp inputs either break down or have protection circuits that forward bias when large negative voltages are applied to the inputs.

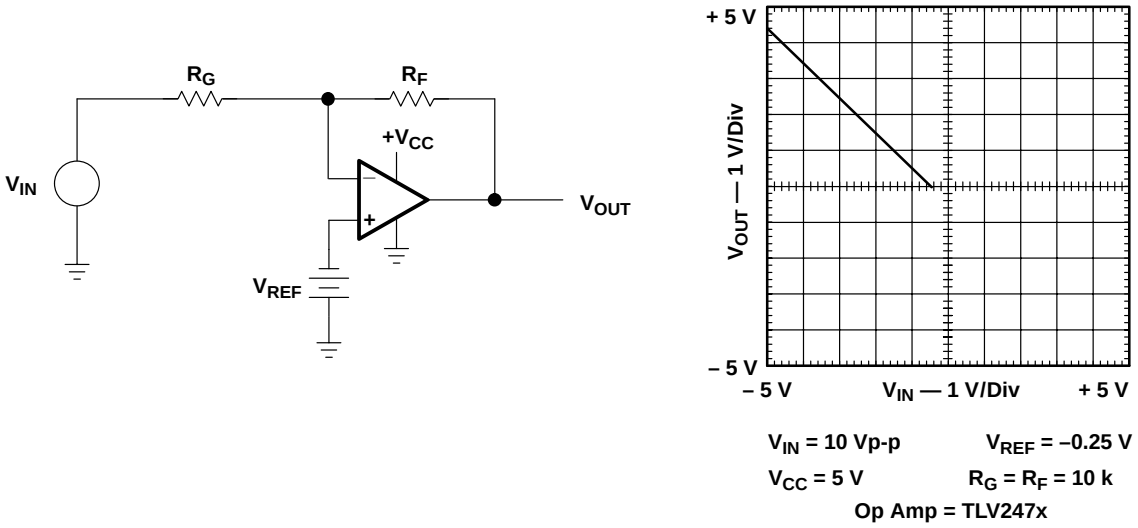


Figure A–7. Inverting Op Amp with Noninverting Negative Reference

A.3.4 Inverting Op Amp with Inverting Positive Reference

The transfer equation takes the form of Y = –mX – b and is given in Equation A–4.

$$V_{OUT} = -V_{IN}\frac{R_F}{R_{G1}} - V_{REF}\frac{R_F}{R_{G2}} \qquad \text{A–4}$$

The transfer function slope is negative, and the dc intercept is negative. R_{G1} and R_{G2} are contained in the equation, thus it is easy to obtain the desired slope and dc intercept by adjusting the value of both resistors. Because of the virtual ground at the inverting input, R_{G2} is the terminating impedance for V_{REF}. The slope and dc intercept terms in Equation A–4 are both negative, hence, unless the correct input voltage range is selected, the output voltage will saturate.

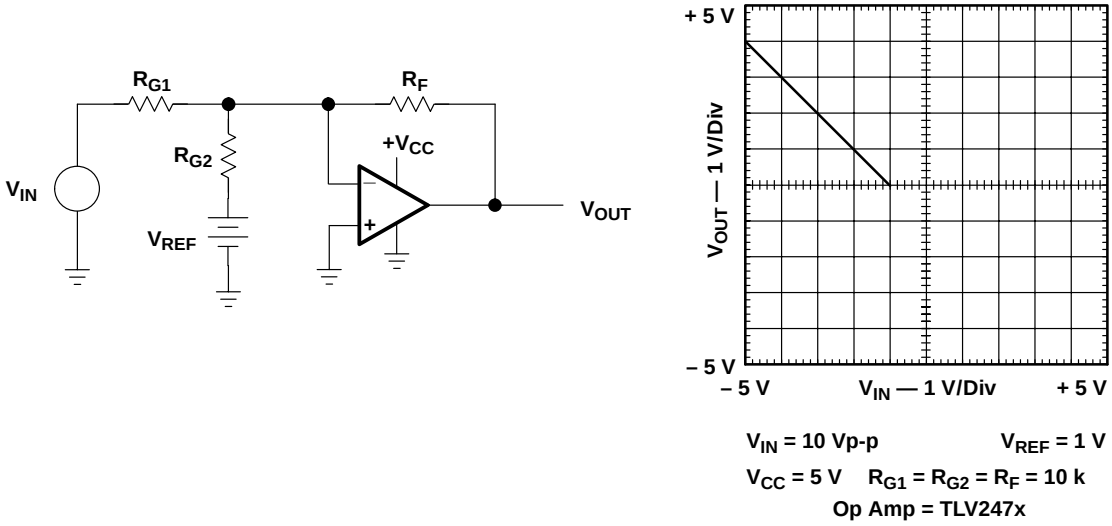


Figure A–8. Inverting Op Amp with Inverting Positive Reference

A.3.5 Noninverting Op Amp with Inverting Positive Reference

The transfer equation takes the form of Y = mX – b and is given in Equation A–5.

$$V_{OUT} = V_{IN} \frac{R_F + R_G}{R_G} - V_{REF} \frac{R_F}{R_G} \qquad \text{A–5}$$

The transfer function slope is positive, and the dc intercept is negative. This is the minimum component count configuration for this transfer function. The reference termination resistor is connected to a virtual ground, so R_G is the load across V_{REF}. R_F and R_G are contained in both halves of the equation, thus it is hard to obtain the desired slope and dc intercept without modifying V_{REF} or placing an attenuator in series with V_{IN}.

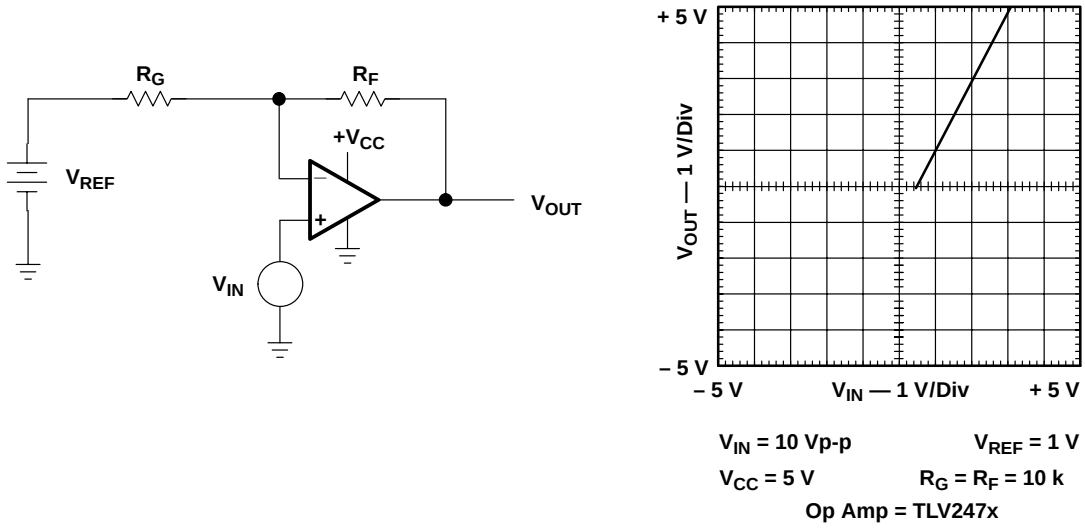


Figure A–9. Noninverting Op Amp with Inverting Positive Reference

A.3.6 Noninverting Op Amp with Noninverting Negative Reference

The transfer equation takes the form of Y = mX – b and is given in Equation A–6.

$$V_{OUT} = V_{IN}\left(\frac{R_2}{R_1 + R_2}\right)\frac{R_F + R_G}{R_G} - V_{REF}\left(\frac{R_1}{R_1 + R_2}\right)\frac{R_F + R_G}{R_G} \qquad \text{A–6}$$

The transfer function slope is positive, and the dc intercept is negative. The reference is terminated in R_1 and R_2. R_1 and R_2 can be selected independent of R_F and R_G to obtain the desired slope and dc intercept. The price for the extra degree of freedom is two resistors.

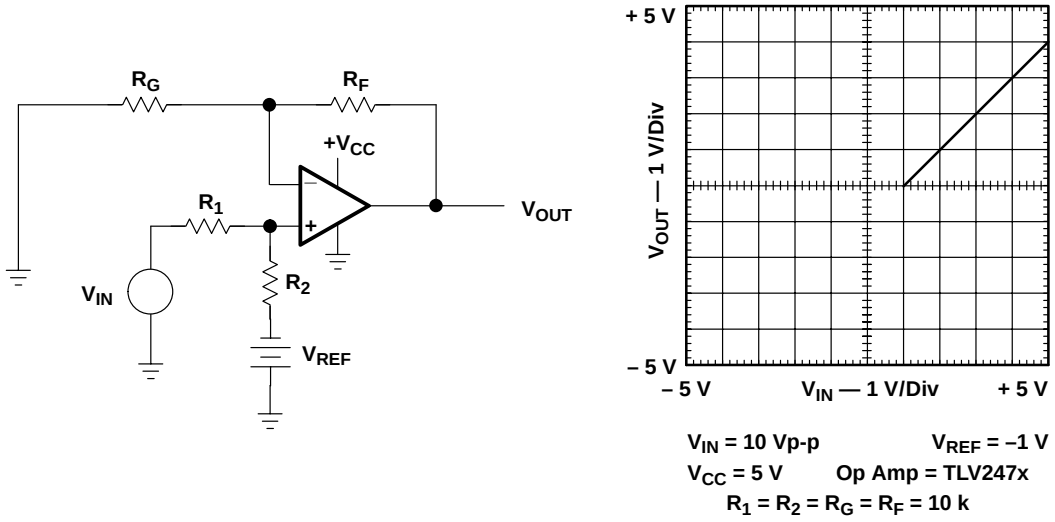


Figure A–10. Noninverting Op Amp with Noninverting Negative Reference

A.3.7 Noninverting Op Amp with Inverting Negative Reference

The transfer equation takes the form of Y = mX + b and is given in Equation A–7.

$$V_{OUT} = V_{IN} \frac{R_F + R_G}{R_G} + V_{REF} \frac{R_F}{R_G} \qquad \text{A–7}$$

The transfer function slope is positive, and the dc intercept is positive. This is the minimum component count configuration for this transfer function. The reference termination resistor is connected to a virtual ground, so R_G is the load across V_{REF}. R_F and R_G are contained in both halves of the equation. Thus it is hard to obtain the desired slope and dc intercept without modifying V_{REF} or placing an attenuator in series with V_{IN}.

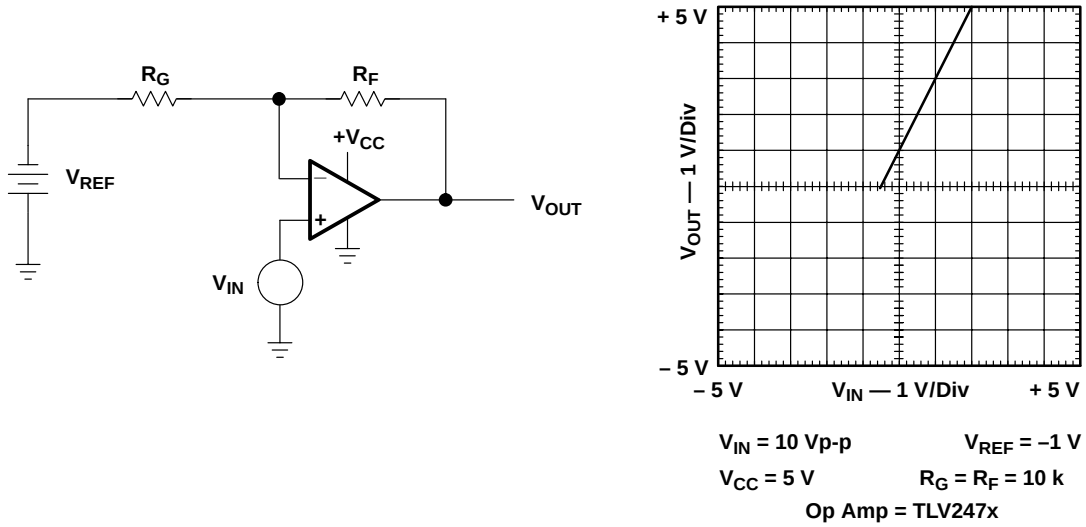


Figure A–11. Noninverting Op Amp with Inverting Positive Reference

A.3.8 Noninverting Op Amp with Noninverting Positive Reference

The transfer equation takes the form of Y = mX + b and is given in Equation A–8.

$$V_{OUT} = V_{IN}\left(\frac{R_2}{R_1 + R_2}\right)\frac{R_F + R_G}{R_G} + V_{REF}\left(\frac{R_1}{R_1 + R_2}\right)\frac{R_F + R_G}{R_G} \qquad \text{A–8}$$

The transfer function slope is positive, and the dc intercept is positive. The reference is terminated in R_1 and R_2. R_1 and R_2 can be selected independent of R_F and R_G to obtain the desired slope and dc intercept. The price for the extra degree of freedom is two resistors.

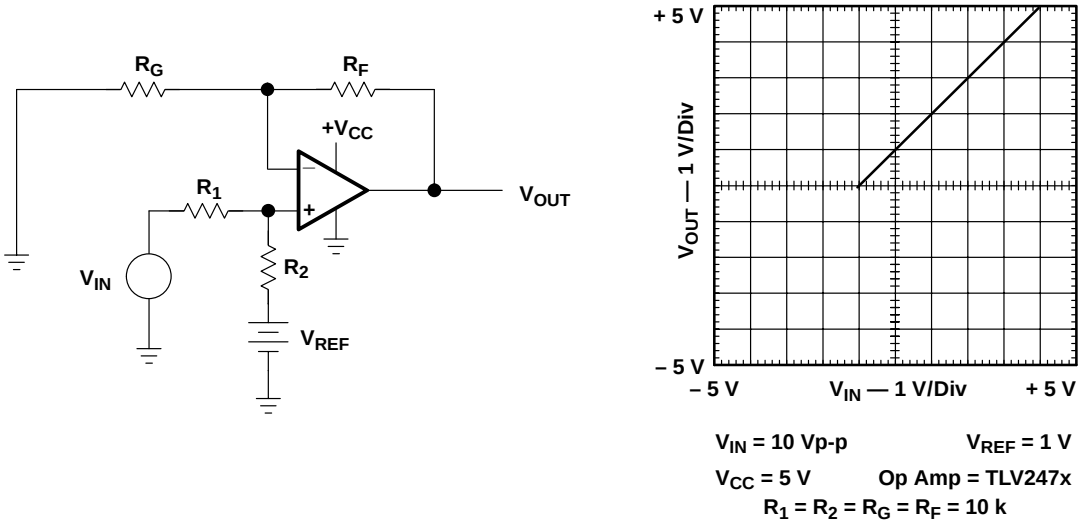


Figure A–12. Noninverting Op Amp with Noninverting Positive Reference

A.3.9 Differential Amplifier

When R_F is set equal to R_2 and R_G is set equal to R_1, Equation A–9 reduces to Equation A–10.

$$V_{OUT} = V_{IN2}\,\frac{R_F + R_G}{R_G}\left(\frac{R_2}{R_1 + R_2}\right) - V_{IN1}\,\frac{R_F}{R_G} \qquad \text{A–9}$$

$$V_{OUT} = \left(V_{IN2} - V_{IN1}\right)\frac{R_F}{R_G} \qquad \text{A–10}$$

These resistors must be matched very closely to obtain good differential performance. The mismatch error in these resistors reduces the common-mode performance, and the mismatch shows up in the output as an amplified common-mode voltage.

Consider Equation A–10. Note that only the difference signal is amplified, thus this configuration is called a differential amplifier. The differential amplifier is a popular circuit in precision applications where it is used to amplify sensor outputs while rejecting common-mode noise.

The inverting input impedance is R_G because of the virtual ground at the inverting op amp input. The noninverting input impedance is $R_F + R_G$ because the noninverting op amp input impedance approaches infinity. The two input impedances are different, and this leads to two problems with this circuit.

First, mismatched input impedances preclude any attempts to cancel input bias currents through resistor matching. Often R_2 is set equal to $R_F \,||\, R_G$ so that the bias currents develop equal common-mode voltages which the op amp rejects. This is not possible when $R_2 = R_F$ and $R_1 = R_G$ unless the source impedances are matched. Second, high output impedance sensors are often used, and when high output sensors work into mismatched input impedances, errors occur.

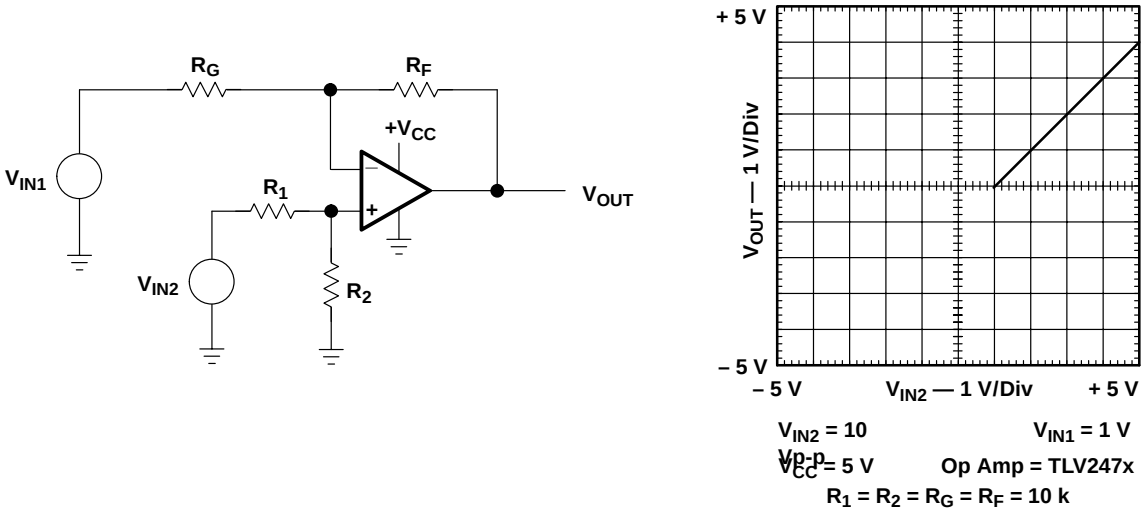


Figure A–13. Differential Amplifier

A.3.10 Differential Amplifier With Bias Correction

When R_F is set equal to R_2 and R_G is set equal to R_1, Equation A–11 reduces to Equation A–12.

$$V_{OUT} = V_{IN2}\frac{R_F + R_G}{R_G}\left(\frac{R_2}{R_1 + R_2}\right) + V_{REF}\frac{R_F + R_G}{R_G}\left(\frac{R_1}{R_1 + R_2}\right) - V_{IN1}\frac{R_F}{R_G} \quad \text{A–11}$$

$$V_{OUT} = \left(V_{IN2} - V_{IN1}\right)\frac{R_F}{R_G} + V_{REF} \quad \text{A–12}$$

When an offset voltage must be eliminated from or added to the input signal, this differential amplifier circuit is employed. The reference voltage can be positive or negative depending upon the polarity offset required, but care must be taken to protect the op amp inputs and not exceed the output range.

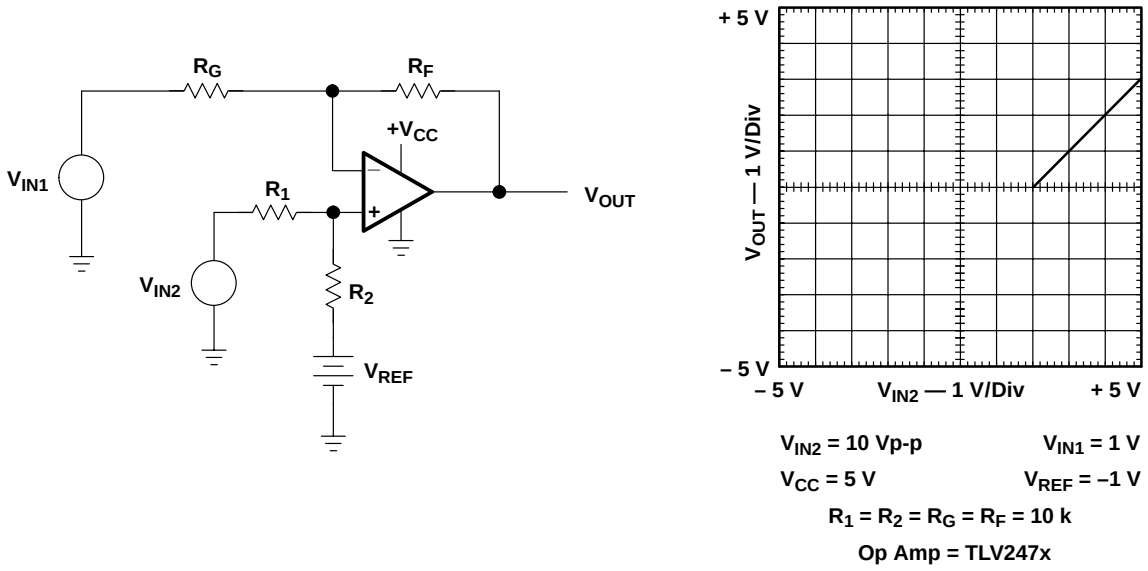


Figure A–14. Differential Amplifier with Bias Correction

A.3.11 High Input Impedance Differential Amplifier

When R_F is set equal to R_1 and R_G is set equal to R_2, Equation A–13 reduces to Equation A–14.

$$V_{OUT} = V_{IN1} \frac{R_1 + R_2}{R_1}\left(-\frac{R_F}{R_G}\right) + V_{IN2} \frac{R_F + R_G}{R_G} \qquad \text{A–13}$$

$$V_{OUT} = \left(V_{IN2} - V_{IN1}\right)\frac{R_F + R_G}{R_G} \qquad \text{A–14}$$

Each input signal is connected to an op amp noninverting input that is very high impedance. The input impedance of the circuit is very high, and it is matched, so this circuit is often used to interface to high-impedance sensors. Each op amp has a signal propagation time, and V_{IN1} experiences two propagation delays versus V_{IN2}'s one propagation delay. At high frequencies, the propagation delay becomes a significant portion of the signal period, and this configuration is not usable at that frequency.

R_F and R_2, and R_G and R_1 should be matched to achieve good common-mode rejection capability. Bias current cancellation resistors equal to $R_F \parallel R_G$ should be connected in series with the input sources for precision applications.

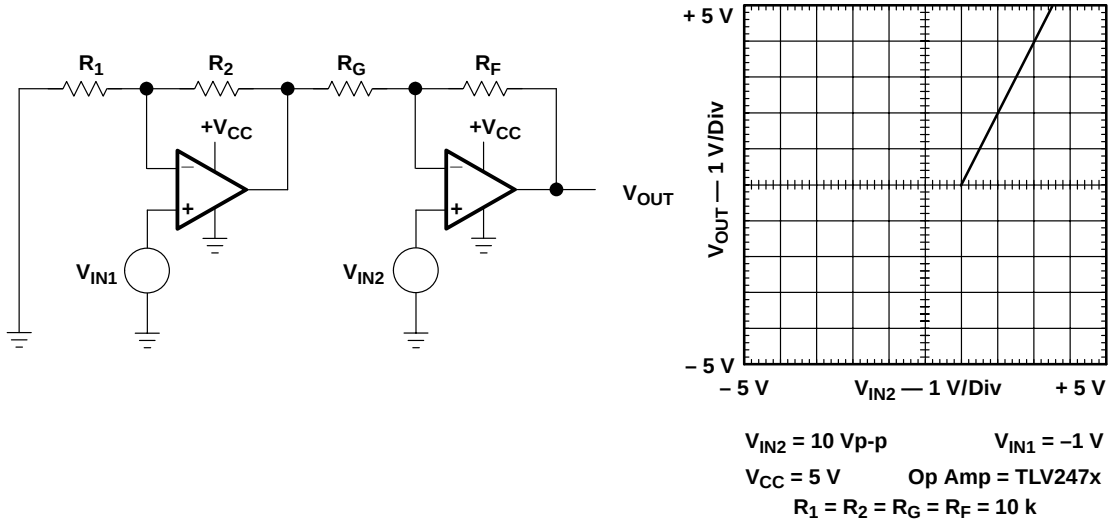


Figure A–15. High Input Impedance Differential Amplifier

A.3.12 High Common-Mode Range Differential Amplifier

When all resistors are equal, Equation A–15 reduces to Equation A–16.

$$V_{OUT} = V_{IN2}\left(-\frac{R_2}{R_1}\right)\left(-\frac{R_F}{R_{G1}}\right) + V_{REF1}\left(\frac{R_1 + R_2}{R_1}\right)\left(-\frac{R_F}{R_{G1}}\right) + V_{IN1}\left(-\frac{R_F}{R_{G2}}\right) + V_{REF2}\left(\frac{R_F + R_{G1} \parallel R_{G2}}{R_{G1} \parallel R_{G2}}\right) \quad \text{A–15}$$

$$V_{OUT} = \left(V_{IN2} - V_{IN1}\right) - 2V_{REF1} + 3V_{REF2} \quad \text{A–16}$$

R_1 and R_{G2} are equal-value resistors terminated into a virtual ground, hence, the input sources are equally terminated. This configuration has high common-mode capability because R_1 and R_{G2} limit the current that can flow into or out of the op amp. Thus, the input voltage can rise to any value that does not exceed the op amp's drive capability. The voltage references, V_{REF1} and V_{REF2}, are added for bias purposes. Without bias, the output voltage of the op amps would saturate at ground, and the bias voltages keep the output voltage of the op amp positive.

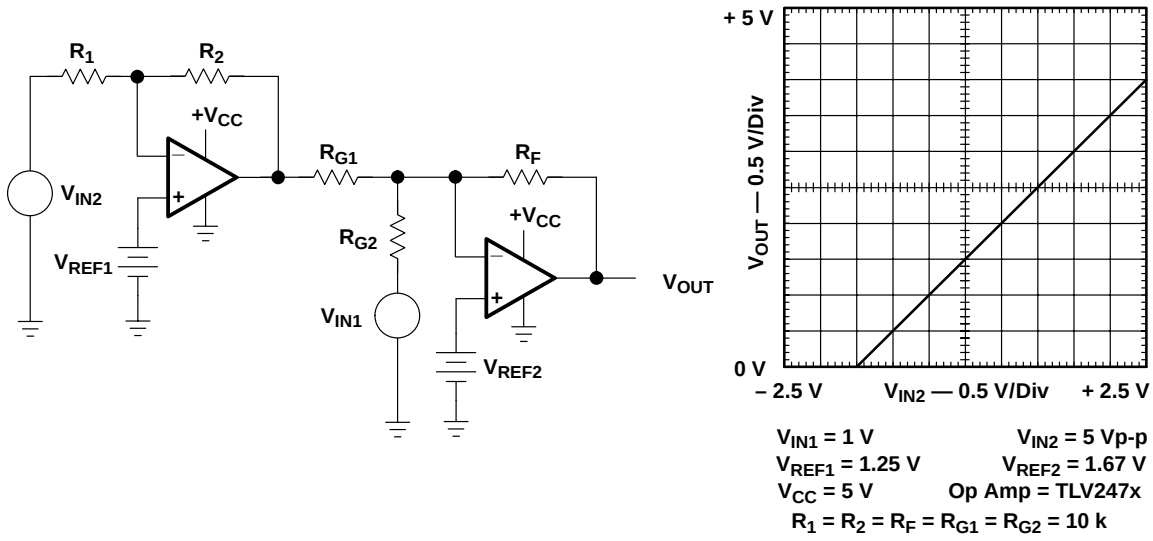


Figure A–16. High Common-Mode Range Differential Amplifier

A.3.13 High-Precision Differential Amplifier

When $R_7 = R_6$, $R_5 = R_2$, $R_1 = R_4$, and $V_{REF1}=V_{REF2,}$ Equation A–17 reduces to Equation A–18.

$$V_{OUT} = \left(V_{IN2} + V_{REF2}\right)\left(\frac{2R_4 + R_3}{R_3}\right)\left(\frac{R_7}{R_5 + R_7}\right)\left(\frac{R_6 + R_2}{R_2}\right) - \left(V_{IN1} + V_{REF1}\right)\left(\frac{2R_1 + R_3}{R_3}\right)\frac{R_6}{R_2} + V_{REF3}\left(\frac{R_5}{R_5 + R_7}\right)\left(\frac{R_6 + R_2}{R_2}\right) \quad \text{A–17}$$

$$V_{OUT} = \left(V_{IN2} - V_{IN1}\right)\left(\frac{2R_1}{R_3} + 1\right)\left(\frac{R_6}{R_2}\right) + V_{REF3} \quad \text{A–18}$$

In this circuit configuration, both sources work into the input impedance of a noninverting op amp. This impedance is very high, and if the op amps are identical, both impedances are very nearly equal. The propagation delay is still equal to two op amp propagation delays, but the propagation delay is very nearly equal, so any distortion resulting from unequal propagation delays is minimized.

The equal resistors should be matched with more precision than is expected from the circuit. Resistor matching eliminates distortion due to unequal gains, and it reduces the common-mode voltage feed through. Resistors equal to $(R_1 \parallel R_3)/2$ may be placed in series with the sources to reduce errors resulting from bias currents. This differential amplifier has the unique feature that the gain can be changed with only one resistor, and if the gain setting resistor is R3, no resistor matching is required to change gain.

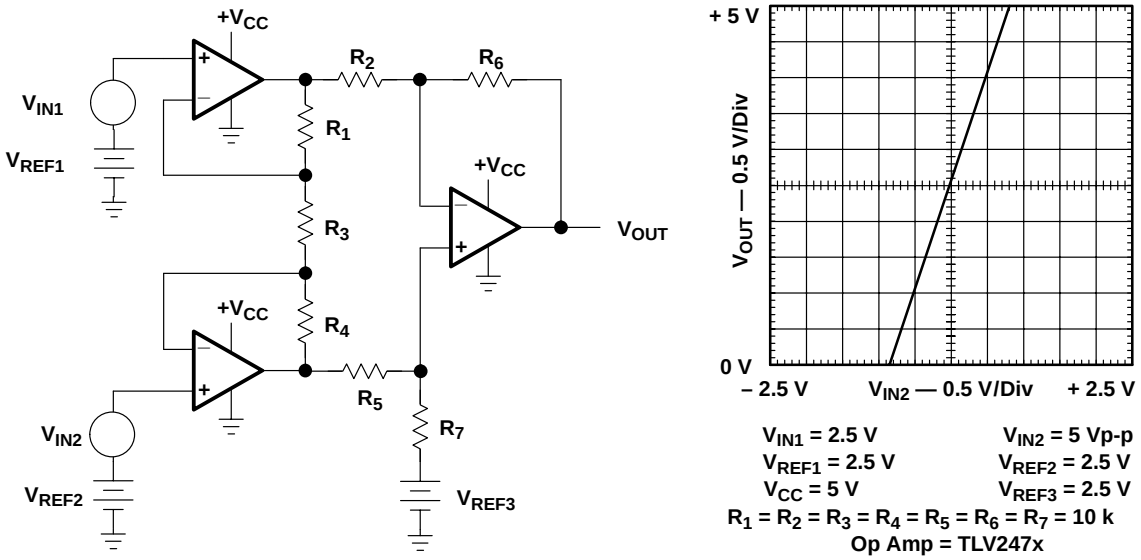


Figure A–17. High-Precision Differential Amplifier

A.3.14 Simplified High-Precision Differential Amplifier

When R_F is set equal to R_2, R_G is set equal to R_1, and $V_{REF1} = V_{REF2}$, Equation A–19 reduces to Equation A–20.

$$V_{OUT} = (V_{IN2} + V_{REF2})\left(\frac{R_F + R_G}{R_G}\right)\left(\frac{R_2}{R_1 + R_2}\right) - (V_{IN1} + V_{REF1})\left(\frac{R_F}{R_G}\right) + V_{REF3}\left(\frac{R_1}{R_1 + R_2}\right)\left(\frac{R_F + R_G}{R_G}\right) \quad \text{A–19}$$

$$V_{OUT} = (V_{IN2} - V_{IN1})\left(\frac{R_F}{R_G}\right) + V_{REF3} \quad \text{A–20}$$

Both input sources are loaded equally with very high impedances in the simplified high precision differential amplifier. This configuration eliminates three resistors, two of which are matched, but it sacrifices flexibility in gain setting capability because the gain must be set with a matched pair of resistors.

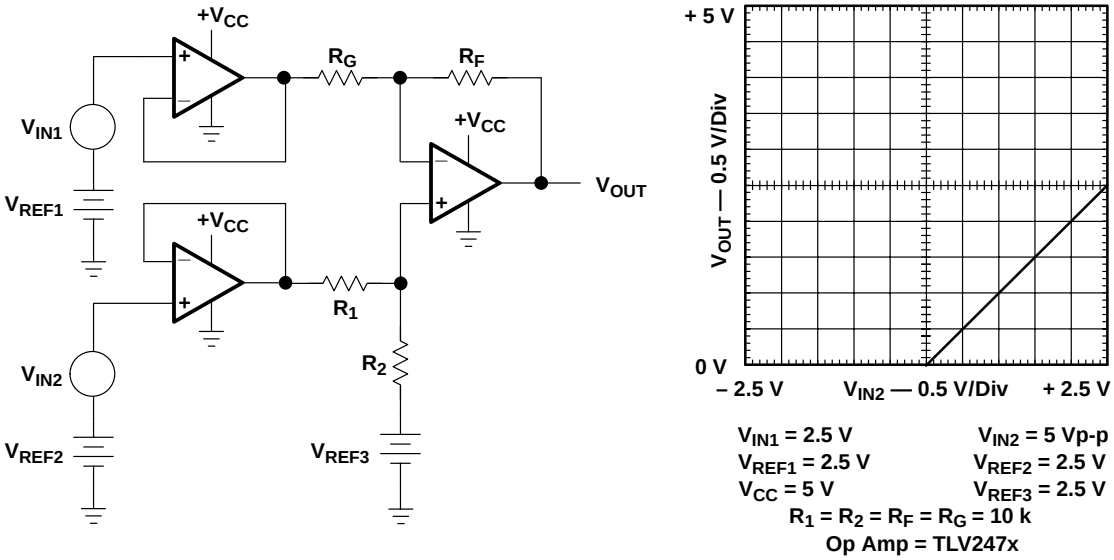


Figure A–18. Simplified High-Precision Differential Amplifier

A.3.15 Variable Gain Differential Amplifier

When R_1 is set equal to R_3 and R_2 is set equal to R_4, Equation A–21 reduces to Equation A–22.

$$V_{OUT} = V_{IN1}\left(\frac{R_2}{R_1 + R_2}\right)\left(\frac{R_3 + R_4}{R_3}\right)\left(\frac{R_G}{R_F}\right) - V_{IN2}\left(\frac{R_4}{R_3}\right)\left(\frac{R_G}{R_F}\right) \qquad \text{A–21}$$

$$V_{OUT} = \left(V_{IN2} - V_{IN1}\right)\left(\frac{R_4}{R_3}\right)\left(\frac{R_F}{R_G}\right) \qquad \text{A–22}$$

When a function is enclosed in a feedback loop, the function acts inverted on the closed loop transfer function. Thus, the gain stage R_F/R_G ends up being an attenuator. The circuit shown in Figure A–19 can be used with any of the differential amplifiers to change gain without affecting matched resistors. R_1, R_3 and R_2, R_4 must be matched to reduce the common-mode voltage.

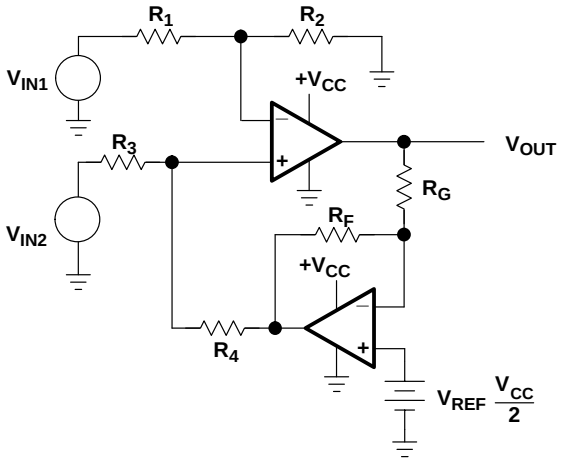


Figure A–19. Variable Gain Differential Amplifier

A.3.16 *T* Network in the Feedback Loop

Sometimes it is desirable to have a low-resistance path to ground in the feedback loop. Standard inverting op amps cannot do this when the driving circuit sets the input resistor value and the gain specification sets the feedback resistor value. Inserting a *T* network in the feedback loop yields a degree of freedom that enables both specifications to be met with a low dc resistance path to ground in the feedback loop.

$$V_{OUT} = -V_{IN}\left(\frac{R_2 + R_3 + \frac{R_2R_3}{R_4}}{R_1}\right) + V_{REF}\left(1 + \frac{R_2 + R_3 + \frac{R_2R_3}{R_4}}{R_1}\right) \quad \text{A–23}$$

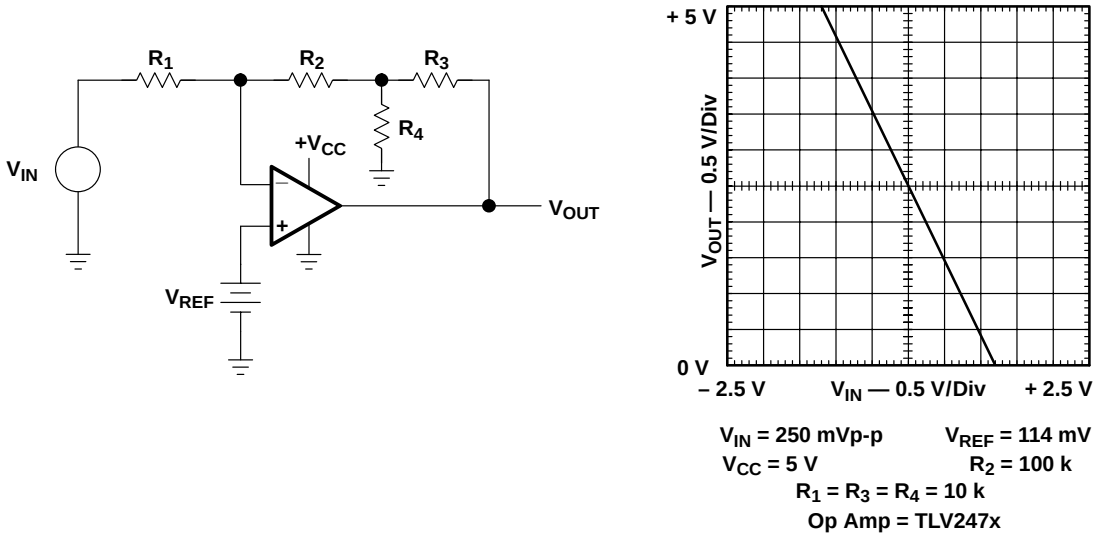


Figure A–20. T Network in the Feedback Loop

A.3.17 Buffer

The buffer input signal polarity must be unipolar because the output voltage swing is unipolar. When this limitation precludes the buffer, a differential amplifier with the negative input correctly biased is used, or a reference voltage is added to the buffer to offset the output voltage. R_F must be included when the op amp inputs are not rated for the full supply voltage. In that case, R_F limits the current into the op amp inputs, thus preventing latch up. Most new op amp inputs can withstand the full supply voltage, so they often leave R_F out as cost savings. The main attraction of the buffer is that it has very high input impedance and very low output impedance. The impedance transformation capability is why buffers are often added to the input of other circuits.

$$V_{OUT} = V_{IN} + V_{REF} \qquad \text{A–24}$$

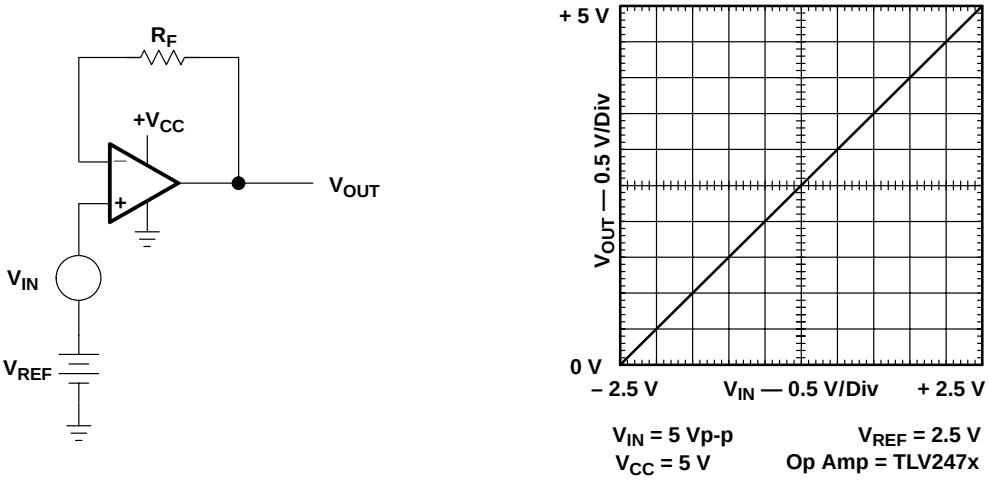


Figure A–21. Buffer

A.3.18 Inverting AC Amplifier

V_{CC} and resistors R set a dc level of $V_{CC}/2$ at the inverting input. R_G is connected to ground through a capacitor, thus the circuit functions as a buffer for dc. This causes the dc output voltage to be $V_{CC}/2$, so the quiescent output voltage is the middle of the supply voltage, and it is ready to swing to either rail as the input signal commands.

The ac gain is given in Equation A–25. R_G and C form a coupling network for the ac signal. Good coupling networks should be constant low impedance at the signal frequencies, so Equation A–26 should be satisfied to get good low-frequency performance. The lowest frequency component of the input signal, f_{MIN}, is determined by completing a Fourier series on the input signal. Then, setting $f_{MIN} = 100f$ in Equation A–26 ensures that the 3-dB breakpoint introduced by R_G and C is two decades lower than f_{MAX}.

$$V_{OUT} = -V_{IN}\frac{R_F}{R_G} + \frac{V_{CC}}{2} \qquad \text{A–25}$$

$$f = \frac{1}{200\pi R_G C} \qquad \text{A–26}$$

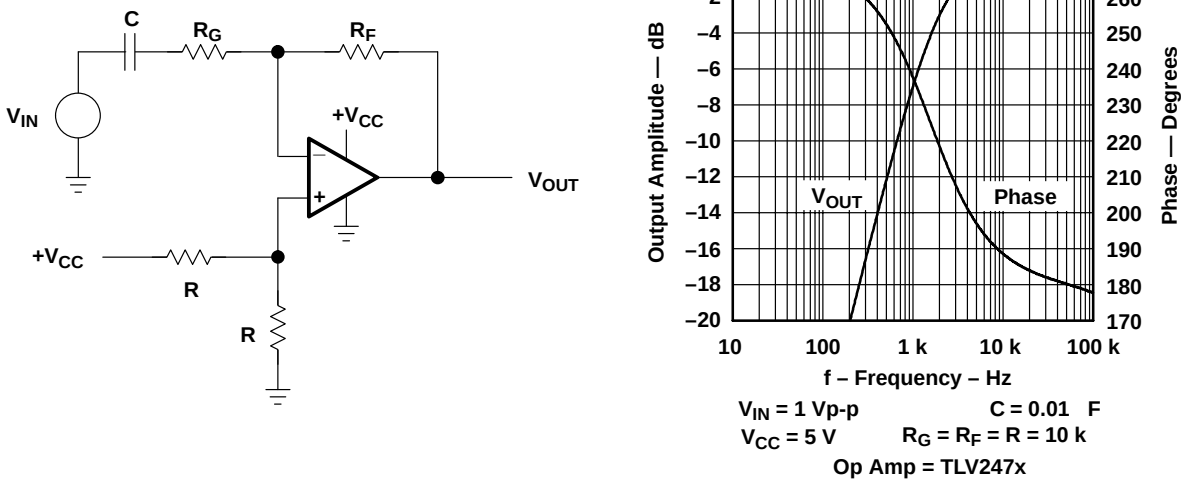


Figure A–22. Inverting AC Amplifier

A.3.19 Noninverting AC Amplifier

V_{CC} and the resistors (R) set a dc level of $V_{CC}/2$ at the inverting input. R_G is connected to ground through a capacitor, thus the circuit functions as a buffer for dc. This causes the dc output voltage to be $V_{CC}/2$, so the quiescent output voltage is the middle of the supply voltage, and it is ready to swing to either rail as the input signal commands.

The ac gain is given in Equation A–27. R_G and C create a coupling network for the ac signal. Good coupling networks should be a constant low impedance at the signal frequencies, so Equation A–28 should be satisfied to get good low frequency performance. The lowest frequency component of the input signal, f_{MIN}, is determined by completing a Fourier series on the input signal. Then, setting $f_{MIN} = 100f$ in Equation A–28 ensures that the 3-dB breakpoint introduced by R_G and C is two decades lower than f_{MAX}. The breakpoint for R_G and C_1 is set in a similar manner.

$$V_{OUT} = V_{IN}\frac{R_F + R_G}{R_G} + \frac{V_{CC}}{2}\left(\frac{R_F + R_G}{R_G}\right) \qquad \text{A–27}$$

$$f = \frac{1}{200\pi R_G C} \qquad \text{A–28}$$

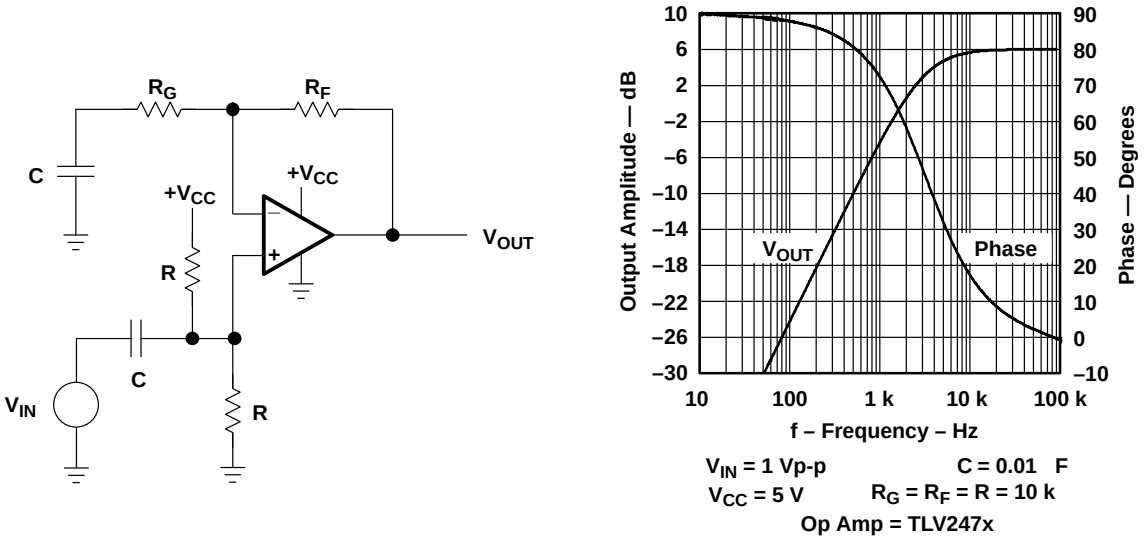


Figure A–23. Noninverting AC Amplifier

A.4 Computing Circuits

Four versions of the inverting op amp and four versions of the noninverting op amp were given in the previous section. During the concept stage of the design, one of these eight op amp circuits is selected. Specifications for the input and output voltage are the selection criteria that determines which circuit configuration is used.

There are four versions of most of the circuits given in this and following sections, but just the simplest version of any circuit is included in this appendix because of space limitations. Each circuit configuration can be modified as required to fit specific applications. Look back to the first section to determine what bias is required to fit the application, and adapt that bias to the new circuit.

A.4.1 Inverting Summer

The three input voltages are inverted and added as Equation A–29 shows. R_B should be made equal in value to the parallel combination of R_F, R_{G1}, R_{G2}, and R_{G3} to convert the input bias current to a common-mode voltage so the op amp can reject it. V_{REF} sets the output voltage somewhere between the supply limits, and this allows negative addition (subtraction) to take place.

$$V_{OUT} = -\left(V_{IN1}\frac{R_F}{R_{G1}} + V_{IN2}\frac{R_F}{R_{G2}} + V_{IN3}\frac{R_F}{R_{G3}} + \ldots\right) + V_{REF}\left(1 + \frac{R_F}{R_{G1} \parallel R_{G2} \parallel R_{G3} \ldots}\right) \quad \text{A–29}$$

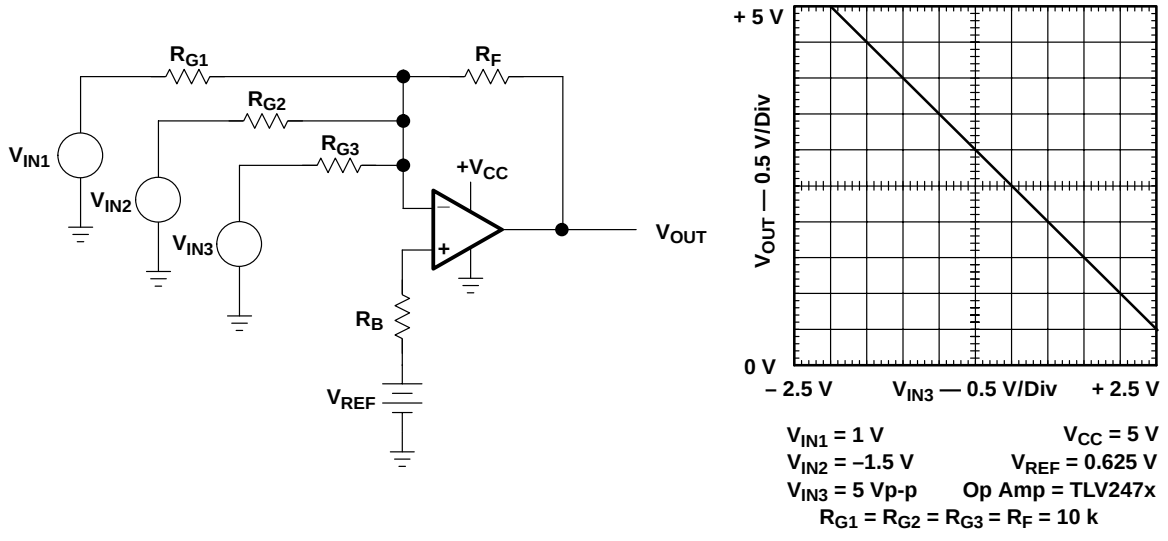


Figure A–24. Inverting Summer

A.4.2 Noninverting Summer

This circuit adds the input voltages and multiplies them by the stage gain. R_{G1}, R_{G2}, and R_{G3} in parallel should be equal to R_F in parallel with R_G to cancel the input bias current using the common-mode input voltage rejection technique. V_{REF} is added to the circuit to enable the addition of negative values.

$$V_{OUT} = + \left(\frac{V_{IN1}\left(R_{G2} \parallel R_{G3}\right)}{R_{G1} + R_{G2} \parallel R_{G3}} + \frac{V_{IN2}\left(R_{G1} \parallel R_{G3}\right)}{R_{G2} + R_{G1} \parallel R_{G3}} + \frac{V_{IN3}\left(R_{G2} \parallel R_{G1}\right)}{R_{G3} + R_{G2} \parallel R_{G1}} \right) \times \left(\frac{R_F + R_G}{R_G} \right) - V_{REF} \frac{R_F}{R_G} \qquad \text{A–30}$$

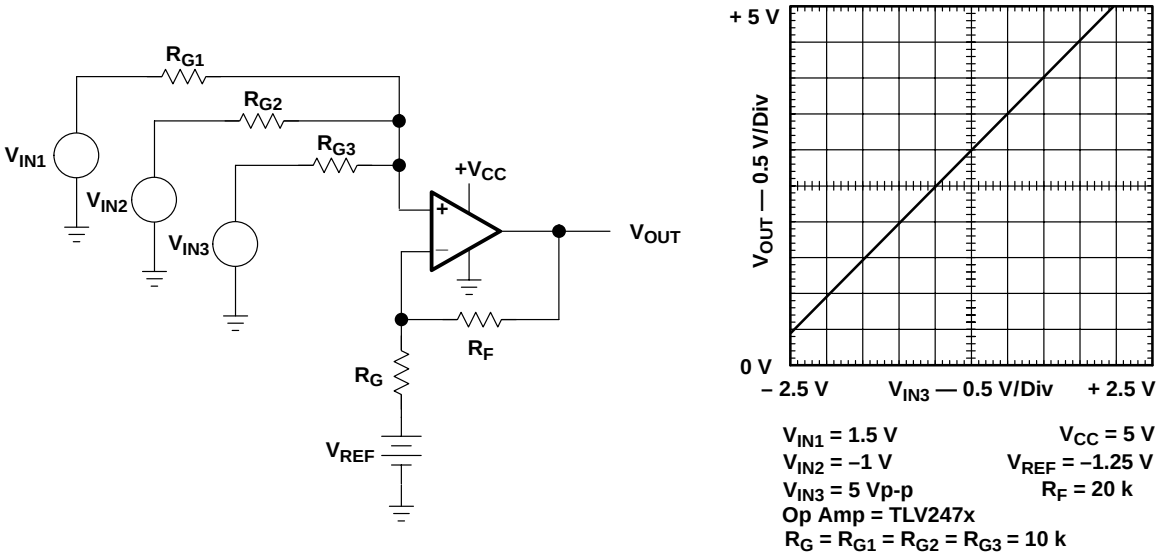


Figure A–25. Noninverting Summer

A.4.3 Noninverting Summer with Buffers

V_{REF1} and V_{REF2} are added to enable the buffers to handle positive input voltages. Their output contribution to the last stage is cancelled out by V_{REF3}. This configuration uses fewer resistors at the expense of two op amps. R_{G1}, R_{G2}, and R_F in parallel should be made equal to R_B to cancel the input bias current.

$$V_{OUT} = V_{IN1}\frac{R_F}{R_{G1}} + V_{IN2}\frac{R_F}{R_{G2}} - V_{REF1}\frac{2R_F}{R_{G1}} - V_{REF2}\frac{2R_F}{R_{G2}} + V_{REF3}\left(1 + \frac{R_F}{R_{G1} \parallel R_{G2}}\right) \quad \text{A–31}$$

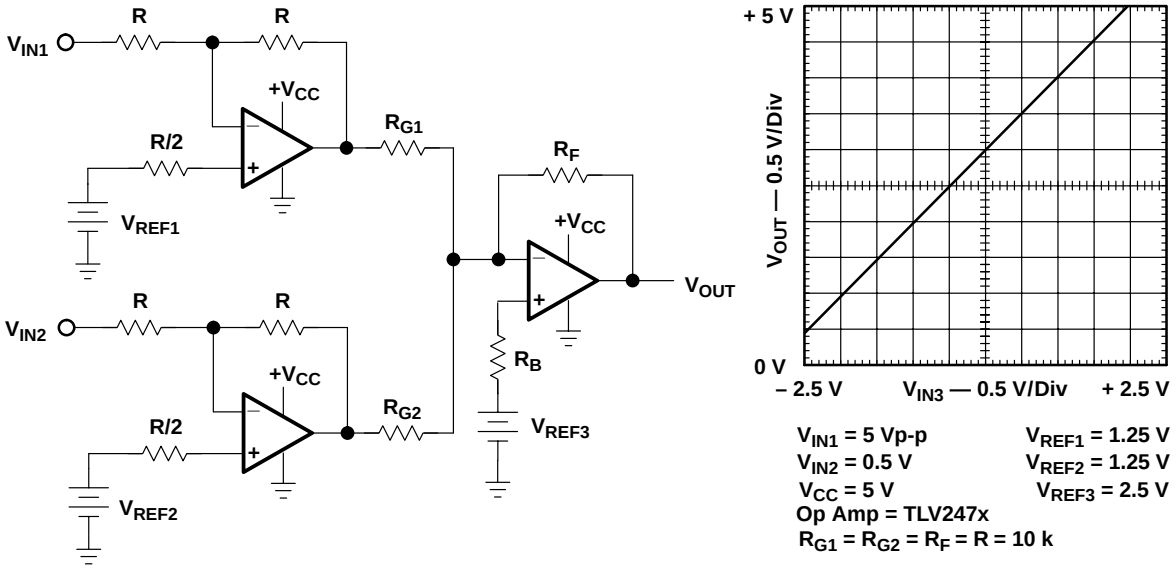


Figure A–26. Noninverting Summer with Buffers

A.4.4 Inverting Integrator

The Laplace operator, s= jω, is used in Equation A–32, and the mathematical operation 1/s constitutes an integration. Differentiation circuits are shown later, and the mathematical operation, s, constitutes a differentiation. The integration time constant is RC, thus the magnitude crosses 0 dB on a log plot when RC = 1. Also the phase is –45° when RC = 1.

$$V_{OUT} = - V_{IN} \frac{1}{RCs} \qquad \text{A–32}$$

This integrator is not very practical because there is no method of discharging the capacitor; hence, any leakage current will eventually charge the capacitor until the circuit becomes saturated. The positive input of the integrator is biased at V_{CC} / 2 to center the output voltage at V_{CC} / 2; thus allowing for positive and negative voltage swings. The bias resistors are selected as 2R so that the parallel combination equals R. This offsets the input current drawn through R.

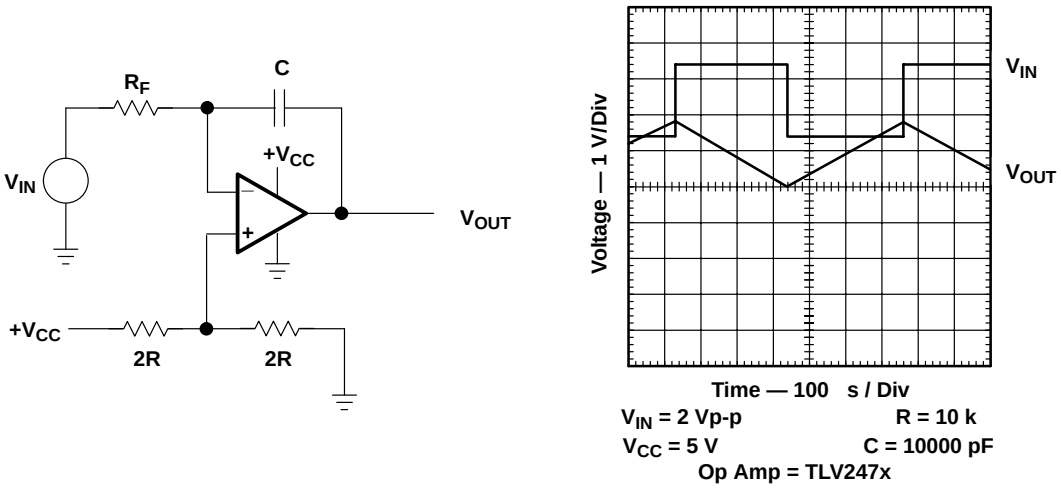


Figure A–27. Inverting Integrator

A.4.5 Inverting Integrator with Input Current Compensation

Functionally, this circuit is the same as that shown in Figure A–27, but a current compensation network has been added to offset the input current. V_{CC}, R_1, and R_2 bias the positive input at V_{CC} / 2 to center the output voltage at V_{CC} / 2; thus allowing for positive and negative voltage swings.

R_1 and R_2 are selected as relatively small values because the current flowing through R_A also flows through the parallel combination of R_1 and R_2. R_A forward biases the diode with a constant current, thus the diode acts like a small voltage regulator. The diode voltage drop is temperature sensitive, and this factor works in our favor because the input transistors are temperature sensitive. The two temperature sensitivities cancel out if the diode current is selected correctly. R_B is a large-value resistor that acts like a current source, so it is selected such that it supplies the input bias current. Selecting R_B correctly ensures that no input current flows through the integration resistor, R.

This integrator is not very practical because there is no method of discharging the capacitor. Hence, any input current will eventually charge the capacitor until the circuit becomes saturated. The bias circuit drastically reduces the input current flowing through R, thus it extends the integration time. A reset circuit is needed to make the integrator more practical.

This bias compensation scheme is set up for an op amp that has NPN input transistors. The diode must be reversed and connected to ground for op amps with PNP input circuits.

$$V_{OUT} = -V_{IN1}\frac{1}{RCs} \qquad \text{A–33}$$

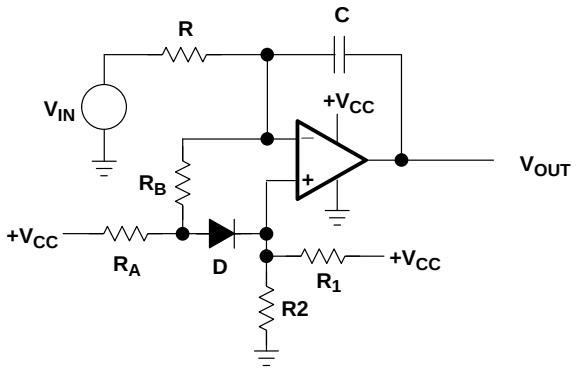


Figure A–28. Inverting Integrator with Input Current Compensation

A.4.6 Inverting Integrator with Drift Compensation

Functionally, this circuit is the same as that shown in Figure A–27, but it uses an RC circuit in the positive lead to obtain drift compensation. The voltage divider is made from a series string of resistors (R_A), and V_{CC} biases the input in the center of the power supply.

Positive input current flows through R and C in parallel, so the positive input current drops the same voltage across the parallel RC combination as the negative input current drops across its series RC combination. The common-mode rejection capability of the op amp rejects the voltages caused by the input currents. Much longer integration times can be achieved with this circuit, but when the input signal does not center around $V_{CC}/2$, the compensation is poor.

$$V_{OUT} = - V_{IN1} \frac{1}{RCs} \qquad \text{A–34}$$

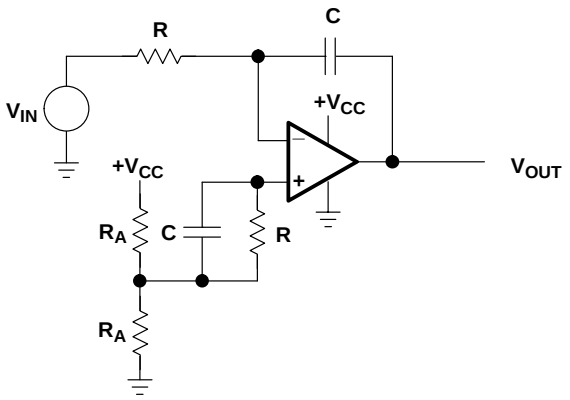


Figure A–29. Inverting Integrator with Drift Compensation

A.4.7 Inverting Integrator with Mechanical Reset

Functionally, this circuit is the same as that shown in Figure A–27, but a method has been provided to discharge (reset) the capacitor. S_1 is a mechanical switch or relay and when the contacts close, they short the integrating capacitor forcing it to discharge. Some capacitors are sensitive to fast discharge cycles, so R_S is put in the discharge path to limit the initial discharge current. When R_S is absent from the circuit, the impulse of current that occurs at the first instant of discharge causes considerable noise, so the selection of R_S is also based on noise considerations. For all practical purposes, the time constant formed by R_S and C determines the discharge rate.

One advantage of mechanical discharge methods is that they are isolated from the remainder of the circuit. Their size, weight, time delay, and uncertain actuating time offset this advantage. When the disadvantages of mechanical reset outweigh the advantages, circuit designers go to electronic reset circuits.

$$V_{OUT} = - V_{IN1} \frac{1}{RCs} \qquad \text{A–35}$$

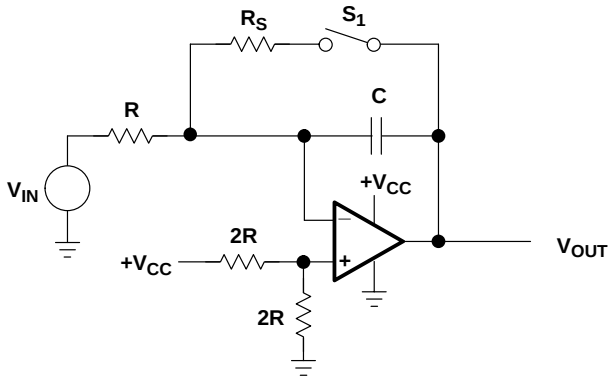


Figure A–30. Inverting Integrator with Mechanical Reset

A.4.8 Inverting Integrator with Electronic Reset

Functionally, this circuit is the same as that shown in Figure A–27, but an electronic method has been provided to discharge (reset) the capacitor. Q_1 is controlled by a gate drive signal that changes its state from on to off. When Q_1 is on, the gate-source resistance is low, less than 100 Ω. And when Q_1 is off, the gate-source resistance is high — about several hundred MΩ.

The source of the FET is at the inverting lead that is at ground, so the Q_1 gate-source bias is not affected by the input signal. Sometimes, the output signal can get large enough to cause leakage currents in Q_1, so the designer must take care to bias Q_1 correctly. Consult a transistor book for more detailed information on transistor reset circuits. A major problem with electronic reset is the charge injected through the transistor's stray capacitance. This charge can be large enough to cause integration errors.

$$V_{OUT} = -V_{IN1}\frac{1}{RCs} \qquad \text{A–36}$$

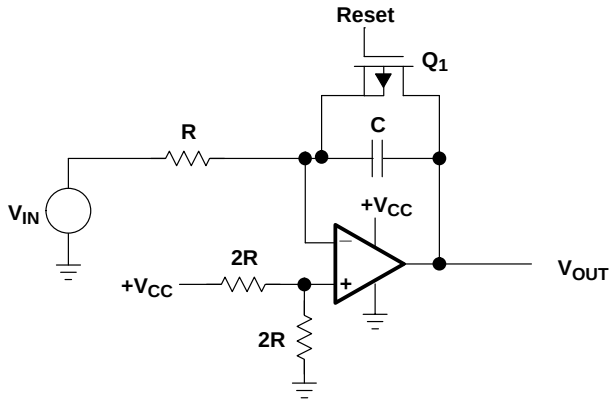


Figure A–31. Inverting Integrator with Electronic Reset

A.4.9 Inverting Integrator with Resistive Reset

This circuit differs from that shown in Figure A–27 because it yields a breakpoint rather than a pure integration. On a log plot, the integrator slope is –6 dB per octave at the 0 frequency intercept, and the 0 dB intercept occurs when f = 1/2πRC. A breakpoint plots flat on a log plot until the breakpoint where it breaks down at –6 dB per octave. It is –3 dB when f = 1/2πRC.

R_F is in parallel with the integrating capacitor, C, so it is continually discharging C. The low frequency attenuation that is the best attribute of the pure integrator is sacrificed for the reset circuit complexity

$$V_{OUT} = -V_{IN1}\left(\frac{R_F}{R_G}\right)\frac{1}{R_F C + 1} \qquad \text{A–37}$$

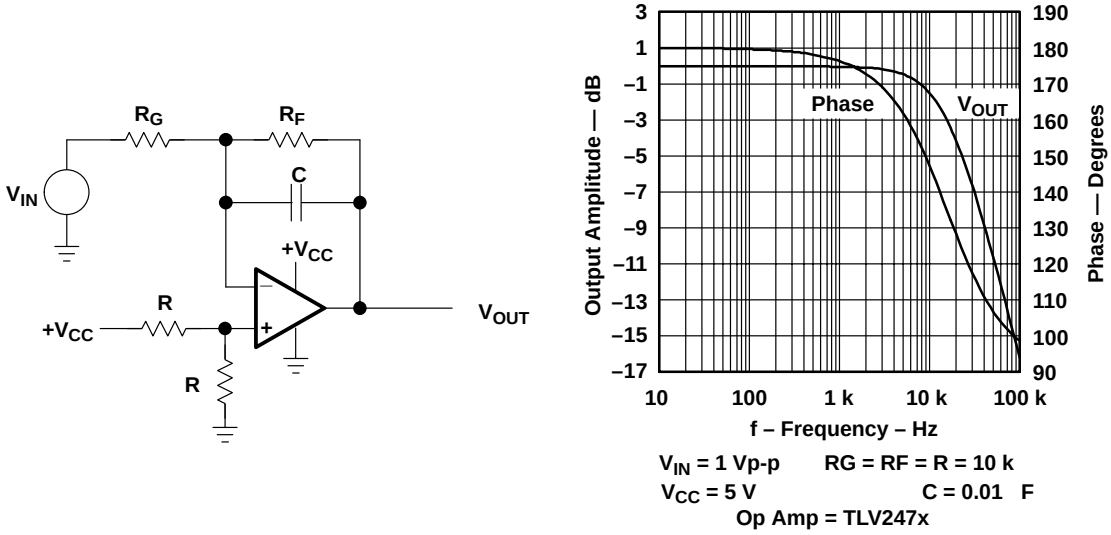


Figure A–32. Inverting Integrator with Resistive Reset

A.4.10 Noninverting Integrator with Inverting Buffer

This circuit is an inverting integrator preceded by an inverting buffer. Eliminating the signal inversion costs an op amp and four resistors, but this is the easiest way to get true noninverting integrator performance.

$$V_{OUT} = \frac{R_A}{R_A}\left(-V_{IN}\,\frac{1}{RCs}\right) = V_{IN}\,\frac{1}{RCs} \qquad \text{A–38}$$

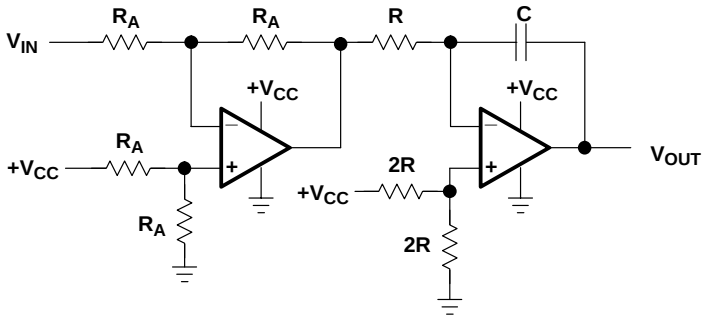


Figure A–33. Noninverting Integrator with Inverting Buffer

A.4.11 Noninverting Integrator Approximation

This circuit has fewer parts than the *Noninverting Integrator With Inverting Buffer* (Figure A–33), but it is not a true integrator because there is a zero in the transfer equation. The log plot starts rolling off at a –6 dB per octave rate at low frequencies, but when f = 1/2πRC, the zero cuts in. The zero causes the log plot to flatten out because the slope decreases to 0 db per decade.

This circuit functions as an integrator at very low frequencies, but at frequencies higher than f = 1/2πRC, it functions as a buffer.

$$V_{OUT} = \frac{RCs + 1}{RCs} \qquad \text{A–39}$$

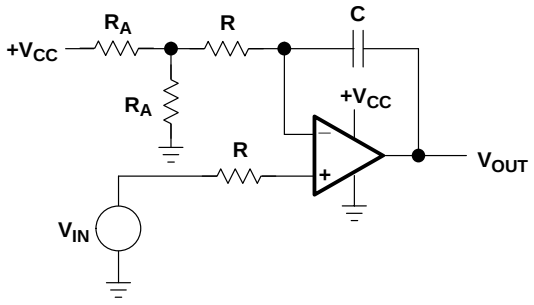


Figure A–34. Noninverting Integrator Approximation

A.4.12 Inverting Differentiator

The log plot of the differentiator is a positive slope of 6-dB per octave passing through 0 dB at f = 1/2πRC. At extremely high frequencies, the capacitive reactance goes to very low values, thus the circuit gain approaches the op amp open-loop gain. This performance emphasizes any system noise or noise generated by the op amp. The poor noise performance of this circuit limits its application to a very few specialized situations.

This configuration has a pole in the feedback loop. If the op amp has more than one pole, and most op amps have several poles, this configuration can become oscillatory. The V_{CC} and R_A circuit bias the output in the center of the power supplies. $R_A/2$ should be selected equal to $R_G||R_F$ so that input currents are canceled out.

$$V_{OUT} = - V_{IN} RCs \qquad \text{A–40}$$

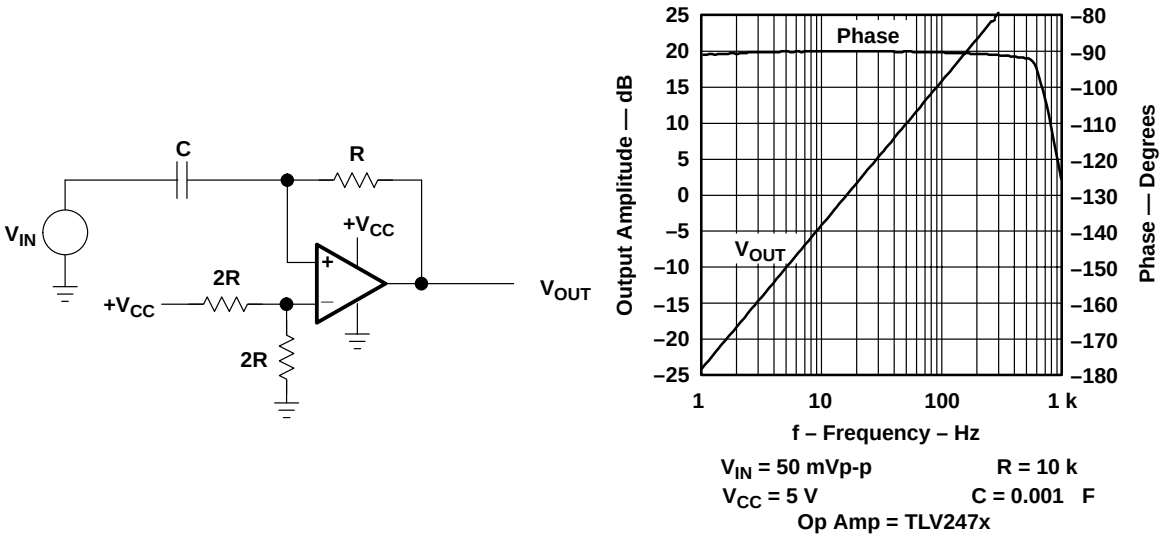


Figure A–35. Inverting Differentiator

A.4.13 Inverting Differentiator with Noise Filter

This circuit has a pure differentiator that rises at a 6-dB per octave slope from zero frequency. At $f = 1/2\pi R_F C_F$, the pole kicks in and the slope is reduced to zero. The pole has two effects. First, it stabilizes the circuit by canceling zero's phase shift. Second, it limits the circuit gain to 1 at high frequencies, so it acts like a noise filter.

R/2 should equal R_F for good input current cancellation, and V_{CC} coupled with R centers the output voltage.

$$V_{OUT} = - V_{IN} \frac{R_F C s}{R_F C_F s + 1} \qquad \text{A–41}$$

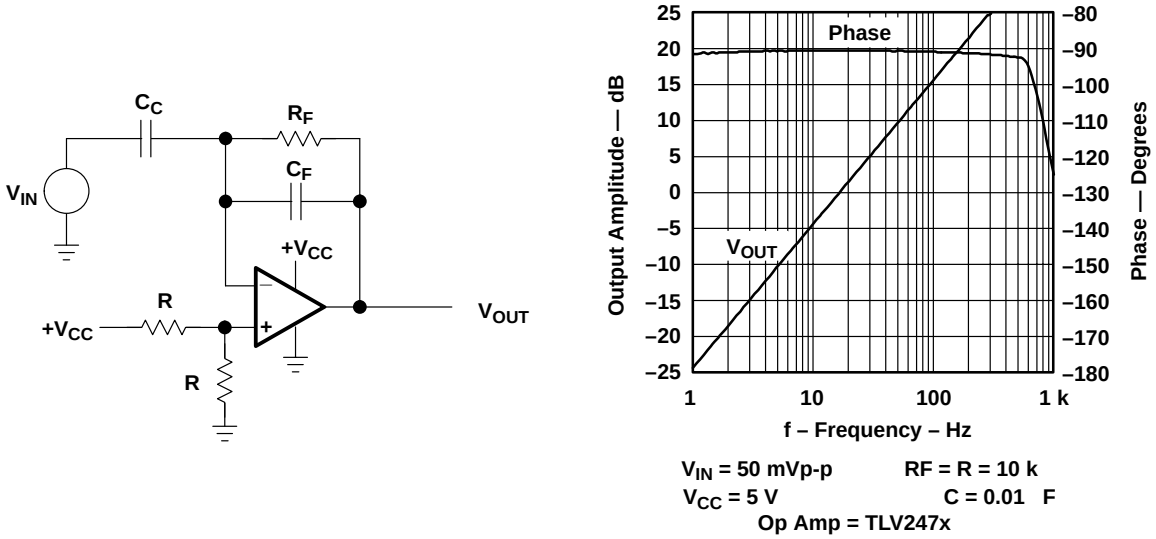


Figure A–36. Inverting Differentiator with Noise Filter

A.5 Oscillators

This section describes some general op amp sinewave oscillator circuits that fall under three main categories: Wien bridge, phase shift, and quadrature. A brief description and of each type is provided, along with one or two variations. Op amp sinewave oscillators are used to create references in applications such as audio and function/waveform generators.

A.5.1 Basic Wien Bridge Oscillator

When $\omega = 2\pi f = 1/RC$, the feedback is in phase (this is positive feedback), and the gain is 1/3, so oscillation requires an amplifier with a gain of 3. When $R_F = 2R_G$ the amplifier gain is 3 and oscillation occurs at $f = 1/2\pi RC$. Normally, the gain is larger than 3 to ensure oscillation under worst case conditions.

V_{REF} sets the output dc voltage in the center of the span.

The output sine wave is highly distorted because limiting by saturation and cutoff is controlling the output voltage excursion. The distortion decreases when the gain is decreased, but the circuit may not oscillate under worst-case low gain conditions.

$$V_{REF} = \frac{\frac{V_{CC}}{2}}{1 + \frac{R_F}{R_G}} \qquad \text{A–42}$$

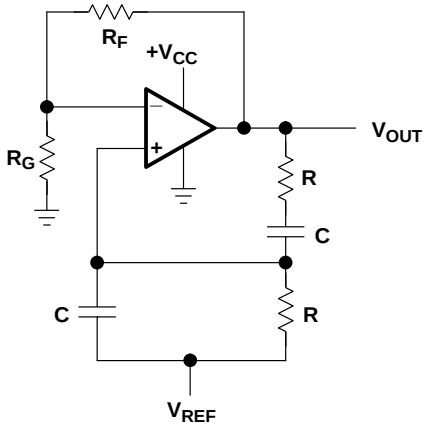


Figure A–37. Basic Wien Bridge Oscillator

A.5.2 Wien Bridge Oscillator with Nonlinear Feedback

When the circuit gain is 3, $R_L = R_F/2$.

Substituting a lamp (R_L) for the gain setting resistor reduces distortion because the non–linear lamp resistance adjusts the gain to keep the output voltage smaller than the power supply voltage. The output voltage never approaches the power supply rail, so distortion doesn't occur. R_F and R_L determine the lamp current (see Equations A–43 and A–44).

$$I_{LAMP} = \frac{V_{OUT(RMS)}}{R_F + R_L} \qquad \text{A–43}$$

$$R_F = \frac{2\left(V_{OUT(RMS)}\right)}{3\left(I_{OUT(RMS)}\right)} \qquad \text{A–44}$$

The lamp is selected by examining lamp resistance curves until a lamp with a resistance approximately equal to $R_F/2$ at $I_{OUT(RMS)}$ is found. The output voltage swing should be less than 75% of the maximum guaranteed voltage swing, and 3 R_L must be greater than the load resistance specified for the voltage swing specification. V_{REF} should be $V_{CC}/5$.

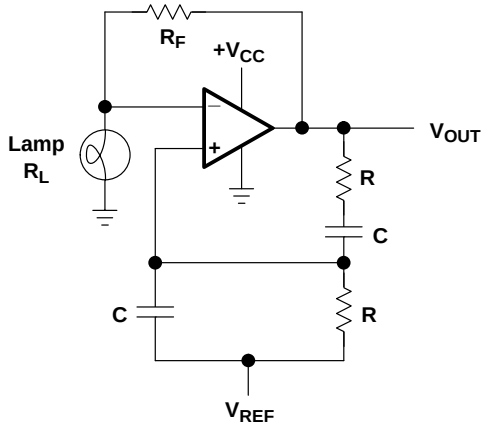


Figure A–38. Wien Bridge Oscillator with Nonlinear Feedback

A.5.3 Wien Bridge Oscillator with AGC

The op amp is configured as an ac amplifier to ease biasing problems. The gain equation for the op amp is given below. R_{G1} or R_{G2}, but not both resistors, is required depending on the selection of the Q_1.

The diode, D_1, half-wave rectifies the output voltage and applies it to the voltage divider formed by R_1 and R_2. The voltage divider biases Q_1 in its linear region, and they eventually set the output voltage. C_1 filters the rectified sine wave with a long time constant so that the output voltage stays constant. C_2 must be selected large enough to act as a short at the oscillation frequency.

As the output voltage increases, the negative voltage across the gate of Q_1 increases. The increased negative gate voltage causes Q_1 to increase its drain-to-source resistance. This results in increased op amp gain and an output voltage decrease. When the voltage divider and FET are selected properly, the output voltage swing is less than the guaranteed maximum swing, so distortion doesn't occur.

$$G = \frac{R_F}{R_{G1} + R_{G2} \parallel R_{FET}} \qquad \text{A–45}$$

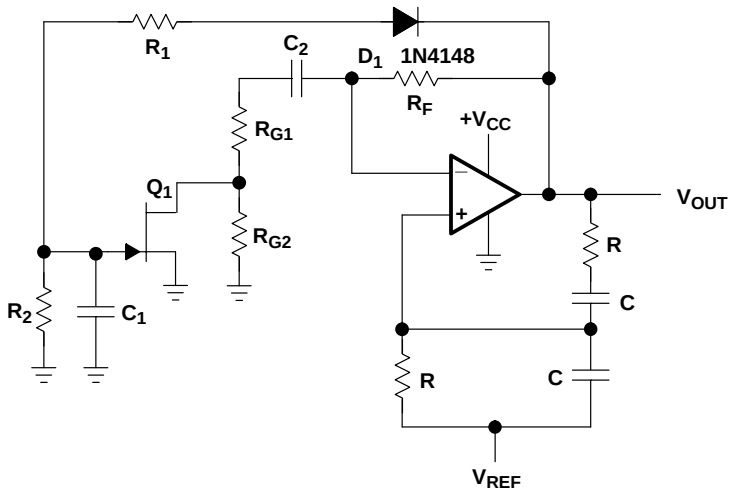


Figure A–39. Wien Bridge Oscillator with AGC

A.5.4 Quadrature Oscillator

Quadrature oscillators produce sine waves 90° out of phase, so they output sine/cosine, or quadrature waves.

When $R_1C_1 = R_2C_2 = R_3C_3$, the circuit oscillates at $\omega = 2\pi f = 1/RC$. Both op amps act as integrators causing two poles at 1/RC, thus the circuit oscillates when the loop gain crosses the 0-dB axis. The integrators ensure that gain is always sufficient for oscillation. There is a slight bit of distortion at the sine output, and it is very hard to eliminate this distortion.

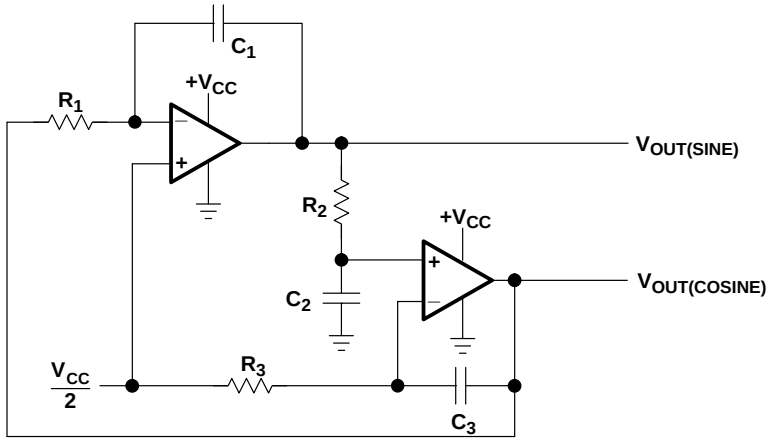


Figure A–40. Quadrature Oscillator

A.5.5 Classical Phase Shift Oscillator

Theoretically, the three RC sections do not load each other, thus the loop gain has three identical poles multiplied by the op amp gain.

The loop phase shift is –180° when the phase shift of each section is –60°, and this occurs when $\omega = 2\pi f = 1.732/RC$ because the tangent of 60° = 1.73. The magnitude of β at this point is $(1/2)^3$, so the gain, $A = R_F/R_G$, must be greater or equal to 8 for the system gain to be equal to 1.

The assumption that the RC sections do not load each other is not entirely valid, thus the circuit does not oscillate at the specified frequency, and the gain required for oscillation is more than 8. This circuit configuration was very popular when an active component was large and expensive, but now that op amps are inexpensive, small, and come quad packages, the classical phase shift oscillator is losing popularity.

The classical phase shift oscillator has an undistorted sine wave available at the output of the third RC section. This is not a low-impedance output, and the signal amplitude is smallest here, but these sacrifices have to be made to get away from distortion. An undistorted output can be obtained from the op amp if an AGC circuit similar to the one shown in Figure A–39 is employed. The reference voltage is set according to the equation $V_{REF} = V_{CC}/ 2(1+R_F/R_G)$ to center the output voltage at $V_{CC}/2$.

$$A\beta = \left(\frac{1}{RCs + 1}\right)^3 \qquad \text{A–46}$$

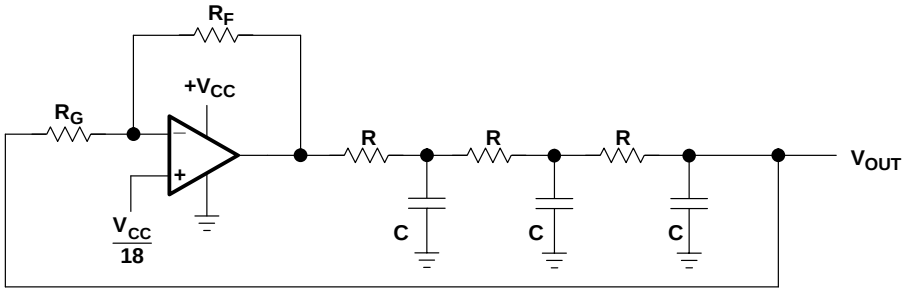


Figure A–41. Classical Phase Shift Oscillator

A.5.6 Buffered Phase Shift Oscillator

A noninverting op amp buffers each RC section in this oscillator. Equation A–46, repeated below, truly represents the transfer function of this circuit if $R_G >> R$.

$$A\beta = \left(\frac{1}{RCs + 1}\right)^3 \qquad \text{A–47}$$

The loop phase shift is –180° when the phase shift of each section is –60°, and this occurs when $\omega = 2\pi f = 1.732/RC$ because the tangent 60° = 1.73. The magnitude of β at this point is $(1/2)^3$, so the gain, $A = R_F/R_G$, must be greater or equal to 8 for the system gain to be equal to one.

The buffered phase shift oscillator has an undistorted sine wave available at the output of the third RC section. This is not a low-impedance output, and the signal amplitude is smallest here, but these sacrifices have to be made to get away from distortion. An undistorted output can be obtained from the op amp if an AGC circuit similar to the one shown in Figure A–39 is employed.

There are three op amps, so the gain can be distributed among the op amps at the expense of a few resistors, and the distortion is reduced. Another method of reducing distortion is to limit the output voltage swing softly with external components. The limiting technique does not yield as good results as the AGC technique does, but it is less expensive. The reference voltage is set according to the equation $V_{REF} = V_{CC}/ 2(1+R_F/R_G)$ to center the output voltage at $V_{CC}/2$.

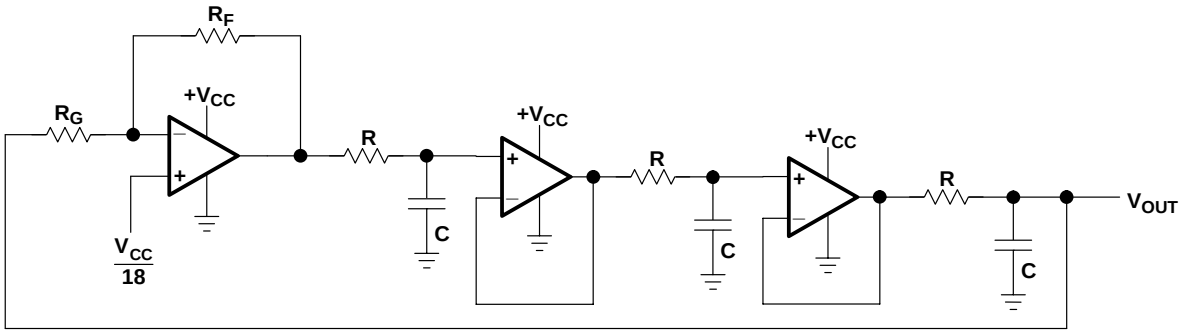


Figure A–42. Buffered Phase Shift Oscillator

A.5.7 Bubba Oscillator

The Bubba oscillator is another phase shift oscillator, but it takes advantage of the quad op amp package to yield some unique advantages. Each RC section is buffered by an op amp to prevent loading. When R_G >> R there is no loading in the circuit, and the circuit yields theoretical performance.

Four RC sections require –45° phase shift per section to accumulate –180° phase shift. Each RC section contributes –45° phase shift when $\omega = 1/RC$. The gain required for oscillation is $G \geq (1/0.707)^4 = 4$. Taking outputs from alternate sections yields low-impedance quadrature outputs. When an output is taken from each op amp, the circuit delivers four 45° phase-shifted sine waves.

The gain, A, must equal 4 for oscillation to occur. Very low distortion sine waves can be obtained from the junction of R and R_G. When low-distortion sine waves are required at all outputs, the gain should be distributed among the op amps. Gain distribution requires biasing of the other op amps, but it has no effect on the oscillator frequency. This oscillator has the best $d\phi/df$ of the phase shift oscillators, so it has minimum frequency drift. The reference voltage is set according to the equation $V_{REF} = V_{CC}/ 2(1+R_F/R_G)$ to center the output voltage at $V_{CC}/2$.

$$A\beta = \left(\frac{1}{RCs + 1}\right)^4 \qquad \text{A–48}$$

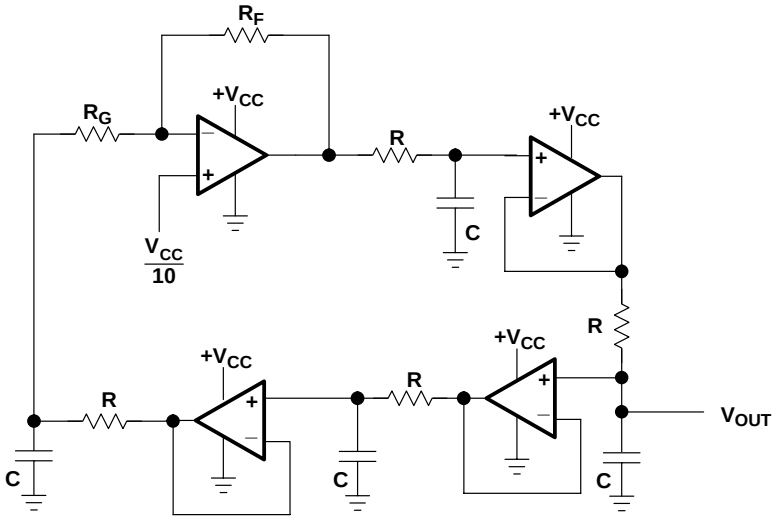


Figure A–43. Bubba Oscillator

A.5.8 Triangle Oscillator

The triangle oscillator produces triangle waves and square waves. The op amp functions as an integrator. When the output voltage of the comparator is low, the output of the op amp charges C until the output voltage exceeds the hysteresis voltage set by R_1 and R_F and the reference voltage ($V_{CC}/2$). At this point, the comparator output switches to a high state and the op amp integrates the voltage in a negative direction. The triangle wave (op amp output voltage swing) is given in Equation A–49. The frequency of oscillation is given in Equation A–50.

$$V_{OUT} = \frac{V_{CC}}{2} \pm \frac{V_{CC}R_1}{2R_F} \qquad \text{A–49}$$

$$f = \frac{R_F}{4CRR_1} \qquad \text{A–50}$$

The op amp reference voltage can be adjusted to equalize the triangle rise and fall times.

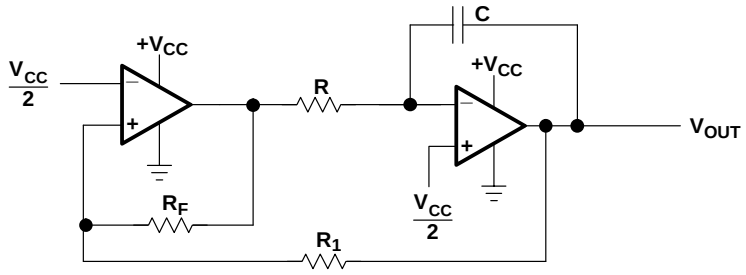


Figure A–44. Triangle Oscillator

Appendix B

Single-Supply Op Amp Selection Guide

This appendix contains a selection guide for single-supply op amps and abstracts from the data sheets for the TLC080, TLV2401, TLV2470, and TLV2760 single-supply op amps.

Table B–1. Single-Supply Operational Amplifier Selection Guide

DEVICE FAMILY†	AMPS in PKG‡	V_{DD} (V)		I_{DD} (μA/Ch)		V_{IO} (μV)		I_{IB} (pA)	GBW (MHz)	SR (V/μs)	CMRR (dB)	$V_{n, 1kHz}$ (nV/√Hz)
		min	max	typ	max	typ	max	typ	typ	typ	typ	typ
V_{DD} (min) ≥ 1.8 V												**V_{DD} (min) ≥ 1.8 V**
TLV278x(A)	S/D/Q	1.8	3.6	650	770	250	3000	2.5	8	2.8	100	18
TLV276x	S*/D/Q*	1.8	3.6	20	28	550	3500	3	0.5	0.2	70	95
V_{DD} (min) ≥ 2.5 V												**V_{DD} (min) ≥ 2.5 V**
TLV277x(A)	S/D/Q	2.5	5.5	1000	2000	360	2500	2	5.1	10.5	96	17
TLV411x	S*/D	2.5	6	400	600	175	3500	0.3	2	1.57	63	26
TLV240x	S/D/Q	2.5	16	0.88	0.95	390	1200	100	0.0055	0.0025	120	n/a
TLV224x	S/D/Q	2.5	12	1	1.2	600	3000	100	0.0055	0.002	100	n/a
TLV246x(A)	S/D/Q	2.7	6	500	575	150	2000	1300	5.2	1.6	80	11
TLV247x(A)	S/D/Q	2.7	6	550	750	250	2200	2	2.8	1.4	78	15
TLV2231	S	2.7	10	750	1200	750	3000	1	1.9	1.25	70	16
TLV2731	S	2.7	10	750	1200	750	3000	1	1.9	1.25	70	16
TLV244x(A)	D/Q	2.7	10	725	1100	300	2000	1	1.75	1.3	75	18
LMV321#	S/D/Q	2.7	5.5	130	250	1700	7000	11000	1	1	65	39
TLV226x(A)	D/Q	2.7	8	200	250	300	2500	1	0.67	0.55	75	12
TLV243x(A)	D/Q	2.7	10	98	125	300	2000	1	0.5	0.25	83	22
TLV2221	S	2.7	10	100	150	610	3000	1	0.48	0.18	82	20
TLV2721	S	2.7	10	100	150	610	3000	1	0.48	0.18	82	20
TLV245x(A)	S/D/Q	2.7	6	23	35	20	1500	900	0.2	0.11	86	49
TLV225x(A)	D/Q	2.7	8	34	62	200	1500	1	0.187	0.1	75	19
TLV2211	S	2.7	10	11	25	470	3000	1	0.056	0.025	83	22
TLV2711	S	2.7	10	11	25	470	3000	1	0.056	0.025	83	22
TLV2422(A)	D	2.7	10	50	75	300	2000	0.5	0.046	0.02	83	23
V_{DD} (min) ≥ 3.8 V												**V_{DD} (min) ≥ 3.8 V**
TLC2652(A)	S	3.8	16	1500	2400	0.6	3	4	1.9	2.8	140	23
TLC2654(A)	S	3.8	16	1500	2400	5	20	50	1.9	2	125	13
TLC450x(A)	S/D	4	6	1000	1500	–	80	1	4.7	2.5	100	12
TLC227x(A)	D/Q	4.4	16	1100	1500	300	2500	1	2.18	3.6	75	9
TLC226x(A)	D/Q	4.4	16	200	250	300	2500	1	0.82	0.55	83	12
TLC225x(A)	D/Q	4.4	16	40	62	200	1500	1	0.21	0.12	88	19
TLC07x(A)	S/D/Q	4.5	16	1900	2500	60	1000	1.5	10	16	140	7
TLC08x(A)	S/D/Q	4.5	16	1800	2500	60	1000	3	10	16	140	8.5
TLC220x(A)	S/D	4.6	16	1000	1500	80	500	1	1.9	2.5	110	8

Note: Single-supply op amp selection guide parameters specified at T_A = 25°C, V_{DD} = 5 V, 2.7 V, and/or 1.8 V.

* Devices in Product Preview stage of development at press time. Please contact TI or visit www.ti.com for current status.

† (A) refers to precision grades with lower V_{IO} and are part of the device number, e.g. TLV2470A.

‡ S = single-channel amplifier, D = dual-channel amplifier, Q = quad-channel amplifier (in the device name, replace x with 1, 2, or 4 for S, D, or Q respectively).

§ Variants of family with shutdown (in the device name, replace x with 0, 3, or 5 for single, dual, or quad amplifiers with shutdown).

¶ Measured at V_O = 0.5 V from rails, except TLC07x and TLC08x where V_O = 0.5 V from negative rail.

Includes LMV358 and LMV324 devices.

Table B–1. Single-Supply Operational Amplifier Selection Guide (Continued)

I_O¶ (mA) typ	SHUT DOWN §	RAIL to RAIL	FEATURES
V_{DD} (min) ≥ 1.8 V			
10	Y	I/O	WBW, LV
4.8	Y	I/O	µP, LV
V_{DD} (min) ≥ 2.5 V			
6	Y	O	VHSR, HD
220	Y	O	VHD, SD
0.2		I/O	nP, LV
0.2		I/O	µP, LV
30	Y	I/O	WBW, LN, VHD
22	Y	I/O	WBW, LN, HD, LIBC
2.5		O	WBW, LN
2.5		O	WBW
2.5		O	LV, 600Ω
–		O	SS
1.1		O	LP, LN
2.5		O	LP
1.8		O	LP, SOT
1.8		O	LP, SOT
4	Y	I/O	µP, R-R I/O
0.25		O	µP, R-R O
0.38		O	µP, WVSR
0.38		O	µP, WVSR
2.5		O	LP, 600Ω
V_{DD} (min) ≥ 3.8 V			
1.4		O	PC
1.4		O	PC, LN
22		O	WBW, P, LN, HD
2		O	WBW, LN
1.3		O	LP, LN
0.3		O	µP
55	Y		WBW, FSR, LN
55	Y		WBW, FSR
1		O	P, LN

Single-Supply Operational Amplifier Selection Guide Features Key:

FSR	fast slew rate	nP	nanopower
HD	high drive	P	precision
LIBC	low input bias current	PC	precision (chopper)
LN	low noise	SOT	SOT–23 package
LP	low power	SS	second sourced
µP	micropower	WBW	wide bandwith

A

AC loads, DAC, 14-2

AC parameters, DAC, 14-14 to 14-18

ACG, *See* Automatic gain control

Active filter, design techniques, 16-1 to 16-61

Actuator, analog interface to DAC, 18-16 to 18-21

Actuator interface
- from DAC current sink circuit, 18-19
- from DAC current source circuit, 18-17

ADC
- analog interface from transducer, 18-13 to 18-14
- characterization, 12-15
- external ref circuits, 13-15 to 13-18
- external voltage reference circuit, 13-17
- input voltage, 12-15
- selection, 13-6 to 13-10
- single-ended input drive circuit, 13-20

ADC interface, from transducer circuit, 18-15, 18-16

Adder, 3-5
- circuit, 3-5

Adjustable low-pass filter, circuit, 16-17

AIA circuit, 12-17

All-pass filter
- design, 16-41 to 16-47
- first order, 16-44
- first-order circuit, 16-44
- high order, 16-45 to 16-47
- second order, 16-44 to 16-46
- second-order circuit, 16-44
- seventh-order circuit, 16-46

Amplifier
- photodiode, 12-8
- phototransistor, 12-9
- transistor, 2-10 to 2-13
- photovoltaic cell, 12-9

Analog input drive circuits, high-speed, 13-18 to 13-22

Analog interface circuit, 12-21

Analog plane placement, 17-8

Antenna, loop and slot PCB layout, 17-15

Anti-aliasing filter, 13-11 to 13-13

Automatic gain control, Wien bridge oscillator circuit, 15-13

Avalanche noise, 10-9 to 10-11

B

Band-pass filter
- Butterworth gain response graph, 16-36
- design, 16-27 to 16-37
- gain response graph, 16-29
- low-pass filter transition, 16-28, 16-37
- Sallen-Key circuit, 16-30
- second order, 16-29 to 16-39

Band-rejection filter, comparison graph active/passive, 16-41

Band-rejection filter
- active twin-T filter, 16-37 to 16-40
- active Wien-Robinson filter, 16-39 to 16-42
- design, 16-36 to 16-42

Bandwidth
- current feedback amp, 9-3 to 9-7
- current feedback amp graph, 8-10
- voltage feedback amp, 9-3 to 9-7

Bessel high-pass filter, circuit, 16-27

Bessel low-pass filter, 16-7 to 16-10

Bias circuit, 12-13

Binary bits, versus significant bits, 12-2

Black, Harry, Bell Labs, 1-1

Block diagram
- CDMA cellular base station, 13-5
- cellular base station receiver, 13-2
- dual IF receiver, 13-4
- feedback system, 6-2
- GSM cellular base station, 13-2
- loop gain, 6-4
- oscillator with positive feedback, 15-8

Block diagram math, 5-1 to 5-6
- control system terms, 5-3
- definition, 5-2
- feedback circuit, 5-3
- multiloop feedback system, 5-4
- summary points, 5-3
- transforms, 5-5

Blocks, block diagram math, 5-2

Bode analysis, feedback, 5-7 to 5-13

Bode plot
- compensated attenuator, 7-18
- current feedback amp, 8-7
- dominant-pole compensation circuit, 7-11
- feedback capacitance, 8-13
- lead compensation, 7-14
- lead-lag compensation, 7-19

Bode, H. W, 1-2

Broadcasting, from PCB traces, 17-9
Bridge, wheatstone, 12-8
Broadband noise, 11-18
Brown noise, 10-12
Bubba oscillator, 15-17 to 15-19
  circuit, 15-17
Burst noise, 10-9 to 10-11
Butterworth band-pass filter, gain response graph, 16-36
Butterworth low-pass filter, 16-6
  amplitude response graph, 16-6
  circuit with unity gain, 16-21
  graph, 16-51
  Tschebyscheff-type characteristic, 16-51

C

Cable amplifier, 3-9
Canonical equations, 6-2 to 6-5
Canonical feedback, 15-2
Capacitance
  effect on inverting op amp, 17-18
  trace to plane, 17-17
Capacitively-loaded
  loop-gain calculation circuit, 7-9
  op amp circuit, 7-9
Capacitor, 3-9 to 3-11
  high-frequency circuit, 17-12
  self resonance, 17-22
Cascading filter stages, 16-12
CDMA cellular base station, transmitter block diagram, 13-5
Cellular base station receiver, block diagram, 13-2
CFA. *See* current feedback amplifier
Characterization, transducer, 12-13 to 12-16
Circuit
  active full-wave rectifier and filter, 12-10
  active Wien-Robinson filter circuit, 16-39
  ADC singled-ended input drive, 13-20
  adder, 3-5
  AIA, 12-17
  analog interface, 12-21
  analysis done the hard way, 2-7
  automatic gain control, 15-13
  band reject filter, 5-9
  Bessel high-pass filter, 16-27
  bubba oscillator, 15-17
Butterworth low-pass filter circuit with unity gain, 16-21
canonical feedback, 5-6
capacitively-loaded op amp, 7-9
compensated attenuator, 7-17
complex feedback networks, 3-7
control canonical feedback, 5-6
current divider rule, 2-4
current feedback amplifier, 8-2
DAC current booster, 14-20
DAC current sink to actuator interface, 18-19
DAC current source to actuator interface, 18-17
DAC deglitch, 14-17
DAC voltage booster, 14-21, 14-22
differential amplifier, 3-6, 3-7
differential op amp, 6-8
dual-supply filter, 16-47
electronic canonical feedback, 5-6
feedback, 5-6
feedback loop broken, 6-7
first-order all-pass filter, 16-44
first-order high-pass filter, 16-23
first-order inverting high-pass filter, 16-23
first-order inverting low-pass filter, 16-13
first-order low-pass filter, 16-12
first-order noninverting low-pass filter, 16-14
first-order passive RC low pass filter, 16-2
fourth-order passive RC low pass filter, 16-3
high-frequency capacitor, 17-12
high-frequency inductor, 17-13
high-pass filter, 3-10
high-performance resistor, 17-11
high-speed analog input drive, 13-18 to 13-22
ideal current feedback amp, 9-3
ideal op amp, 3-2
input current, 11-10
input for an RRI op amp, 18-8
input for nonRRI op amp, 18-8
input offset voltage, 11-9
input offset voltage adjustment, 11-9
input parasitic elements, 11-13
inverting, 6-6, 6-7
inverting current feedback amp, 8-5
inverting op amp, 3-4
Kirchoff's current law, 2-3
Kirchoff's voltage law, 2-2
lead-lag compensation, 7-19
logic gate output, 17-21
long-tailed pair, 9-2
loop gain calculation, 7-9

low pass filter, 16-1
low-pass filter, 3-10, 5-8
low-pass filter circuit with unity gain, 16-19
maximum output voltage swing, 11-12
MFB band-pass filter, 16-31
MFB high-pass filter biasing, 16-50
MFB low-pass filter biasing, 16-49
Miller effect compensation, 7-2
noninverting current feedback amp, 8-4
noninverting op amp, 3-3, 18-7
nonlinear feedback, 15-12
Ohm's law, 2-2
op amp design, 12-16 to 12-23
op amp terminal model, 17-5
op amp with negative feedback, 15-8
op amp with positive feedback, 15-8
original amplifier, 2-5
passive twin-T filter, 16-37
passive Wien-Robinson bridge circuit, 16-39
phase shift oscillator, 15-14
phase shift oscillator, buffered, 15-16
photodiode amplifier, 12-8
phototransistor amplifier, 12-9
photovoltaic cell amplifier, 12-9
quadrature oscillator, 15-19
reference bias, 12-13
RRO output stage, 18-11
Sallen-Key band-pass filter, 16-30
Sallen-Key high-pass filter, 16-24
Sallen-Key high-pass filter biasing, 16-50
Sallen-Key low-pass filter, 16-15
Sallen-Key low-pass filter biasing, 16-48
Sallen-Key low-pass filter circuit, 16-20
saturated transistor, 2-9
second-order all-pass filter, 16-44
second-order low-pass filter, 16-17
second-order MFB high-pass filter, 16-25
second-order MFB low-pass filter, 16-18
second-order Tschebyscheff low-pass filter, 16-16
seventh-order all-pass filter, 16-46
simplified op amp, 11-17
simultaneous equation , 4-12, 4-15, 4-17, 4-20
single supply, 4-1, 4-3
common-mode voltage, 4-2
reference voltage input, 4-2
single supply inverting op amp, 4-4, 4-5
single supply noninverting, 4-7
single-ended to diff output, 13-18
single-pole reconstruction filter, 13-15
single-supply DAC, 14-23
single-supply filter, 16-47
slew rate, 11-16
split-supply op amp, 10-19
stability analysis, 8-2, 8-3
stray capacitance on inverting input, 7-16
superposition example, 2-8
T network, 3-8
Thevenim equivalent of the base circuit, 2-11
Thevenim's equivalent, 2-5, 2-6
TLC2201, 10-21, 10-23
traducer bias, 12-13
transducer to ADC interface, 18-15, 18-16
transistor, 2-10
video amplifier, 3-9
voltage divider rule, 2-3
voltage feedback amp, 9-2
voltage reference filter, 13-16
Wein bridge oscillator, 15-12
Wein bridge oscillator with AGC, 15-13
wheatstone bridge, 12-8
Wien bridge oscillator, 15-9, 15-11
Wien-Robinson bridge circuit, 16-39
Wien-Robinson circuit, 16-39
Circuit board layout, 17-1 to 17-30
component of op amp design, 17-1
noise sources, 17-2
prototype, 17-1 to 17-3
Circuit Theory, op amps, 2-1 to 2-13
Closed loop response, diff amp graph, 13-19
Colors, noise, 10-10 to 10-13
Common-mode range, op amp, 18-6 to 18-10
Common-mode rejection ratio, 11-15
Compensated attenuator, 7-16 to 7-19
circuit, 7-17
Compensation
comparison of schemes, 7-20 to 7-22
DAC, 14-18 to 14-20
dominant pole graph, 7-11
dominant-pole, 7-9 to 7-13
external, 7-8
feedback capacitance, 8-13 to 8-15
gain, 7-12 to 7-14
gain capacitance, 8-13
graph, 7-12, 7-13
internal op amp, 7-2 to 7-9
inverting op amp, 7-15
lead, 7-13 to 7-17
lead-lag, 7-18 to 7-21
Miller effect, 7-2 to 7-4

Miller effect circuit, 7-2
noninverting op amp, 7-16
Complex feedback network, 3-7 to 3-9
Thevenin's theorem, 3-7 to 3-9
Control system terms, block diagram math, 5-3
Corner frequency, 10-13 to 10-15
Current booster
circuit for DAC, 14-20
op amps, 14-20
Current divider rule, 2-4
Current drain, op amp, 18-12 to 18-14
Current feedback amplifier
analysis, 8-1 to 8-15
bandwidth, 9-3 to 9-7
bandwidth graph, 8-10
Bode plot, 8-7
circuit, 8-2
compensation, 8-13 to 8-15
equation comparison, 9-8 to 9-10
feedback capacitance, 8-12 to 8-14
feedback capacitance Bode plot, 8-13
feedback capacitance compensation, 8-13 to 8-15
feedback resistor graph, 8-10
feedback resistor selection, 8-9 to 8-12
frequency graph, 9-5
gain capacitance compensation, 8-13 to 8-15
gain graph, 8-10, 9-5
ideal circuit, 9-3
impedance, 9-7 to 9-9
input capacitance, 8-11 to 8-13
inverting, 8-5 to 8-7
inverting circuit, 8-5
model, 8-1 to 8-3
nonnverting, 8-3 to 8-6
noninverting circuit, 8-4
precision, 9-2 to 9-4
stability, 8-11 to 8-14, 9-6 to 9-8
stability analysis, 8-7 to 8-10
stability analysis circuit, 8-2, 8-3
stability equation, 8-2 to 8-4
stray capacitance graph, 8-12
voltage feedback amp comparison, 9-1 to 9-10
Current source, 12-7
Curve
band reject filter, 5-10
bode plot of low-pass filter, 5-9
Bode plot, no poles, 5-11
Bode plot, with single pole, 5-11
damping ratio vs overshoot, 5-16
damping ratio vs phase margin, 5-16
loop gain plot, 5-13, 5-14
low-pass filter, 5-9
overshoot vs damping ratio, 5-16
phase margin vs damping ratio, 5-16
single supply inverting op amp, 4-5
single supply noninverting op amp, 4-7
transfer, 4-12, 4-15, 4-18, 4-21

D

DAC
AC application error budget, 14-8
AC errors, 14-14 to 14-18
AC loads, 14-2
AC parameters, 14-14 to 14-18
accuracy vs resolution, 14-7
actuator analog interface, 18-16 to 18-21
compensation, 14-18 to 14-20
current booster circuit, 14-20
DC application error budget, 14-7
DC errors, 14-10 to 14-14
DC loads, 14-1 to 14-3
DC offsets, 14-22
DC parameters, 14-10 to 14-14
deglitch circuit, 14-17
differential nonlinearity error, 14-12
differential nonlinearity error graph, 14-13
dynamic range, 14-9 to 14-11
ENOB, 14-14
error budget, 14-6 to 14-11
errors, 14-10 to 14-16
external ref circuits, 13-15 to 13-18
external voltage reference circuit, 13-17
gain error, 14-11 to 14-13
gain error graph, 14-12
integral nonlinearity error, 14-13 to 14-15
intermodulation distortion, 14-15 to 14-17
interpolation filter, 13-5
nonlinearity error, 14-12 to 14-15
offset error, 14-10 to 14-12
offset error graph, 14-11
parameters, 14-10 to 14-16
power booster, 14-22
power supply rejection ratio, 14-14
R/2R, 14-4 to 14-6
R/2R circuit, 14-5
R/R2 resistor array, 14-4
reconstruction filter, 13-5, 13-13 to 13-16

resistor ladder, 14-2 to 14-4
resistor ladder circuit, 14-2
RF application budget, 14-10
selection, 13-6 to 13-10
settling time, 14-16 to 14-18
sigma-delta , 14-5 to 14-7
SINAD, 14-14
single-supply operation, 14-22
specifications, 14-2 to 14-6
spurious free dynamic range, 14-15
THD, 14-8 to 14-10, 14-14
types, 14-2
voltage booster circuit, 14-21, 14-22
weighted resistor, 14-3 to 14-5
weighted resistor circuit, 14-3
DAC current sink, actuator interface circuit, 18-19
DAC current source, actuator interface circuit, 18-17
Damping ratio, graph, 7-4
DC loads, DAC, 14-1 to 14-3
DC offsets, DAC, 14-22 to 14-24
DC parameters, DAC, 14-10 to 14-14
Deglitch circuit, DAC, 14-17
Design, op amp circuit, 12-16 to 12-23
Design hints, 16-47 to 16-55
Design procedure, 12-11 to 12-13
Differential amp, closed loop response, 13-19
Differential input voltage range, parameter, 11-11
Differential nonlinearity error, 14-12 to 14-14
graph, 14-13
Differential op amp, 3-6 to 3-8, 6-8 to 6-10
circuit, 3-6, 6-8
common-mode input signal circuit, 3-7
noise, 10-18
Differential output drive, from single-ended circuit, 13-18
Differential voltage amplification, parameter, 11-13
Digital plane placement, 17-8
Distortion, total harmonic, 11-18 to 11-20
Dominant-pole compensation, 7-9 to 7-13
graph, 7-11
Dominant-pole compensation circuit, bode plot, 7-11
Dual IF receiver, block diagram, 13-4
Dual supply op amp, vs single supply, 4-1 to 4-4
Dual-supply filter, circuit, 16-47
Dynamic range
DAC, 14-9 to 14-11
low-voltage op amp designs, 18-3 to 18-6

E

Effective number of bits, DAC, 14-14
Elapsed time
graph for TL07x, 7-5
graph for TL08x, 7-6
ENOB, DAC, 14-14
Equivalent input noise, parameter, 11-17 to 11-19
Error sources, op amp, 18-4
External compensation, 7-8

F

Fairchild
Robert J. Wilder, 1-3
uA709, 1-3
Feedback, 5-6 to 5-8
Bode analysis, 5-7 to 5-13
Wien bridge oscillator circuit, 15-12
Feedback capacitance
Bode plot, 8-13
compensation, 8-13 to 8-15
current feedback amp, 8-12 to 8-14
Feedback circuit, block diagram math, 5-3
Feedback loop, broken circuit, 6-7
Feedback network, complex, 3-7 to 3-9
Feedback resistor
graph, 8-10
selection, 8-9 to 8-12
Feedback system, block diagram, 6-2
Filter
anti-aliasing, 13-11 to 13-13
band reject circuit, 5-9
band reject curve, 5-10
capacitor selection, 16-50 to 16-53
cascading filter stages, 16-12
characteristics of reconstruction, 13-14
coefficients, 16-55 to 16-63
component values, 16-52
DAC recontracting, 13-13 to 13-16
description, 16-1 to 16-3
dual-supply circuit, 16-47
low pass fundamentals, 16-2 to 16-11
low-pass design, 16-11 to 16-21
op amp selection, 16-53 to 16-55
quality factor, 16-9 to 16-11

single-pole reconstruction circuit, 13-15
single-supply circuit, 16-47
voltage reference circuit, 13-16
Filter circuit biasing, 16-47 to 16-51
First-order all-pass filter, 16-44
First-order low-pass filter, 16-12 to 16-15
Flicker noise, 10-8 to 10-10
Flux residue resistance, 17-20
Fourth-order band-pass filter, 16-32 to 16-37
Frequency
corner, 10-13 to 10-15
current feedback amp graph, 9-5
graph for TL07x, 7-5
graph for TL08x, 7-6
graph for TLV277x, 7-7
voltage feedback amp graph, 9-4
Full-wave rectifier, 12-10

G

Gain
current feedback amp, 9-5
current feedback amp graph, 8-10
oscillator, 15-4 to 15-6
voltage feedback amp, 9-4
Gain capacitance, compensation, 8-13 to 8-15
Gain compensation, 7-12 to 7-14
graph, 7-12
Gain error
DAC, 14-11 to 14-13
graph, 14-12
Gain error budget, 12-19
Gate output, circuit, 17-21
Gaussian distribution, noise graph, 10-2
Glossary, parameter, 11-2 to 11-10
Graph
band-rejection filter, 16-41
Bassel low-pass filter, 16-8, 16-9
Bode plot for dominant-pole compensation, 7-11
Bode plot for lead compensation, 7-14
Butterworth band-pass filter gain response, 16-36
amplitude of Butterworth low-pass filter, 16-6
Butterworth low-pass filter, 16-51
comparison of active/passive band-rejection filter, 16-41
comparison of normalized group delay graph, 16-8
current feedback amp, 8-10
current feedback amp bandwidth, 8-10
current feedback amp frequency, 9-5
current feedback amp gain, 8-10, 9-5
DAC gain error, 14-12
DAC offset error, 14-11
damping ratio, 7-4
diff amp closed loop response, 13-19
differential nonlinearity error, 14-13
dominant-pole compensation, 7-11
elapsed time for TL07x, 7-5
elapsed time for TL08x, 7-6
feedback resistor, 8-10
filter response, 16-53
frequency for TL07x, 7-5
frequency for TL08x, 7-6
frequency for TLV277x, 7-7
frequency response for group delay, 16-43
frequency response of low-pass filter, 16-4
gain compensation, 7-12
gain of Tschebyscheff low-pass filter, 16-7
gain response band-pass filter, 16-29
group delay frequency response, 16-43
high-pass filter gain response, 16-22
input bias current changes, 18-9
input common-mode voltage, 18-9, 18-10
input noise spectrum, 11-18
input output voltage changes, 18-10
integral nonlinearity error, 14-13
intermodulation distortion, 14-16
lead compensation, 7-13
lead-lag compensation, 7-20
low pass to band pass transition, 16-28
maximum output voltage swing, 11-12
noise energy, 10-2
op amp bandwidth, 15-7
op amp frequency response, 15-6
open loop response, 16-53
oscillator output, 15-7
output spectrum, 11-19
output voltage for TL07x, 7-5
output voltage for TL08x, 7-6
percent overshoot, 7-4
phase margin, 7-4
phase margin for TLV277x, 7-7
phase response of phase response of low-pass filter, 16-4
phase shift, 11-21
phase shift for RC, 15-3
phase shift for TL03x, 7-3

phase shift for TL07x, 7-5
phase shift for TL08x, 7-6
power spectral density, 13-14
pulse response for TL03x, 7-3
pulse response for TLV277x, 7-7
settling time, 11-22, 14-17
spurious free dynamic range, 14-15
stray capacitance, 8-12
time response for TL07x, 7-5
time response for TL08x, 7-6
TL03x, 7-3
TL07x, 7-5
TL08x, 7-6
TLC2201 noise, 10-20
TLV2772 noise, 10-13
Tschebyscheff low-pass filter quality factor, 16-10
voltage amplification, 11-21
voltage amplification for TL03x, 7-3
voltage amplification for TL07x, 7-5
voltage amplification for TL08x, 7-6
voltage amplification for TLV277x, 7-7
voltage feedback amp frequency, 9-4
voltage feedback amp gain, 9-4
voltage follower frequency, 13-17
Wein bridge oscillator, 15-11
Grounds, 17-8
Group delay, frequency response graph, 16-43
GSM cellular base station, block diagram, 13-2
GSM receiver block system budget, 13-3

H

High-frequency capacitor, circuit, 17-12
High-frequency inductor, circuit, 17-13
High-frequency resistor, circuit, 17-11
High-pass filter
Bessel circuit, 16-27
circuit, 3-10
design, 16-21 to 16-27
first order, 16-23 to 16-25
from a low-pass filter, 16-21
gain response graph, 16-22
higher order, 16-26 to 16-28
inverting circuit, 16-23
MFB circuit, 16-25
noninverting circuit, 16-23
Sallen-Key circuit, 16-24
Sallen-Key circuit with unity gain, 16-24
second order, 16-24 to 16-27
High-speed analog input drive circuits, 13-18 to 13-22

I

Ideal op amp
assumptions, 3-1
characteristics, 3-2
circuit, 3-2
IF receiver, block diagram, 13-4
Impedance
current feedback amp, 9-7 to 9-9
voltage feedback amp, 9-7 to 9-9
Inductive vias, 17-19 to 17-21
Inductor, high-frequency circuit, 17-13
Input capacitance, 11-14
current feedback amp, 8-11 to 8-13
Input common-mode range, op amp, 18-6 to 18-8
Input common-mode voltage range, parameter, 11-11
Input current, 11-10
circuit, 11-10
Input noise spectrum, graph, 11-18
Input offset voltage
adjust circuit, 11-9
circuit, 11-9
parameter, 11-8 to 11-10
Input parasitic elements, 11-13 to 11-15
circuit, 11-13
input capacitance, 11-14
input resistance, 11-14
output impedance, 11-14
output impedance circuit, 11-14
Input resistance, 11-14
Instrumentation amaplifiers, 12-1 to 12-24
Integral nonlinearity error, DAC, 14-13
Intermodulation distortion, DAC, 14-15 to 14-17
Internal compensation, op amp, 7-2 to 7-9
Inverting
current feedback amp, 8-5 to 8-7
current feedback amp circuit, 8-5
Inverting high-pass filter, circuit, 16-23
Inverting input, stray capacitance circuit, 7-16
Inverting low-pass filter, circuit, 16-13
Inverting op amp, 3-4 to 3-6, 6-6 to 6-8
capacitance effect, 17-18
circuit, 3-4, 6-6

feedback back loop broken circuit, 6-7
noise, 10-16 to 10-18
single supply, 4-3 to 4-6
single supply circuit, 4-4, 4-5
transfer curve, 4-5
with lead compensation, 7-15

K

Kirchoff's voltage law, 2-2 to 2-4

L

Large signal differential voltage amplification, 11-13
Law, Ohm's, 2-1 to 2-3
Lead compensation, 7-13 to 7-17
Bode plot, 7-14
graph, 7-13
inverting op amp, 7-15
noninverting op amp, 7-16
Lead-lag compensation, 7-18 to 7-21
Bode plot, 7-19
circuit, 7-19
closed-loop plot, 7-20
Loading capacitance, making an op amp unstable, 7-4
Logic gate output, circuit, 17-21
Long-tailed pair, circuit, 9-2
Loop antenna, PCB layout, 17-15
Loop gain, block diagram, 6-4
Loop gain plots, 5-12 to 5-16
curve, 5-13, 5-14
Low-pass filter
adjustable circuit, 16-17
band-pass filter transition, 16-28
band-pass transition, 16-37
Bessel, 16-7 to 16-10
Butterworth, 16-6
Butterworth amplitude response graph, 16-6
Butterworth circuit with unity gain, 16-21
circuit, 5-8
comparison of gain responses graph, 16-9
comparison of normalized group delay graph, 16-8
comparison of phase responses graph, 16-8
design, 16-11 to 16-21
first-order, 16-12 to 16-15
fourth-order RC circuit, 16-3
frequency graph, 16-4
fundamentals, 16-2 to 16-11
high order, 16-19 to 16-22
inverting circuit, 16-13
MFB circuit, 16-18
noninverting circuit, 16-12
Noninverting with unity gain circuit, 16-14
phase response graph, 16-4
RC first-order circuit, 16-2
Sallen-Key circuit, 16-15
Sallen-Key circuit with unity gain, 16-15
Sallen-Key circuit with unity gain circuit, 16-20
second-order, 16-14 to 16-20
second-order active, 16-1
second-order passive, 16-1 to 16-3
to a high-pass filter, 16-21
Tschebyscheff, 16-7
Tschebyscheff gain response graph, 16-7
Tschebyscheff unity gain circuit, 16-16
Tschebyscheff quality factor graph, 16-10
unity gain circuit, 16-19
Low-pass filter circuit, 3-10

M

Math, block diagram, 5-1 to 5-6
Maximum output voltage swing
graph, 11-12
parameter, 11-12
MFB band-pass filter, circuit, 16-31
MFB high-pass filter
biasing circuit, 16-50
circuit, 16-25
MFB low-pass filter
biasing circuit, 16-49
circuit, 16-18
Miller effect compensation, 7-2 to 7-4
circuit, 7-2
Multiloop feedback system, block diagram math, 5-4
Multiple feedback, low-pass filter circuit, 16-18
Multiple feedback band-pass filter, 16-31 to 16-33
circuit, 16-31
Multiple feedback high-pass filter
biasing circuit, 16-50
circuit, 16-25
Multiple feedback low-pass filter, 16-18 to 16-20
biasing circuit, 16-49

N

Noise, 10-8 to 10-10
  avalanche, 10-9 to 10-11
  broadband, 11-18
  burst, 10-9
  characterization, 10-1
  colors, 10-10 to 10-13
  corner, 10-12 to 10-14
  differential op amp, 10-18
  equivalent input, 11-17 to 11-19
  floor, 10-3
  Gaussian distribution graph, 10-2
  graph, 11-18
  inverting op amp, 10-16
  multiple sources, 10-3
  noninverting op amp, 10-17
  op amp, 10-1 to 10-24
  op amp model, 10-14 to 10-16
  pink, 10-11
  red/brown, 10-12
  rms vs P-P, 10-1 to 10-3
  shot, 10-5 to 10-8
  spot, 11-17 to 11-19
  thermal, 10-7 to 10-9
  total, 10-12 to 10-14
  types, 10-4 to 10-11
  units, 10-4
  white, 10-11

Noise corner, 10-12 to 10-14

Noise floor, 10-3

Noise model, op amp, 10-14 to 10-16

Noise sources, PCBs, 17-2

Noninverting, current feedback amplifier, 8-3 to 8-6

Noninverting high-pass filter, circuit, 16-23

Noninverting low-pass filter
  circuit, 16-12
  circuit with unity gain, 16-14

Noninverting op amp, 3-3 to 3-5, 4-6 to 4-8, 6-5 to 6-7
  circuit, 3-3, 18-7
  noise, 10-17
  open-loop, schematic, 6-6
  schematic, 6-5
  single supply circuit, 4-7
  transfer curve, 4-7
  with lead compensation, 7-16

Nonlinear feedback, Wien bridge oscillator circuit, 15-12

Nonlinearity error, 14-12 to 14-15

Nonrail to rail op amp, input circuit, 18-8

Norton's theorem, 2-5

O

Offset budget, 12-19

Offset error
  DAC, 14-10 to 14-12
  graph, 14-11

Ohm's Law, 2-1 to 2-3
  applied to a component, 2-2
  applied to the total circuit, 2-2

Open-loop noninverting op amp, schematic, 6-6

Operational amplifier
  cable, 3-9
  canonical equations, 6-2 to 6-5
  capacitively-loaded circuit, 7-9
  circuit design, 12-16 to 12-23
  circuit theory, 2-1 to 2-13
  comparison, 18-20 to 18-23
  compensated attenuator, 7-16 to 7-19
  current booster, 14-20
  current drain, 18-12 to 18-14
  current feedback analysis, 8-1 to 8-15
  designing low voltage, 18-1 to 18-24
  differential, 3-6 to 3-8, 6-8 to 6-10
  dominant-pole compensation, 7-9 to 7-13
  dynamic range, 18-3 to 18-6
  error sources, 18-4
  external compensation, 7-8
  frequency graph, 15-6
  gain error budget, 12-19
  gain compensation, 7-12 to 7-14
  H. W. Bode, 1-2
  Harry Black, Bell Labs, 1-1
  history of, 1-1 to 1-5
   impact on oscillator, 15-5 to 15-8
  ideal, 3-1
  increasing current, 14-19 to 14-24
  input common-mode range, 18-6 to 18-10
  internal compensation, 7-2 to 7-9
  inverting, 3-4 to 3-6, 6-6 to 6-8, 7-15
  inverting circuit, 6-6
  lead compensation, 7-13 to 7-17
  lead-lag compensation, 7-18 to 7-21
  loading capacitance, 7-4
  loop gain calculation circuit, 7-9

noise, 10-12 to 10-19
noise characterization, 10-1
noise model, 10-14 to 10-16
noise of differential, 10-18
noise of inverting, 10-16 to 10-18
noise of noninverting, 10-17
noise theory, 10-1 to 10-24
non-ideal, 6-1 to 6-11
noninverting, 3-3 to 3-5, 6-5 to 6-7, 7-16
noninverting circuit, 18-7
noninverting, schematic, 6-5
nonrail to rail input circuit, 18-8
offset budget, 12-19
open-loop noninverting, schematic, 6-6
output voltage swing, 18-11 to 18-13
parameter glossary, 11-2 to 11-10
parameters, 11-1 to 11-23
pinouts, 17-24
proper termination, 17-28
rail to rail input circuit, 18-8
rail to rail output circuit, 18-11
rms vs P-P noise, 10-1 to 10-3
selection, 12-15 to 12-17, 13-10 to 13-12
signal-to-noise ratio, 18-5 to 18-7
single-supply circuit design, 18-13
split-supply circuit, 10-19
terminal model circuit, 17-5
TL03x, 7-3
video, 3-9
uA709, 1-3
uA741, 1-3
voltage booster, 14-20 to 14-23

Oscillation, requirements, 15-1 to 15-3

Oscillator
analysis of circuit, 15-7 to 15-9
bubba, 15-17 to 15-19
bubba circuit, 15-17
buffered phase shift circuit, 15-16
circuits for sine wave type, 15-9 to 15-21
description of sine wave, 15-1
gain, 15-4 to 15-6
impact of active element op amp, 15-5 to 15-8
output graph, 15-7
phase shift, 15-3 to 15-5, 15-14 to 15-17
phase shift circuit, 15-14
positive feedback diagram, 15-8
quadrature, 15-18 to 15-21
quadrature circuit, 15-19
sine wave, 15-1 to 15-21
Wein bridge, 15-9 to 15-15
Wien bridge circuit, 15-11, 15-12
Wien bridge graph, 15-11

Output impedance, 11-14
circuit, 11-14

Output spectrum, graph, 11-19

Output voltage
graph for TL07x, 7-5
graph for TL08x, 7-6

Output voltage swing
graph, 11-12
op amp, 18-11 to 18-13
parameter, 11-12

Overshoot predictions, 5-15 to 5-17

P

P-P noise, vs rms noise, 10-1 to 10-3

Parameter
broadband noise, 11-18
common-mode rejection ratio, 11-15
differential input voltage range, 11-11
equivalent input noise, 11-17 to 11-19
glossary, 11-2 to 11-10
input common-mode voltage range, 11-11
input current, 11-10
input offset voltage, 11-8 to 11-10
large signal differential voltage amplification, 11-13
maximum output voltage swing, 11-12
phase margin, 11-19 to 11-22
settling time, 11-22
slew rate, 11-16 to 11-18
spot noise, 11-17 to 11-19
supply current, 11-16
supply voltage rejection ratio, 11-15
total harmonic distortion, 11-18 to 11-20
understanding, 11-1 to 11-23
unity gain bandwidth, 11-19 to 11-22

Parameters, DAC, 14-10 to 14-16

Parasitic elements, 11-13 to 11-15
circuit, 11-13

PCB design
capacitors, 17-12 to 17-14
decoupling, 17-20 to 17-24
grounding, 17-7 to 17-11
inductors, 17-13
input isolation, 17-23 to 17-25
ordering of layers, 17-6 to 17-8
output isolation, 17-23 to 17-25

packages, 17-24 to 17-27
parallel signal traces, 17-19
passive component characteristics, 17-11 to 17-21
resistors, 17-11
surface mount, 17-27
through-hole considerations, 17-26 to 17-29
trace antennas, 17-14 to 17-17
trace characteristics, 17-14
trace reflections, 17-16 to 17-21
PCB layout, 17-10
loop and slot antenna, 17-15
trace corners, 17-16
PCB mechanical conduction, 17-3 to 17-8
double sided, 17-5
layers, 17-4 to 17-7
materials, 17-3 to 17-5
multi-layer, 17-6
single layer, 17-4 to 17-6
PCB traces
broadcasting from, 17-9
trace to plane capacitance, 17-17
Percent overshoot, graph, 7-4
Phase margin, 11-19 to 11-22
graph, 7-4
graph for TLV277x, 7-7
Phase shift
buffered oscillator circuit, 15-16
graph, 11-21
graph for TL07x, 7-5
graph for TL08x, 7-6
oscillator, 15-3 to 15-5
Phase shift oscillator
buffered, 15-15 to 15-17
circuit, 15-14
single amplifier, 15-14 to 15-16
Photodiode amplifier, 12-8
Phototransistor amplifier, 12-9
Photovoltaic cell amplifier, 12-9
Pin outs, op amp, 17-24
Pink noise, 10-11
Plane, to trace capacitance, 17-17
Plane placement, digital and analog, 17-8
Power booster, DAC, 14-22
Power supply rejection ratio, DAC, 14-14
Precision
current feedback amp, 9-2 to 9-4
voltage feedback amp, 9-2 to 9-4
Precision current source, 12-7
Pulse response, graph for TLV277x, 7-7

Q

Quadrature oscillator, 15-18 to 15-21
circuit, 15-19
Quality factor, filter, 16-9 to 16-11

R

R/2R
circuit for DAC, 14-5
DAC, 14-4 to 14-6
resistor array, 14-4
Rail-to-rail op amp, input circuit, 18-8
Rail-to-rail output op amp, output stage circuit, 18-11
RC
first-order passive low pass circuit, 16-2
fourth-order pass low pass circuit, 16-3
low pass filter frequency graph, 16-4
low pass phase response graph, 16-4
RC , phase shift graph, 15-3
Reconstruction filter
characteristics, 13-14
circuit for single pole, 13-15
DAC, 13-13 to 13-16
Red noise, 10-12
Reference bias circuit, 12-13
Reference circuit
ADC, 13-15 to 13-18
DAC, 13-15 to 13-18
Reference voltage characterization, 12-12 to 12-14
Resistance, flux residue, 17-20
Resistor
current feedback amp, 8-9 to 8-12
high-performance circuit, 17-11
Resistor ladder
circuit for DAC, 14-2
DAC, 14-2 to 14-4
Resistive transducer, 12-6, 12-7
Resonance, capacitor, 17-22
Ringing predictions, 5-15 to 5-17
RMS noise, vs P-P noise, 10-1 to 10-3
Rule
current divider, 2-4
voltage divider, 2-3

S

Sallen-Key band-pass filter, circuit, 16-30
Sallen-Key high-pass filter
  biasing circuit, 16-50
  circuit, 16-24
  circuit with unity gain, 16-24
Sallen-Key low-pass filter
  biasing circuit, 16-48
  circuit, 16-15
  circuit with unity gain, 16-15, 16-20
Sallen-Key topology, 16-14 to 16-18, 16-24 to 16-26
  band-pass filter, 16-30 to 16-32
Saturated transistor circuit, 2-9 to 2-11
Schematic, 6-5
  simultaneous equation, 4-9, 4-13, 4-16, 4-19
Second order equation, 5-15 to 5-17
Second-order all-pass filter, 16-44 to 16-46
Second-order band-pass filter, 16-29 to 16-33
  multiple feedback topology, 16-31 to 16-33
  Sallen-Key topology, 16-30 to 16-32
Second-order high-pass filter, 16-24 to 16-27
  multiple feedback topology, 16-25 to 16-27
  Sallen-Key topology, 16-24 to 16-26
Second-order low-pass filter, 16-14 to 16-20
  coefficients, 16-17
  multiple feedback topology, 16-18 to 16-20
  Sallen-Key topology, 16-14 to 16-18
Self resonance, capacitor, 17-22
Settling time, 11-22
  DAC, 14-16 to 14-18
  graph, 11-22
Shot noise, 10-5 to 10-8
Shutdown, op amp, 18-12 to 18-14
Sigma-delta, DAC, 14-5 to 14-7
Signal conditioning, for IF sampling, 13-1 to 13-23
Signal-to-noise and distortion, DAC, 14-14
Signal-to-noise ratio, 10-3
  op amp, 18-5 to 18-7
Significant bits, versus binary bits, 12-2
Simultaneous equations, 4-8 to 4-19
SINAD, DAC, 14-14
Sine wave oscillator, 15-1 to 15-21
  circuits, 15-9 to 15-21
  description, 15-1
Single supply op amp, 4-1 to 4-23
  circuit, 4-1
  inverting circuit, 4-4, 4-5
  circuit analysis, 4-3 to 4-8
  inverting, 4-3 to 4-6
  noninverting, 4-6 to 4-8
  noninverting circuit, 4-7
  transfer curve, 4-5, 4-7
  vs dual supply, 4-1 to 4-4
Single-ended, to diff output drive circuit, 13-18
Single-ended input drive, circuit for ADC, 13-20
Single-supply circuit design, 18-13
Single-supply filter, circuit, 16-47
Single-supply operation, DAC, 14-22 to 14-24
Slew rate
  at unity gain, 11-16
  circuit, 11-16
Slot antenna, PCB layout, 17-15
SNR. *See* signal-to-noise ratio
Spans, requiring correction, 12-4
Spectral density, graph, 13-14
Split-supply op amp, circuit, 10-19
Spot noise, 11-17 to 11-19
Spurious free dynamic range, DAC, 14-15
Stability
  current feedback amp, 8-11 to 8-13, 9-6 to 9-8
  feedback capacitance, 8-12 to 8-14
  op amp loading capacitance, 7-4
  voltage feedback amp, 9-6 to 9-8
Stability analysis
  circuit, 8-2, 8-3
  current feedback amp, 8-7 to 8-10
Stability equation, 8-2 to 8-4
Stray capacitance
  graph for current feedback amp, 8-12
  on inverting input circuit, 7-16
Summary points, block diagram math, 5-3
Superposition, 2-8 to 2-10
Supply current, 11-16
Supply voltage rejection ratio, 11-15
System specifications, review, 12-12

T

T network circuit
  feedback loop, 3-7
  Thevenim's theorem, 3-8
Termination, unused op amps, 17-28

Theorem
Norton's, 2-5
Superposition, 2-8 to 2-10
Thevenin's, 2-5 to 2-8
Thermal noise, 10-7 to 10-9
Thevenin's theorem, 2-5 to 2-8
complex feedback networks, 3-7 to 3-9
THS3201
gain vs frequence graph, 13-21
phase vs frequency graph, 13-21
TL03x, 7-4
frequency response graph, 7-3
op amp, 7-3
phase shift graph, 7-3
small-signal pulse response graph, 7-3
voltage amplification graph, 7-3
voltage follower graph, 7-3
TL07x, 7-4 to 7-6
graph, 7-5
TL08x, 7-5 to 7-7
graph, 7-6
TLC2201, 10-19 to 10-24
circuit, 10-21, 10-23
noise graph, 10-20
TLE2027, 10-19
TLV227x, 7-6 to 7-9
graphs, 7-7
TLV2772, 10-12 to 10-14
noise graph, 10-13
typical noise, 10-14
Total harmonic distortion, 11-18 to 11-20
DAC, 14-8 to 14-10, 14-14
Total noise, 10-12 to 10-14
Trace
length for inverting op amp, 17-25
length for quad op amp, 17-25
parallel signal coupling, 17-19
quad op amp with half-supply generator, 17-26
to plane capacitance, 17-17
Trace corners, PCB layout, 17-16
Transducer, 12-1 to 12-6
ADC analog interface, 18-13 to 18-17
ADC interface circuit, 18-15, 18-16
characterization, 12-13 to 12-16
measurement system, 12-1
output voltage, 12-14
resistive, 12-6, 12-7
types, 12-6 to 12-12
Transducer bias circuit, 12-13
Transforms, block diagram math, 5-5
Transistor amplifier, 2-10 to 2-13
Tschebyscheff low-pass filter, 16-7
Butterworth response graph, 16-51
gain response graph, 16-7
quality factor graph, 16-10
second order, 16-16
Twin-T filter
active circuit, 16-38
passive circuit, 16-37

U

uA709, op amp, 1-3
uA741, op amp, 1-3
Unity gain, 11-16
Unity gain bandwidth, 11-19 to 11-22

V

VFA. *See* Voltage feedback amplifier; voltage feedback amplifier
Vias, inductive, 17-19 to 17-21
Video amplifier, 3-9
Video amplifier circuit, 3-9
Voltage amplification
graph, 11-21
graph for TL07x, 7-5
graph for TL08x, 7-6
graph for TLV277x, 7-7
parameter, 11-13
Voltage booster,
op amp, 14-20 to 14-23
circuit for DAC, 14-21, 14-22
Voltage divider rule, 2-3 to 2-5
Voltage feedback amplifier, 7-1 to 7-23
bandwidth, 9-3 to 9-7
current feedback amp comparison, 9-1 to 9-10
equation comparison, 9-8 to 9-10
frequency graph, 9-4
gain graph, 9-4
impedance, 9-7 to 9-9
long-tailed pair circuit, 9-2
precision, 9-2 to 9-4
stability, 9-6 to 9-8
Voltage follower frequency, graph, 13-17
Voltage reference circuit

ADC, 13-17
DAC, 13-17
Voltage reference filter, circuit, 13-16

W

Weighted resistor
circuit for DAC, 14-3
DAC, 14-3 to 14-5
Wein bridge oscillator, 15-9 to 15-15
circuit with AGC, 15-13
circuit with nonlinear feedback, 15-12
graph, 15-11
Wheatstone bridge, circuit, 12-8
White noise, 10-11
Wien bridge
circuit, 15-9
oscillator circuit, 15-11
Wien-Robinson filter, 16-39 to 16-42
active circuit, 16-39
passive circuit, 16-39
Wilder, Robert J., Fairchild, 1-3
Wireless communication, 13-1 to 13-23
Wireless systems, 13-1 to 13-6

Elliott Sound Products **Project 48**

Active Sub-Woofer and Controller

Rod Elliott (ESP)
Updated 03 Feb 2004

PCBs are available for this project. Click the image for details.

Introduction

Sub woofers are very popular, with home theatre being one of the driving forces. However, a good sub adds considerably to normal hi-fi program material, and especially so if it is predictable and has good response characteristics.

The majority of sub woofers use a large speaker driver in a large box, with tuning vents and all the difficulties (and vagaries) that conventional operation entails. By conventional, I mean that the speaker and cabinet are operated as a resonant system, using the Thiele-Small parameters to obtain a box which will (if everything works as it should) provide excellent performance.

Photo of Completed Prototype

The principle of Extended Low Frequency™(or ELF™) [1] is surprisingly uncommon, with one manufacturer that I have found using it in their subs [2]. I suspect there is one other, but I am not certain that the same method is used - although the principle is the same. Since ELF is trademarked, I will not be using the term in this project, but will refer to my version as Electronically Assisted Subwoofer (EAS). I briefly thought about Electronic Subwoofer Principle (ESP) but decided that would be silly :-)

The basic principles were discovered by Edward Long and Ronald Wickersham (although there is a

possibility that others have used similar principles beforehand, there is little available literature), and they both point out that there are some major problems in the reproduction of low bass, highlighting the fact that the bass is the foundation upon which the sound image is created, and that the phase response of ported enclosures can cause "smearing" of the sound in the time domain. I don't know about "smearing", but I do know that my prototype provides bass that is deeper and tighter than anything I have heard before. Ported enclosures definitely cause problems with the sound, as the reproduction mechanism relies on two resonant systems, and it takes time for the sound to build up and decay.

Siegfried Linkwitz [3] developed a circuit that equalises the bottom end of the system, but does not affect the higher frequencies. This is shown and described fully in Project 71, and must be used with a crossover. Although it offers several advantages over the EAS system described here, it is also much more reliant on your detailed knowledge of the loudspeaker driver parameters.

The electronics to perform the necessary processing are easy to build, with the only hard part being a suitably high powered amplifier, and the correct choice of loudspeaker drivers. The cabinet is very easy, since it is small and sealed, so there are no issues with resonance and tuning to worry about.

What? A small, sealed box for a sub-woofer - that can't be right. Well, it is, and the principle is quite different from the conventional approach. When a loudspeaker is installed in a sealed box (or any box, for that matter), it will have a resonant frequency that is higher than in free air. The smaller the box, the higher the resonant frequency will be.

With the EAS approach, the idea is to operate the speaker below resonance, where all the impedance peaks have been left behind in the upper frequencies, leaving a very predictable performance driver to handle the low frequency range. With some experimentation, I determined that a 55 litre box was ideal (it turned out to be 60 litres without the speaker, so with the speaker installed it is about right) for the driver I have, a 4 Ohm, 250W 380mm monster, with a free air resonant frequency of 18Hz. In the small box, this is increased to 63Hz, and this defines the maximum frequency of operation. Resonance probably should have been a little higher, but it manages to sound right, so I shall not worry too much.

Below resonance, a loudspeaker in a sealed box has a response that falls at 12dB / octave, so a means is required to provide an amplifier drive signal that increases at the same rate. A very common circuit in electronics is an integrator, and these are used in many signal processing applications. An integrator has a frequency response that falls at 6dB / octave from DC, extending as far as one wishes. By using two integrators, we obtain a response that falls at 12dB / octave, and by adding resistors, we can cause the response to shelve at any frequency we select. By including capacitors, we can create a high pass filter, so that response to DC is not possible (and nor is it desirable - but more on this later).

It appears that for reasons that are a complete mystery to me (and others), the ELF™ process is patented. Since the basic theory is public domain, and has been discussed by others [3], [4], at some length, it is doubtful that a patent challenge would stand up in a court, however I must warn you of this. In theory, the construction of this project is possibly a violation of patent, however for individual use it is impossible to enforce. Commercial production is a different matter, but no-one would do this without my consent anyway (hmmm), naturally having read and understood my disclaimer and copyright notice.

Driver Selection and Hardware

A quick word is warranted here, to allow you to determine if the speaker you have will actually work in a small sealed enclosure. The EAS principle (or Linkwitz transform circuit) will allow any driver to extend to 20 Hz or even lower. A good quick test is to stick the speaker in a box, and drive it to 100W or so at 20 Hz - you should see a lot of cone movement, a few things will rattle, but you shouldn't actually hear a tone. A "bad" speaker will generate 60 Hz (third harmonic) - if you don't hear anything, the speaker will work in an equalised sub.

If a tone is audible, or the speaker shows any signs of distress (such as the cone breaking up with appropriate awful noises), then the driver cannot be used in this manner. Either find a different driver, or use a vented enclosure.

Before you can build your own EAS box, you will need to select a suitable driver, using the above as a guide. Cone excursion will be very high at the lowest frequencies, so the speaker needs to be capable of high power, good excursion, and of reasonable size (there is no substitute for cone area for moving air at low frequencies). I am using a 380mm (15") driver, but two smaller drivers (say 300mm - 12") can be used, or even a larger number of smaller drivers. I have also had excellent results with a single 300mm driver, which has lower sensitivity (as one would expect) but is perfectly adequate for normal usage.

The test methods I used are applicable to any combination, but in general I suggest either a single large driver or a pair of (say) 300mm units. The next hurdle is the amplifier needed to drive the speaker. This is not trivial. If the selected driver has a sensitivity of 93dB / W @ 1 metre, then you can safely assume that the efficiency will be less than this below resonance, by a factor of maybe 6dB or more. If you are used to driving a sub with 100W, this means that you have just increased the power to 400W - although this is an over-simplification.

If we are to operate the sub from 60Hz (my goal from the beginning), we will increase the power by 12dB for each octave, so if 20W is needed at 60Hz, then at 30Hz this has increased to 320W, and at 15Hz, you will need over 5kW.

Fortunately, the reality is a little different, and 400W or so will be more than sufficient for a fairly powerful system, due mainly to the fact that the energy content in the low bass region is not normally all that great. (Although some program material may have very high energy content, in general this is *not* the case). The EAS system augments the existing system, which is allowed to roll off naturally - contrast this with the normal case, where a crossover is used to separate the low bass from the main system, so existing speaker capability is lost.

This method of driving a sub is best suited to systems where the main enclosures are sealed rather than ported. While it will still work with ported systems, the phase response will be unpredictable, and overall performance may not meet your expectations. As an experiment, the port(s) can be blocked. This will not only make the EAS box work better with your main loudspeakers, but will also improve their bass transient response. Make sure that any modification you make is reversible!

One area where you do need to be very careful is the selection of bass driver and box size. These are critically tied to the characteristics of the selected subwoofer, and should be carefully matched for best

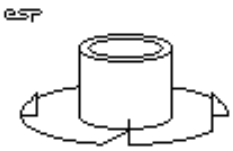
performance (and minimum power). If the box is too small, then the resonant frequency will be too high, causing a dramatic decrease in efficiency and a peaking response from the sub. If the box is too large, resonance will be lower than required, and the system will have a response that does meet the 12dB/ octave characteristics properly.

It is helpful if you know the -3dB frequency of your main speakers, and multiply by 0.75 to obtain the -6dB frequency. This is the frequency where you will actually cross over from the main speakers to the sub-woofer. The EAS box ideally needs to be tuned so that its resonance is between 1 and 1.4 times that of the main speakers. As part of my experiments, I checked the resonance of my EAS box, and found it to be 63Hz. What surprised me a little was the magnitude of the peak - at just over double the off-resonance impedance. This is very low, and is (presumably) a result of the small box and the fibreglass filling. A standard frequency scan shows a slight rise in output level at resonance, but this may be ignored (the room will have a far more profound effect anyway).

Let's assume that your main speakers are 3dB down at 60Hz - this is the resonant frequency of a sealed box. This means they will be 6dB down at 45Hz. The resonant frequency for the EAS box / loudspeaker combination therefore needs to be between 60Hz and 80Hz to obtain a good match. You will either need to experiment with actual boxes, or use a speaker box design program (BassBox, BoxPlot, etc.) - or just build it and have a listen.

There seems to be a fair bit of flexibility in the system, so it is unlikely that the unit you build will be completely unsuited - unless it has resonance at 60Hz and your main speakers roll off above 100Hz, for example. If you make the box a little too large, it is easy to reduce the volume by adding some suitable packing. If it is too small, this will be slightly more difficult :-)

The box I built is made from 25mm (1") MDF (Medium Density Fibreboard), and packed with fibreglass. Apart from the fact that it is very heavy (which is a good thing, because it wants to walk with low frequencies), the cabinet is acoustically dead, with no resonances in the low frequencies at all (completely unlike my house and furniture, dammit !). The woofer is recessed into the baffle, and sealed with weather sealing foam. When attaching the speaker, do NOT use wood screws, or any other screw into the MDF. I used "Tee" nuts. I have no idea what they are called elsewhere in the world, but they look like this ...



Tee Nut

The centre is tapped, and accepts a metal thread screw, and the little spikes mean that you just have to drill a hole, and hammer in the Tee nut. If you use a screw through the hole and screwed lightly into the Tee nut, you can hold it in place as you bash away at it, and can also see that it is straight when you are done. Just make sure that the end of the screw doesn't stick out the end, or you will never remove it again after the hammering! I suggest that you lock the tee nut into place with some construction adhesive (don't get any in the threaded section) so they don't fall out while you are installing the speaker.

The EAS Controller

The controller is quite (actually very) simple, and the circuit is shown in Figure 1. An input buffer ensures that the input impedance of the source does not affect the integrator performance, and allows summing of left and right channels without any crosstalk. The output provides a phase reversal switch, so that the sub can be properly phased to the rest of the system. If the mid-bass disappears as you advance the level control, then the phase is wrong, so just switch to the opposite position.

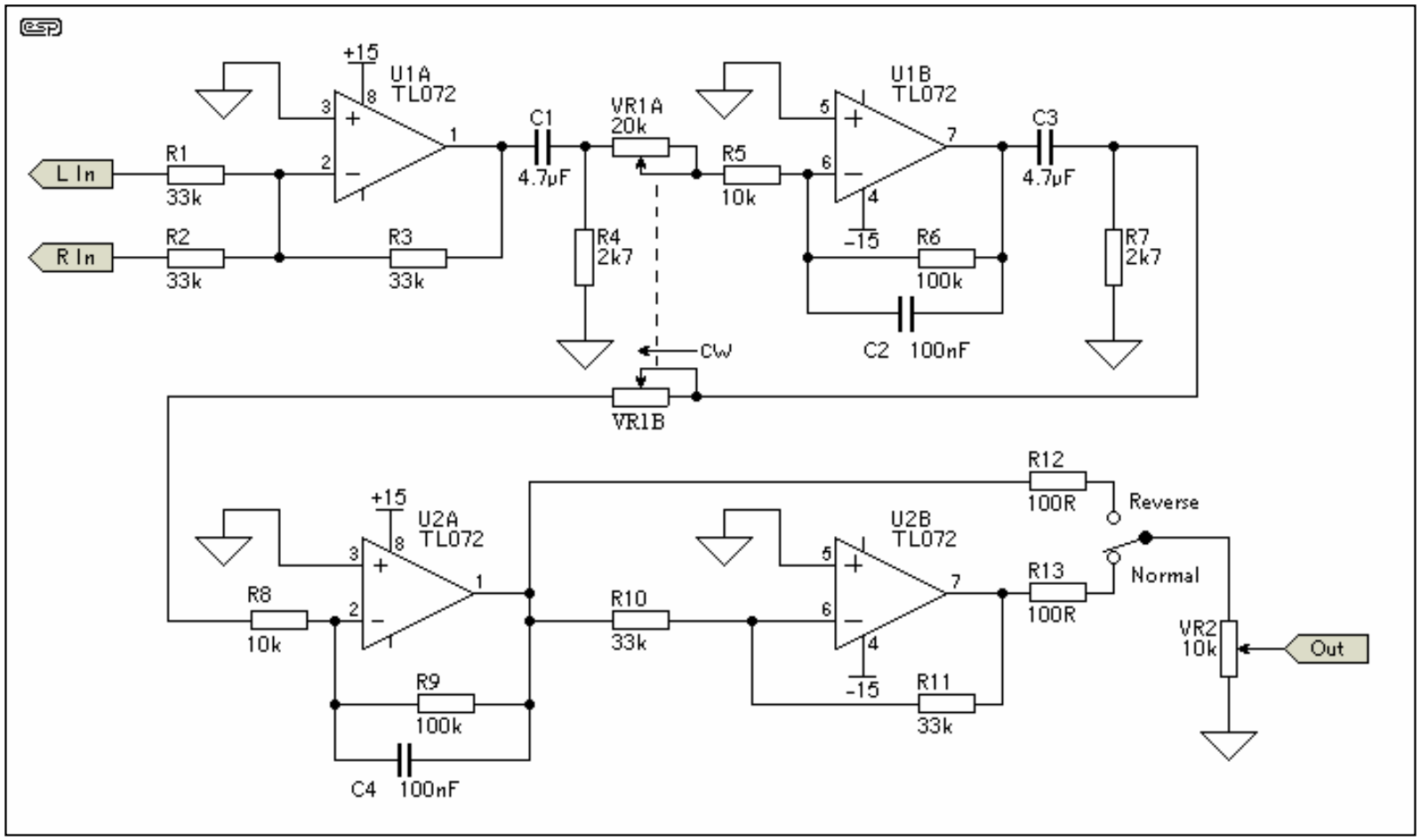


Figure 1 - The Original EAS Filter / Controller

It turns out that the controller can be simplified, but there really is no point. While the dual pot seemed like a good idea when I built my unit, it actually only changes the gain. Now, having experimented some more, this is a very good thing, since it means that the level through the controller can be set to make sure that there is no distortion - there can be a huge amount of gain at low frequencies, and if the gain is too high, distortion is assured!

The integrators (U1B and U2A) include shelving resistors (R6 and R9), and the capacitor / resistor networks (C1-R4, C3-R7) ensure that signals below 20Hz are attenuated. If you don't want to go that low, then the value of the caps (or the resistors R4 & R7) can be reduced. I used 4.7uF caps, and these are non-polarised electrolytics - a high value was needed to keep the impedance low to the integrators. I originally included the dual pot (VR1) to allow the upper frequency rolloff to be set - however it does no such thing (as described above). The final output level is set with VR2, which may be left out if your power amp has a level control.

It is OK to substitute different opamps, but there is little reason to do so. Any substitution device should be a FET input opamp, or DC offset may be a problem. Do not be tempted to use a DC coupled amp. If the one you are planning to use is DC coupled, the input should be isolated with a capacitor. Choose a value

to give a -3dB frequency of about 10Hz, as this will have little effect on the low frequency response, but will help to attenuate the subsonic frequencies.

The unity gain range (using a 20k pot as shown) is from 53Hz to 159Hz. This should be sufficient for most systems, but if desired, the resistors (R5 and R8) can be increased in value to 22k, or you can select a larger value pot. Using 22k resistors and the 20k pot will give a range from 36Hz to 72Hz.

The unity gain frequency is important in only one respect - it will determine the internal gain of the system, and needs to be set based on the input signal level. If the unity gain frequency is set to (say) 100Hz and there is a 1V RMS input, then a 1V RMS input at 20Hz will severely clip the integrators. I suggest that VR1 should be internal, and set when you know the signal level you will be using - this is determined by the input sensitivity of your power amplifier(s) used on the main system. It is probably easier to experiment a little than try to measure everything.

To allow lower frequencies, you can increase the 100k shelving resistors (R6 and R9) to 220k, and increase the high pass capacitors (4.7uF) with 10uF (or R4 & R7 may be increased - a maximum of 4.7k is recommended). This will give a turnover frequency of around 8Hz, but expect to use a lot more power, as there will likely be significant sub-sonic energy that will create large cone excursions with no audible benefit.

The input must be a normal full range (or for a biamped system, the complete low frequency signal). Do not use a crossover or other filter before the EAS controller. For final adjustment, and to integrate the system into your listening room, I recommend the Project 84 constant-Q equaliser. The end result using this is extraordinarily good - I have flat in-room response to 20Hz!

For the power supply, use the one in Project 05, or anything else will provide +/-15V at a few milliamps. My supply is not even regulated, and the entire system is as close to noiseless as you will hear (or not hear). Construction is not critical - I built mine on a piece of Veroboard (perforated prototype board), and managed to fit everything (including the power supply rectifier and filter) on a piece about 100 x 40 millimetres with room to spare.

The EAS system is surprisingly easy to set up with no instrumentation. Of course if you have an SPL meter and oscillator you can also verify the settings with measurements. Remember that the room acoustics will play havoc with the results, so unless you want to drag the whole system outdoors, setting by ear might be the easiest. Even if you did get it exactly right in an anechoic environment, this would change completely once it was in your listening room anyway.

It takes a little experimentation to get right, but is surprisingly easy to do. When properly set, a test track (or bass guitar) should be smooth from the highest bass note to the lowest, with no gross peaks or dips. Some are inevitable because of room resonances and the like, but you will find a setting that just sounds "right" with little difficulty.

Power Amplifier

I have had many requests for a high power amplifier (the 400W unit I am using was a kit - on special - so I couldn't pass it up :-) The easiest is to use the subwoofer amp (Project 68) which is designed specifically

for subwoofers.

Now, if you wanted a really powerful sub, use two 300mm (12") woofers, with one 250W amp per speaker. A system such as this would be excellent, especially since the effective cone area is greater than the 380mm woofer I am using. A typical 300mm speaker has a cone area of about 0.05 m² versus 0.085 m² for a 380mm unit, so two 300mm speakers will have a total effective cone area of 0.1 m². You could go all the way, with a 300mm driver on each face of a cube, with a 250W amp for each. 1500W of low bass should do the job, especially with a cone area of 0.3 m².

This idea is not as silly as it sounds. With six 200mm (8") drivers, you would get a cone area of nearly 0.13 m² (about the same as a single 450mm (18") driver), and the box would still only be quite small at about 60 litres (my guess). There is a lot of scope for experimentation, with far less chance of a box being completely useless than with conventional ported designs.

> Had the driver I am using not been on special, I probably would have used a pair of 300mm drivers for my own sub. As you may have guessed, I wanted to keep the cost of this down to a minimum. I even managed to get the MDF for AU$15 for the lot - as offcuts. The paint for the box cost more than that !

You must accept that for this project to work, you will need a bigger amp than you might have done otherwise. On the positive side, you will get performance that exceeds that of most subs - mine appears to give almost flat response down to less than 20Hz, except I can't actually hear it. My sound level meter can, and it confirms that the response is there. There is rolloff below 15Hz, but I suspect that is my meter, rather than the woofer, since the dual integrator was still working. This was before I built the unit properly, and I allowed the test circuit to go down to 1.5Hz - this caused some excessive cone excursions, to say the least !

Performance Of My Prototype

I measured 80dB SPL at 1 metre in my workshop (sub-woofer perched on a stool in more or less the centre of the space) with at 25Hz and 70W. This improved dramatically when the unit was installed in the listening room, but as I said earlier, there is generally not a lot recorded below around 35Hz. The longest pipe on the organ is about 16Hz, but I think it is not often used (they actually stopped the pipe in the Sydney Town Hall if I remember correctly, because when it was used, the interior trimmings in the building fell off !).

A couple of orchestral recordings revealed traffic (or perhaps underground railway) rumble that I was completely unaware of before (however this was before it was set correctly, and the bass was a tad louder than needed). Once set up properly, its presence is very unobtrusive - except I now have about one and a half octaves of additional bottom end.

I eventually decided on a 20Hz minimum frequency (-3dB), and this is reflected in the component values shown in Figure 1. The actual roll-over frequency is 16.5Hz, after which the output is attenuated at about 12dB / octave (see Figure 2). Without the rolloff capacitors, the gain would be 20dB at 20Hz. Unity gain frequencies are about 4Hz and 63Hz with the pot 20k centred.

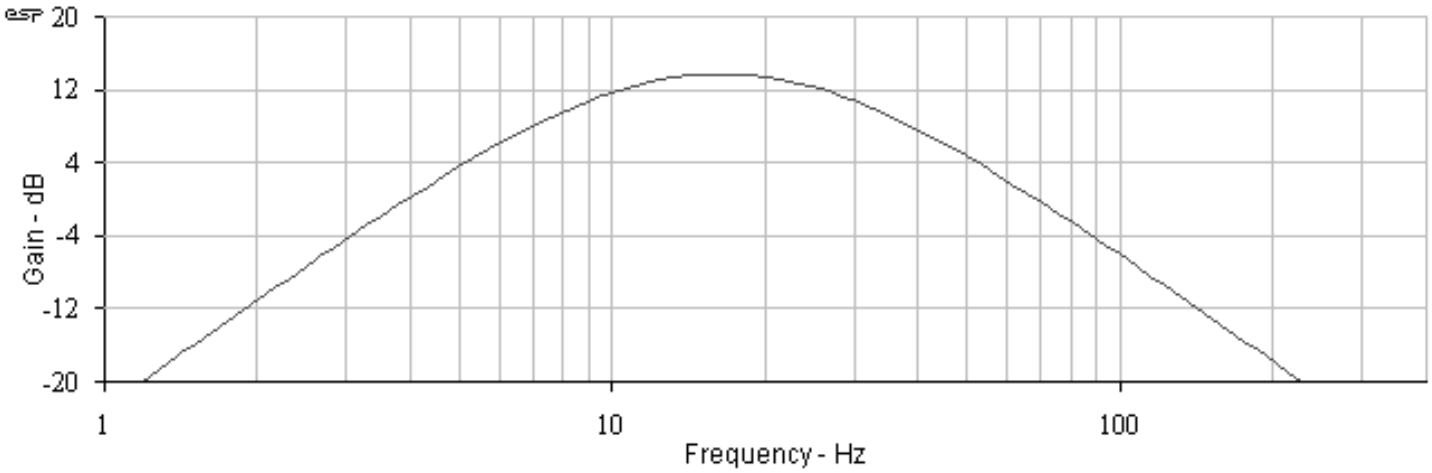


Figure 2 - Frequency Response of EAS Controller

Some Australian readers may recognise the woofer brand in the photo (Figure 3) of my completed unit. The compact size of the box can be seen from the fact that there is very little spacing around the speaker itself, and most of what is there is the top and sides - I used 25mm MDF, so it makes the outside of the box quite a bit bigger than the inside. Outside dimensions are 470W x 450H x 410D (18 1/2"W x 17 1/2"H x 16"D), which gives a capacity of 60 litres (about 2.1 ft³ - excluding the internal space occupied by the speaker. I think you would agree that this is a small box indeed for a 380mm loudspeaker that performs down to 15Hz.

Figure 3 - Photo of Completed EAS Cabinet

Overall, I would have to say that I doubt that any conventional design would be as compact, or would have such clarity and solidarity. Being a sealed box, there is none of the "waffle" that ported designs often give, and the speaker is protected against excessive excursion by the air pressure in the box itself (below the cutoff frequency, anyway).

The bottom end in my system is now staggering. It is rock solid, and absolutely thunders when called

upon. The 400W amp is more than sufficient for the job, considering it has to keep up with a biamped main system capable of very high SPL (up to 120dB at my listening position). In fact a quick test indicates that 200W would have been enough (but ... better to have it and not need it than need it and not have it).

The fact that the EAS design augments the existing speakers rather than taking over from them completely with a crossover goes a long way towards ensuring the power requirements do not get out of hand. As an added benefit, I have found that I get the same aural sensation at much lower SPLs - I can listen quite happily at 90dB, but it sounds much louder. I can even hear the phone ring while listening now !

All in all, I feel it is unlikely that anything other than an isobaric enclosure could give the same performance for a box size even close to the EAS box, and even then would be limited to about 35Hz. Added to this is the rather unpredictable combined response of the main speakers and the sub, which is not an issue with this design. With an EAS system, more power is needed than a conventional design, but for many people, power is much cheaper than space.

Given the performance, I would never consider a conventional sub again, except for the isobaric I have in my car - that only gets down to 40Hz (-3dB), but is acceptable considering the listening environment. The only other sub that takes my fancy is a dipole (i.e. open backed) arrangement [3], but I'm afraid that will have to wait for a while now - I have my EAS system, and am very happy with it.

References

1. ELF™ and Extended Low Frequency™ are trademarks of Long/Wickersham Labs
2. Bag End® is a trademark of Modular Sound Systems, Inc. (http://www.bagend.com)
3. Siegfried Linkwitz - LinkWitz Labs (http://www.linkwitzlab.com)
4. John L. Murphy, Tech Topics #13 - TrueAudio (http://www.trueaudio.com)

My thanks to Siegfried Linkwitz for some very useful pointers and other information sources used for this article, and to various others who have spurred my efforts along, and also supplied some useful information.

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999-2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright (c) 14 Jan 2000./ Updated 07 Jan 2004

The Audio Pages

Elliott Sound Products **Project 63**

Multiple Feedback Bandpass Filter

Rod Elliott - Elliott Sound Products

Introduction

The multiple feedback bandpass filter is a simple looking design, but it is difficult to calculate the values for a given set of parameters. These filters are useful for equalisation, analysis and other tasks such as the Sound to Light converter (Project 62) or even a fully functional Vocoder. For those who have not heard of the vocoder, it is a device that takes a music source as one input and vocals as the other, allowing a guitar, keyboard or complete ensemble to be made to speak or sing. The "speech" from a good vocoder is quite intelligible, and is "ear candy" of the very best kind for experimental musicians.

This is the first in a series of projects using this filter type, and I have included a small calculator programme to make it easier to determine the component values for different filter parameters.

Description

A schematic for the filter is shown in Figure 1. The source impedance must be low with respect to the input resistance, and normally these filters are driven from an opamp buffer. If a high impedance is used, it adds to the total input resistance, causing unpredictable centre frequency and response.

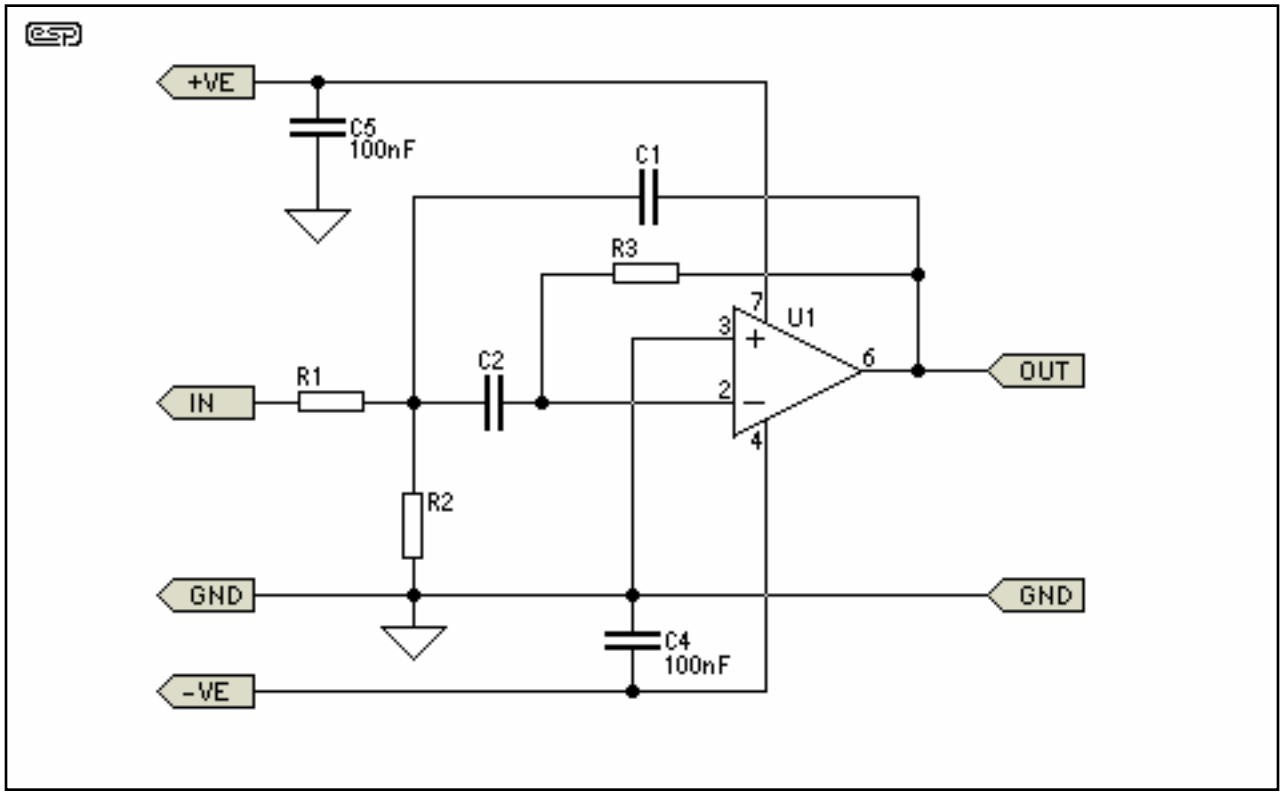
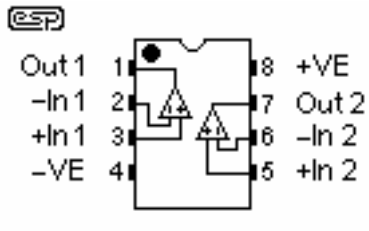


Figure 1 - Multiple Feedback Bandpass Filter

The resistor and capacitor values not shown are calculated from the formulae below, or by using the calculator program (see below for details). The opamp shown is a single device, but most commonly dual or quad opamps will be used for this kind of application. A resistance from the +ve input of the opamp is optional. If used, the resistor should be the same value as R3 to obtain minimum DC offset from the opamp output. A 100nF capacitor is highly recommended to bypass the non inverting input to earth for AC, and helps to reduce noise. If offset is not a problem for you, simply connect the non inverting input to the earth (GND) rail as shown. C4 and C5 are supply bypass capacitors, and should be used at each IC package. Ceramic capacitors are recommended for the most effective high frequency bypass. The opamp can be any common device for low frequencies, but at high frequencies (above about 2kHz) a high speed unit is required for best performance.



The pinout for a typical dual opamp is shown for reference. This is pretty much an industry standard, and nearly all dual opamps use this pin configuration. As always, I suggest that you download the data sheet for the device you intend to use to double check.

Octaves and Filter Q

With any bandpass design (or indeed any filter), one of the important parameters is Q, or 'Quality Factor'. The Q of a filter determines its bandwidth, and this is especially true of bandpass filters. Perhaps unexpectedly, bigger is not better. Most audio applications will require a maximum Q of about 4, which is suited to a 1/3 octave filter set. Few applications require closer filters than this, and to have 1/3 octave band filters covering the audio band requires 30 separate filters. Naturally, for even greater resolution this would increase dramatically.

At this point, a discussion about Q, bandwidth and even division of octaves is needed.

A filter with a Q of 10 (which is too high to be useful in most audio applications), has a bandwidth that is 1/10th of the centre frequency. Thus, a 1,000Hz filter with a Q of 10 has a bandwidth of 100Hz, measured at the -3dB frequencies on either side of the resonant peak. This is shown in Figure 2 for a filter having a centre (resonant) frequency of 123Hz. With -3dB frequencies of 116 and 149Hz, bandwidth is 33Hz, so Q is 132 / 33 = 4.

Note that beyond about 1.5 octaves either side of the resonant peak, the rolloff slope is 6dB / octave. This limits the usable range of the circuit in some respects, as the ultimate slope of 6dB / octave (20dB / decade) is only a first order filter response. Increasing the Q does nothing to solve this, the peak will be narrower and will extend more than the typical 8 to 10dB, but the ultimate slope is unaffected.

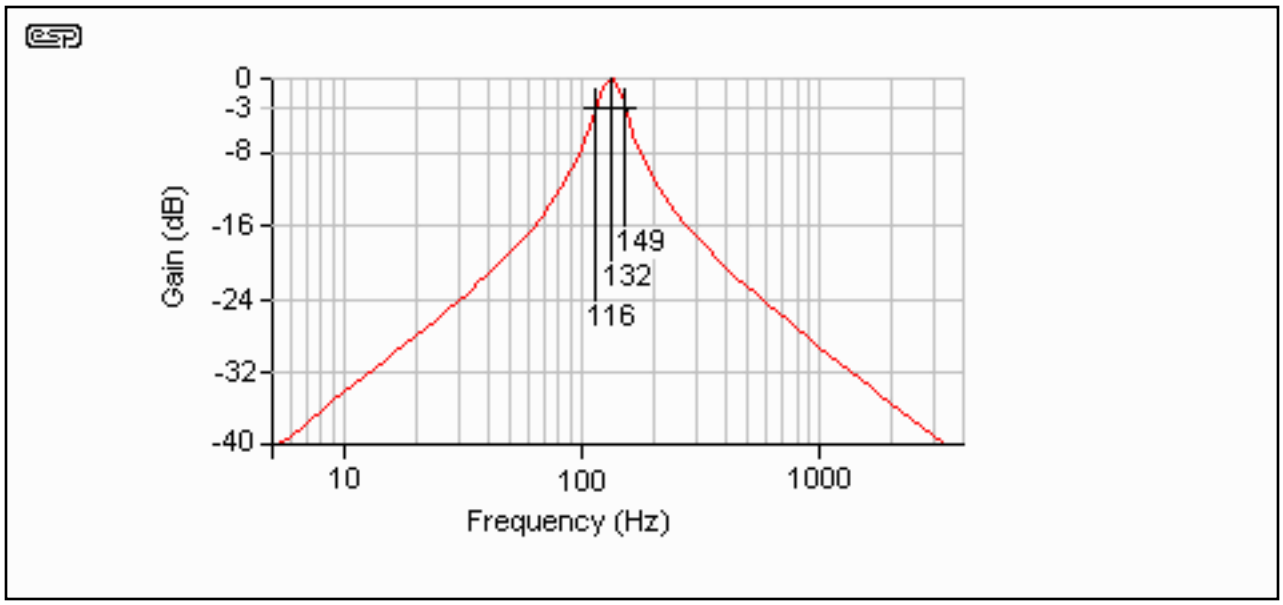


Figure 2 - Typical measurement of Q

Division of an octave is simple, once you see how it is done. An octave is the doubling (or halving) of a frequency, so for A440 (Concert Pitch), an octave above is 880Hz, and an octave below is 220Hz. All notes are "A", but in different octaves. If we want to divide an octave into 12 parts (the equally tempered scale, as used in most musical instruments) we will get 12 semitones. Using A220 as a starting point. we must divide the 220Hz bandwidth into 12 musically related frequencies. The trick here is that you can't divide 220Hz by 12 to find the interval, because our ears have a logarithmic frequency response.

Instead, we must find the 12th root of 2 (1.05946...). If you multiply 220 by 1.05946 twelve times, you will get 440 (well, near enough anyway). To divide an octave in half, we use the square root of 2 (tooth root of 2?) which is 1.414. So (for the sake of simplicity), a half octave division from 100Hz gives the halfway point at 141Hz, and the next octave at 200Hz.

To locate the 1/3 octave intervals, we use the 3rd root of 2 (1.26), so the frequencies will be 100Hz, 126Hz, 158.7Hz and 200Hz. By this means, an octave can be divided into as many frequencies as desired.

Calculating Frequencies

By far the easiest way is to use my program (download here), but since it is only applicable to Windoze machines, some of you will have to do it the hard way. Note that you will also need the Visual Basic 4 (VB4) runtime library, which can be obtained from Annoyances.org (easy) or the Microsoft support Website (less easy).

Based on the circuit above, select an appropriate capacitance first (both capacitors must be the same value), then these are the formulae ...

Input resistance	$R1 = Q / (G * 2 * \pi * f * C)$
Attenuator resistance	$R2 = Q / ((2 * Q^2 - G) * 2 * \pi * f * C)$
Feedback resistance	$R3 = Q / (\pi * f * C)$
Passband Gain	$G = 1 / ((R1 / R3) * 2)$
Centre Frequency	$f = (1 / (2 * \pi * C)) * \sqrt{((R1 + R2) / (R1 * R2 * R3))}$

This process can hardly be regarded as trivial, especially if there is a significant number of filters to design. The program also allows you to apply standard resistor values and

then calculate backwards to see how far out the final design will be from the design ideal. This applies to frequency, gain and Q. Capacitor selection will come with experience, but as a guide, the following table will help ....

$f_{(min)}$ (Hz)	$f_{(max)}$ (Hz)	Capacitance (nF)
20	80	330
80	300	82
300	1,200	22
1200	4,800	5.6
4,800	20k	1.5

Increasing capacitance decreases the resistance values, and especially with multiple filters, care must be taken to ensure that the loading on the input buffer opamp is not excessive. At the same time, low values of capacitance cause problems due to stray capacitance on the board, and resistor values become too high. The above is a reasonable set of values, and will give satisfactory results for most applications.

Equalisers and Analysers

For both equalisers and analysers, there are commonly accepted frequencies that are normally used. The exact frequencies depend on the octave division, the application and some degree of manufacturer preference, but nearly all share the basic octave boundaries which are based on a centre frequency of 1000Hz.

For an octave band EQ or analysis instrument (10 Band) the frequencies are usually as follows ...

31 63 125 250 500 1k0 2k0 4k0 8k0 16k

1/2 Octave band frequencies (20 Band) will be (typically) ...

31 44 63 87 125 175 250 350 500 700 1k0 1k4 2k0 2k8 4k0 5k6 8k0 11k 16k 20k

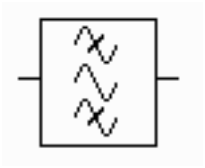
Finally, 1/3 Octave band instruments (30 Band) will typically follow this sequence ...

25 31 40 50 63 80 100 125 160 200 250 315 400 500 630 800 1k0 1k2 1k6 2k0 2k5 3k2 4k0 5k0 6k3

8k0 10k 12k 16k 20k

Naturally, it is possible to use wider bandwidth (and fewer filters) or narrower bandwidth and more filters. Although some 5 band 'graphic' equalisers have been made, they are of limited use for anything other than elaborate tone controls. Anything narrower than 1/3 Octave is rare, since the complexity of the filters increases for higher values of Q. This can get rather expensive, and in reality is of limited use for most applications in audio.

Projects Using MFB Filters



The symbol to the left is a standard representation of a band pass filter, and will be used in any future article or project using this building block. Following this description of the basic multiple feedback bandpass filter building block, projects will be offered that use it. The ones I am currently thinking of including are ...

- Graphic Equaliser (1) - a conventional boost / cut equaliser featuring constant Q at each frequency (see Project 84, a 1/3 octave subwoofer equaliser)
- Graphic Equaliser (2) - a frequency selective filter bank, with each band having its own volume control (project 64)
- Graphic Analyser - based on filters and LED VU meters, gives a display of the signal energy in each band (one day, perhaps)
- Vocoder - a proper vocoder circuit, using filters and voltage controlled amplifiers (VCAs) (maybe one day, maybe not - there is a lot of work involved)

The last two circuits will represent a considerable commitment in time and money, but you will be able to build a basic version of each first, and expand it later as your needs and/or funds allow.

References

1. Active Filter Cookbook, Don Lancaster (Howard W Sams & Co., Inc.) ISBN 0-672-21168-8
2. IC Opamp Cookbook, Walter G Jung (Howard W Sams & Co., Inc.) ISBN 0-672-20969-1

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright (c) Rod Elliott 18 Jul 2000

Reed Electronics Group
EDN: The Voice of the Electronics Engineer

EDN Asia | EDN China | EDN Japan | EDN Europe | EDN Australia

VOICE OF THE ELECTRONICS ENGINEER

Monday, May 29, 2006

- Current Issue
- Archives
- Tools & Services
- Blogs
- RSS
- About EDN
- Contact Us

Register for EDN Newsletters
Enter your e-mail addre
Register Change Profile

Technical Channels

Analog | Communications | Components/Interconnect | Computers & Boards | Design Ideas | Digital Den | EDA/IP | High-Speed Design | Power Sources/Controllers | Processors/Software/Tools | Programmable Logic/Memory | Test & Measurement

Zibb Beta Strictly Electronics

SEARCH >>

ADVERTISEMENT

Feedback Loop

Post a comment on this article

Free Print Subscription | Printer-friendly version / PDF | Email to a Friend

Design Idea

Bandpass filter features adjustable Q and constant maximum gain

Edited by Brad Thompson

Herminio Martínez, Joan Domingo, Juan Gmiz, and Antoni Grau, Technical University of Catalonia, Barcelona, Spain -- EDN, 3/3/2005

There is a self-evident mistake in ...
Ed Learned 06/04/05

Comparison of sensitivity figures: ...
Lutz v. Wangenheim 03/12/05

It is simply not true that "all of...
Lutz v. Wangenheim 03/10/05

Excellent !...
Dick Benson 03/03/05

Related Articles

- **The hot 100 products of 2005**
12/16/2005, EDN
- **Deconstructing gain and impedance from S11**
11/10/2005, EDN
- **One in a million**
07/07/2005, EDN
- **Crosstalk penalty in high-speed, serial-communication modules**
06/23/2005, EDN
- **Beyond the Spice model's dc and ac performance**
06/23/2005, EDN

FROM THE EDN Network

- **Taking Stock: The Unfolding Stock Option Investigation** Electronic News, 5/26/2006
- **Digital Eyewear Could Hurdle Smart Screen Barrier** Electronic Business, 5/26/2006
- **Battery packs get tougher, smarter** EDN, 5/25/2006
- **Six microwave instruments earn LXI certification** Test & Measurement World, 5/23/2006 6:21:00 AM

Applications such as audio equalizers require bandpass filters with a constant maximum gain that's independent of the filter's quality factor, Q. However, all of the well-known filter architectures—Sallen-Key, multiple-feedback, state-variable, and Tow-Thomas—suffer from altered maximum gain when Q varies. **Equation 1** expresses the second-order bandpass transfer function of a bandpass filter:

$$H_{BP}(s) = K\frac{\left(\frac{s}{\omega_0}\right)}{\left(\frac{s}{\omega_0}\right)^2 + \frac{1}{Q}\left(\frac{s}{\omega_0}\right) + 1},$$

EQUATION 1

where K represents the filter's gain constant. When the input frequency equals ω_O, the filter's gain, A_{MAX}, is proportional to the product, KQ. Thus, modifying the quality factor alters the gain and vice versa.

This Design Idea describes a filter structure in which K is inversely proportional to Q. Altering Q also modifies K, producing a magnitude-plot set in which the curves maintain the same maximum gain at the central frequency ω_O —that is, KQ remains constant. **Figure 1** shows the filter, which comprises a twin T cell with an adjustable quality factor and a differential stage. The differential stage comprises op amp IC_3 and resistors R_{5A} through R_{5D}. This stage outputs the difference between the filter's input signal and the twin-T network's output. Capacitors C_1 and C_2 are of equal value, $C=C_1=C_2$, capacitor C_3 equals 2C, resistors R_1 and R_2 are also equal and of value $R=R_1=R_2$, and R_3 equals $R/_2$. **Equation 2** describes the twin-T circuit's transfer-function response as a notch filter producing output V_{BR} (t):

ADVERTISEMENT

$$H_{BR}(s) = \frac{V_{BR}(s)}{V_{IN}(s)} = \frac{(RCs)^2 + 1}{(RCs)^2 + 4RC(1-m)s + 1},$$

EQUATION 2

Equation 3 describes the compete circuit's transfer function, a bandpass-filter response with output V_{OUT} (t):

Sponsored by:

News & New Products

- Japanese power-supply vendors see engineers as competitive advantages
EDN, 5/25/2006
- Frequency-agile WiMax transceiver matches client to band
EDN, 5/25/2006
- Germanium boosts thermophotovoltaic efficiency
EDN, 5/25/2006

More >>

Design Ideas

- Shift registers and resistors deliver multiphase sine waves
EDN, 4/13/2006
- Stealth-mode LED controls itself
EDN, 5/25/2006
- CPLD automatically powers itself off
EDN, 4/13/2006

More >>

Features

- Spread-spectrum clocking: measuring accuracy and depth
EDN, 5/11/2006
- Displays invade embedded-system space
EDN, 5/25/2006
- Dynamic packet classification enables multimedia traffic on WLANs
EDN, 5/25/2006

More >>

Columnists

- Charge your SAR-converter

$$H_{BP}(s)=\frac{V_{OUT}(s)}{V_{IN}(s)}=\frac{4RC(1-m)s}{(RCs)^2+4RC(1-m)s+1},$$ **EQUATION 3**

where m represents the twin-T cell's feedback factor. If you designate R_{XY} as the resistance potentiometer R_4 's upper terminal, Point X; the rotor as Point Y; and R_{YZ} as the resistance between the rotor and the bottom terminal, Point Z, you can express m as the quotient of **Equation 4** :

$$m=\frac{R_{YZ}}{R_{XY}+R_{YZ}}=\frac{R_{YZ}}{R_4}.$$ **EQUATION 4**

Comparing **Equation 3** with the respective normalized transfer functions of a bandpass filter, **Equation 1**, **Equation 5** expresses the central frequency of the filter, ω_O, coincident with the transmission zero of the twin-T network:

$$\omega_0=\frac{1}{RC}.$$ **EQUATION 5**

Equations 6 and **7**, respectively, give quality factor Q and gain constant K:

$$Q=\frac{1}{4(1-m)};$$ **EQUATION 6**

$$K=\frac{1}{Q}=4(1-m).$$ **EQUATION 7**

The maximum gain, A_{MAX}, at $\omega=\omega_O$, always remains constant and equal to 1 (0 dB) and is independent of Q. The minimum quality factor is ¼ for m=0, which corresponds to the potentiometer's

inputs
By Bonnie Baker, EDN, 5/11/2006

- **Hidden schematic**
By Bruce Archambeault, PhD, EDN, 5/25/2006

- **The PC is the workstation**
By Maury Wright, Editor in Chief, EDN, 5/25/2006

More >>

Blogs

- **So Long (But Not For Long)....** - 05/25/2006
- **'Flash'y HDDs: Has Their Time Finally Come?** - 05/25/2006
- **Risky Business: Microsoft's Direct3D Backwards-Compatibility Discard** - 05/25/2006

More >>

Affiliate Network

Tool Center

- **Software**
- **Catalogs & Brochures**
- **White Papers**
- **eLiterature**
- **ECards**

Events

- **Agilent Technologies Power-of-Two Seminar**
Dates: *4/10/2006 - 5/31/2006*
Location: *Tempe, AZ, San Diego, CA, and many other locations*
- **Microchip's 16-bit Microcontroller Seminar**
Dates: *4/18/2006 - 6/27/2006*
Location: *Montreal, Boston, Los Angeles, and many more*
- **Texas Instruments'**

rotor connected to ground. The maximum gain is theoretically infinite, but, in practice, it's difficult to achieve a quality factor beyond 50. In most applications, Q ranges from 1 to 10.

Figure 2 shows the filter's magnitude and phase Bode plots for the frequency-notch output V_{BR} (t) (available at IC_1's output) for values of m from 0.1 to 0.9. **Figure 3** shows Bode plots for the filter's bandpass output, V_{OUT} (t), for the same values of m. In both graphs, frequency f_O equals 1061 Hz. To minimize frequency-response variations and improve response accuracy, you can build the filter with precision metal-film resistors of 1% or better tolerance. Likewise, use close-tolerance mica, polycarbonate, polyester, polystyrene, polypropylene, or Teflon capacitors. For best performance, avoid carbon resistors and electrolytic, tantalum, or ceramic capacitors.

PDF Version

Check out our Best of Design Ideas section!

Sponsor Links

Ads By Google

Microwave & RF filters
Filters, bandpass, ceramic, LC Crystal, cavity, duplexers
www.anatechelectronics.com

Bandpass Filter
The Spot To Find It! It Is All Here.
ebandpassfilter.com

Bandpass Filters
Low Loss LC Filter Modules. Free Samples. Order Online.
www.coilcraft.com

DaVinci Technology Technical Seminar
Dates: *5/31/2006 - 5/31/2006*
Location: *Indianapolis, IN*

More >>

ADVERTISEMENT

Reed Business Information.

1997-2006 Reed Business Information, a division of Reed Elsevier Inc. All rights reserved.
Use of this web site is subject to its Terms of Use | Privacy Policy

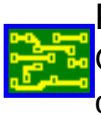
Embedded Secure Document

The file *http://www.cirrus.com/en/pubs/appNote/AN048Rev2.pdf* is a secure document that has been embedded in this document. Double click the pushpin to view.

Elliott Sound Products **Project 37**

Minimalist Discrete Hi-Fi Preamp

Rod Elliott (ESP)


PCBs are available for this project. Click the image for details.
Click here to see a photo of the prototype - the actual boards have a silk screened overlay, but are otherwise identical.

Introduction

A preamp designed for the minimalist, and having no frills at all is the design goal for this project. It is designed as a preamp for the Death of Zen (DoZ) Class-A power amp (Project 36), and has very low levels of noise and distortion, in a minimum component count, fully discrete circuit.

This gain module can be used as the basis for any preamp - performance is exemplary, with low noise, wide bandwidth, and it sounds extremely good indeed.

You can add as many inputs as you need, and the only controls are volume and input selection. A power switch is also a good idea, but if you wanted to you could leave the preamp running all the time. This is not necessary, as it will reach a stable operating condition within a few seconds, and will not change its characteristics to any audible degree. An advantage of a power switch is that a +30V signal can be used to switch the power on the power amplifiers, so only a single switch is needed for the system. This can be expanded to switch power to the entire music centre, including CD player(s), tuner, etc.

The preamp can also be used with other amplifiers, and can drive an impedance of 2k Ohms with ease. Although shown using a single supply (to easily match the DoZ power amp), it can also be operated with a dual +/- 15V supply if desired, although this is not really recommended (and is not a trivial undertaking, either) partly because of the possibility of reverse biasing the output capacitors (which cannot be omitted), and partly because of other modifications that are needed for it to work properly.

As shown, frequency response is absolutely flat from 10Hz to 100kHz, without any frequency stabilisation required. The table below shows the rated performance of the gain module.

Distortion	< 0.01%
Output Voltage	6.0 V RMS
Output Impedance	< 100 Ohms
Minimum Load	2k Ohms
Frequency Response	10Hz - 100kHz (-0.5dB)
Voltage Gain	10dB nominal
Supply Voltage	30V
Supply Current	<10mA

Description

The circuit for the amplifier module is shown in Figure 1, and as can be seen is very simple. Two of these can easily be built on a piece of Veroboard, although the PCB makes it a lot easier. This preamp relies on a completely hum free power supply, as it is not an opamp, and cannot reject supply noise as well. This is not to say that the power supply rejection is especially bad, just that it is not as good as an opamp. The power supply is shown below, and this will have a ripple and noise of less than 10uV if built as specified.

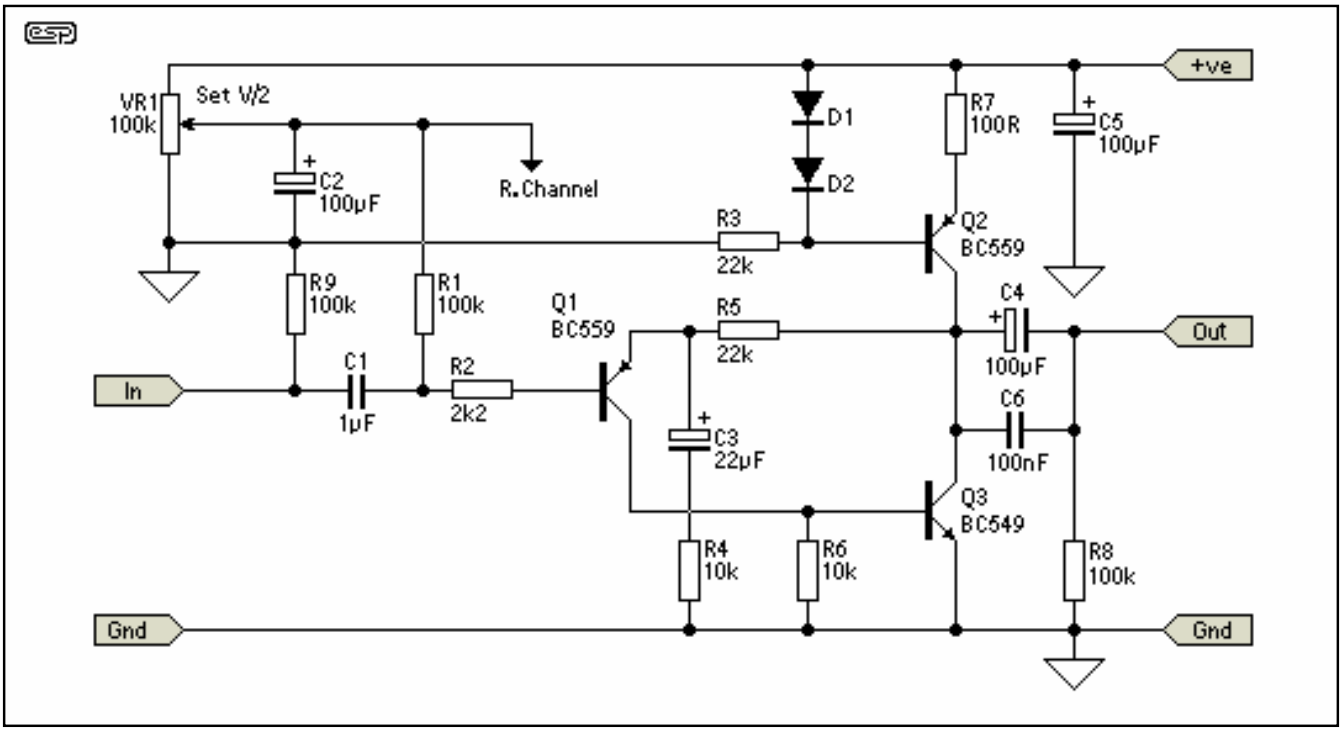


Figure 1 - Gain Module

D1 and D2 are 1N4148 (or 1N914), and together with Q2 form a current source providing a bias current of about 7mA. This stage is not an amplifier, but an active (and very linear) load, allowing the amplifying transistor Q3 to provide a high gain with excellent linearity. Feedback is applied through R5, with R4 setting the AC feedback ratio and thus the voltage gain of the amp. The transistors Q2 and Q3 operate at about 100mW and will get slightly warm in operation.

The trimpot VR1 is used to set the voltage at the collectors of Q2 and Q3 to 1/2 the supply voltage (15V), and a single unit can be used for a stereo pair with no interaction or other undesirable side effects. A 100k resistor is then used to supply bias for each preamp module. If you really don't want to use the pot, you can use a 10k resistor from the +ve of C2 to earth, and a 22k resistor to the supply. This will not allow you to set the quiescent voltage as accurately, but will be acceptable for normal use. R8 ensures that the output capacitor is properly charged with no output connected, and can be left out if the output is connected directly to the volume control as shown below.

The distortion of the preamp is reasonably consistent with frequency, and is less than 0.01%. I measured about 0.0075% - this is just above the residual of my signal generator (0.006%), so it is presumably much better than this, but I can't measure it. A rough guess would be the difference between the two, giving 0.0015%, but I prefer to err on the side of caution.

Frequency response is flat to within -0.2dB from 10Hz to 100kHz, and even at 100kHz, square wave performance is almost perfect, showing slight rounding and no ringing or instability of any kind. Measured output impedance is 84 Ohms, and the circuit can drive 6 Volts RMS into a 2.2k Ohm load easily. These are excellent figures, considering the simplicity of the circuit. The gain is nominally 3.2 (10dB) as shown, and is easily changed by varying R4 - increase the value to

decrease gain and vice versa. Because of the simplicity of the circuit, it can be made very small, even on Veroboard. My prototype measured less than 20mm x 40mm complete.

The preamp modules provide the only gain stage, and may be bypassed for high level signals (such as CD players) that have enough signal to drive the power amp directly. This is a great advantage for those who want the minimum number of components in line with the signal. The circuit cannot be direct coupled, however, since it operates on a single supply. This means that polyester caps can be used at the inputs, but for low impedance outputs, electrolytic capacitors must be used. These may be bypassed with 100nF polyester caps, but the frequency response is not altered to any significant degree (not measurable at 100kHz), and there is no measurable decrease in distortion with the bypass caps.

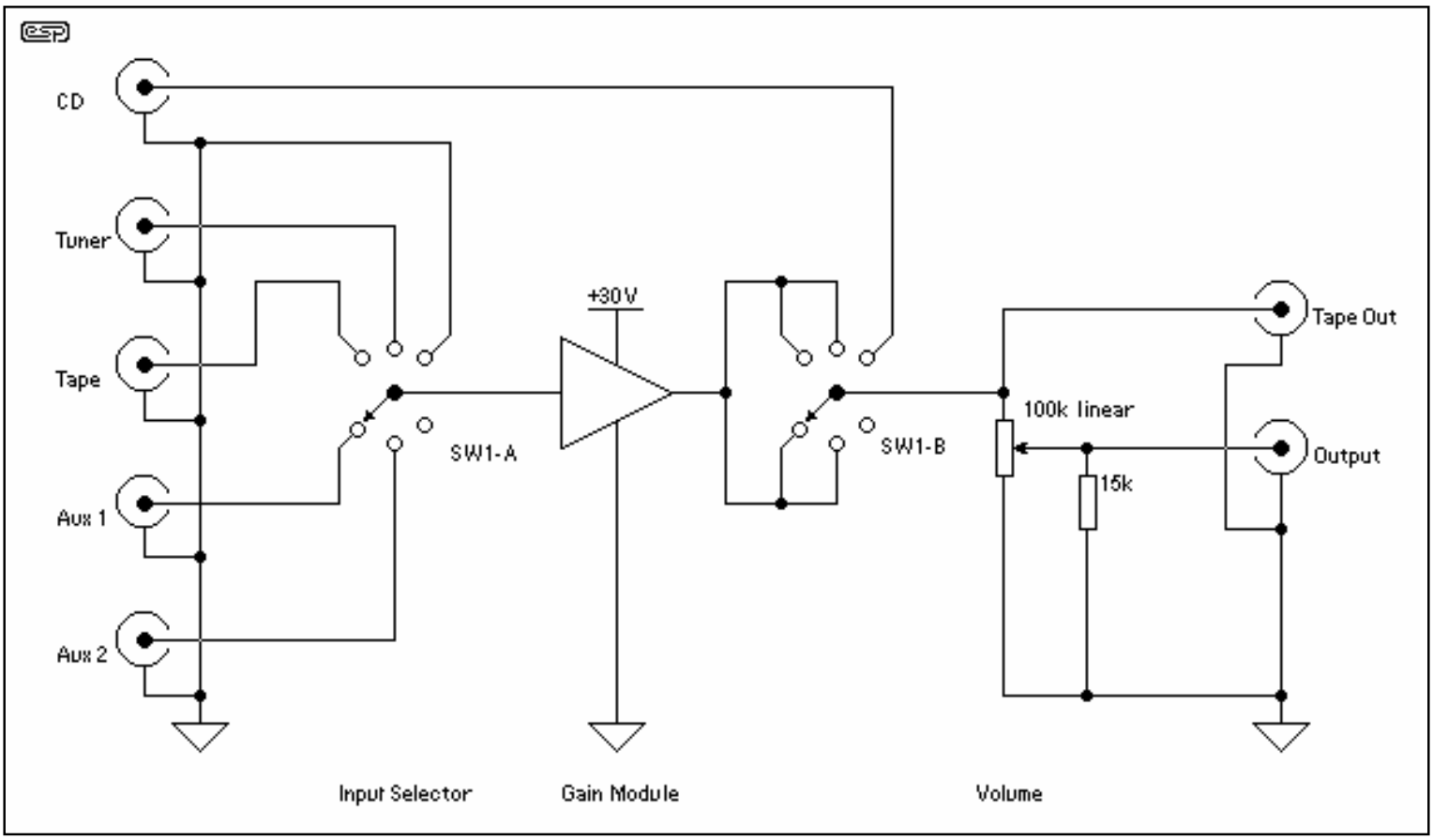


Figure 2 - Preamp Wiring Suggestion

Figure 2 shows one channel of a suggested wiring for the preamp. This is duplicated exactly for the other channel, and the switch and volume control are common to both. The CD input bypasses all electronics and is applied directly to the volume control, and the low level inputs use the gain of the modules to bring their level up to that of the CD player. This introduces a bit more switching, but many audiophiles will prefer this method. Note that when switched to CD, the input of the preamp is earthed to prevent any noise pick-up.

The switching requires a 4-pole, 5 position rotary switch for both channels, and a dual-gang pot. I strongly suggest that a linear pot is used, with the resistor as shown. Although this produces a

variable impedance to the source, it is a better approximation of a log pot than those you can buy at normal outlets, and has much better tracking between channels. If you can get a conductive plastic logarithmic pot, do not use the resistor, as these are generally much better than the normal carbon film types, and the resistor will ruin the log curve.

In some cases - due to an amplifier with unusually low input sensitivity for instance - an output gain stage might be needed. This would typically use another gain module using the circuit above, but generally with lower gain. It is expected that a gain of 2 (6dB) will be enough for any amplifier, since this will provide an output of 4V RMS for a CD players typical 2V output. I have found that with typical sources, a preamp gain of about 10dB is sufficient to drive a typical power amplifier, so the additional gain will probably not be needed.

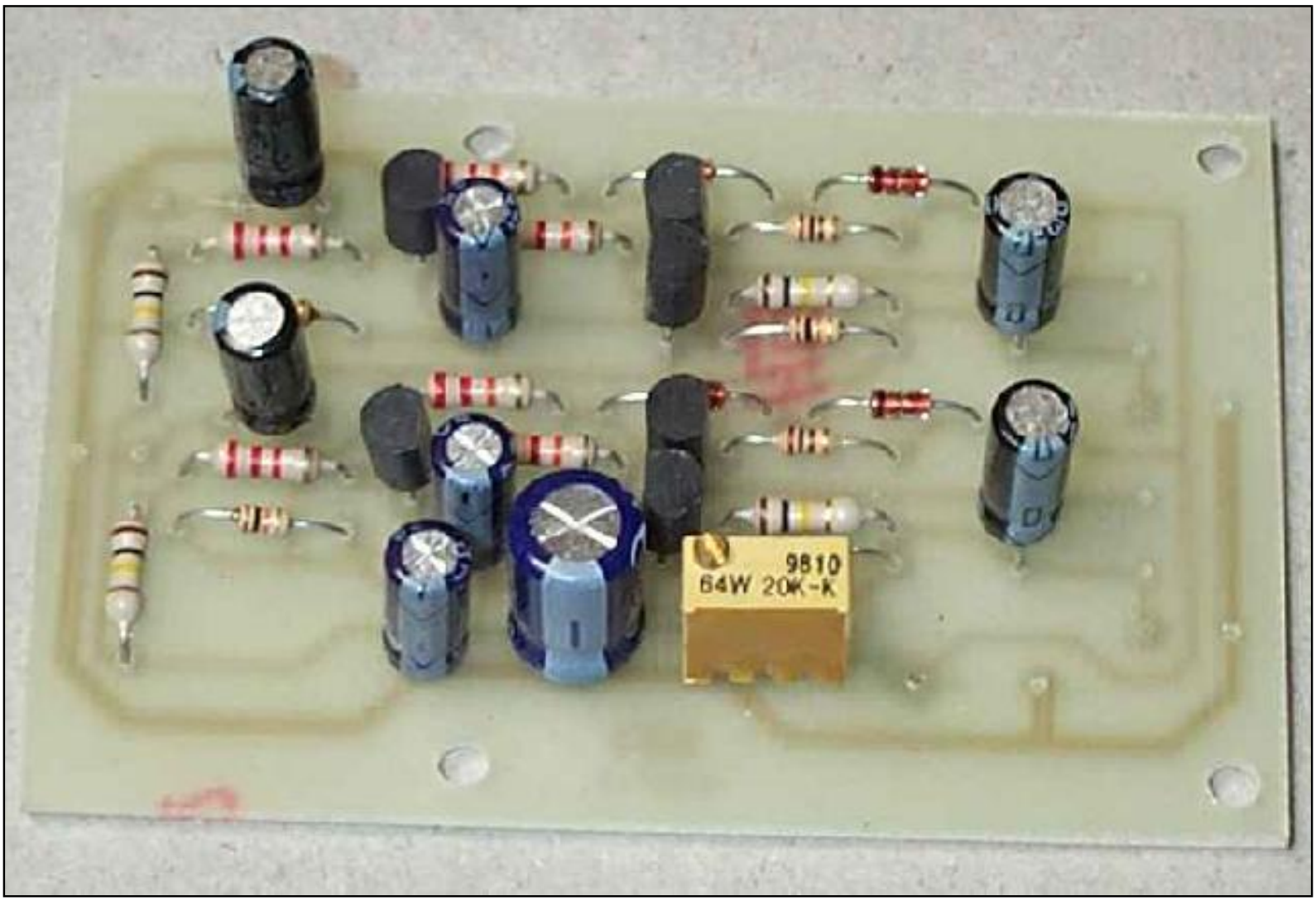


Photo of Prototype

The photo shows what the final unit looks like. Since this was taken of the prototype board, the silk screen component overlay is not present, and I used "ordinary" resistors for all my initial tests. Better noise performance will result from using metal film resistors.

Power Supply

The power supply must be free from hum and noise, and the circuit shown in Figure 3 should be used. Do not use 3 terminal regulators, as the applied voltage is too high, and they will fail. A discrete circuit is (usually) not quite as good, but is simple to build and reliable. The shunt regulator shown here will be hard to beat with any circuit though, mainly because of the filtering used, although the regulator does dispose of a lot of hum by itself.

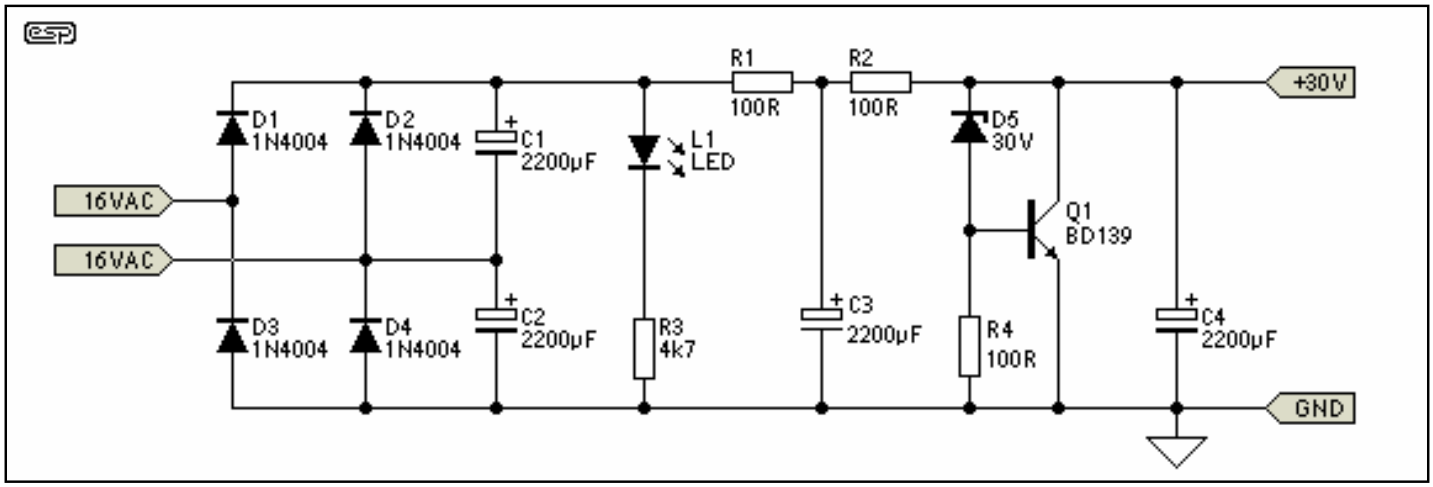


Figure 3 - Power Supply Circuit

This circuit is suitable for two gain modules as shown. If you want to use more, then R1 and R2 will have to be reduced in value. With a 40V input and a total of 200 Ohms, maximum current is 50mA for 30V output. I would suggest that a minimum of 20mA flows in the shunt regulator, so if you wanted to run 4 gain modules, reduce the value of R1 and R2 to about 82 Ohms. You can expect a very small increase in noise, but this will not be audible. The Zener diode must be a 30V device, and should be rated at 1 Watt. Your local electronics supplier will be able to tell you what type number you need. All electrolytic capacitors should be rated at 63V. If you cannot get a 16V transformer, you could use a unit with up to 20V AC output, but you will have to re-calculate the value of R1 and R2.

You can add decoupling between channels to prevent any possibility of crosstalk, but even without this it should be possible to obtain excellent channel separation, since the circuit operates in Class-A and does not vary the supply current with signal (as do opamps and many other circuits). The power transformer does not have to be high power - 20VA is more than enough due to the low current drain (about 50mA).

Because this is a shunt regulator, it has "automatic" protection against short circuits at the output. You should not test this though, because the discharge from the output capacitor will cause a very high peak current, and may damage the cap. Why the shunt regulator? Because in the interests of greatest simplicity, this is the ultimate. Since there is a resistor / capacitor "pi" filter, hum and noise are very low, and should be less than 10uV RMS under normal operation. The shunt regulator also acts to reduce hum, since this is "seen" by the circuit as a varying voltage, and the shunt circuit will try to compensate by modulating its shunt impedance to

minimise the fluctuating voltage.

The series 100 Ohm resistors (R1 & R2) carry the preamp + shunt current, which will typically be around 50mA, so dissipation is quite low. 5W resistors should be used to ensure reliability, especially in the event of an accidental short circuit on the output. In this event, dissipation will be about 7 to 8W, so the resistors will get very hot. The BD139 shunt transistor will operate at about 50mA when the supply is not loaded by the preamp modules, and will dissipate about 1.5 Watts. This must be mounted on a suitable heatsink, having a thermal resistance of no more than 10°C / Watt. This will keep the transistor nice and cool under worst case conditions, with a typical maximum temperature rise of 15°C.

If readers want an alternative with even less hum (the above circuit has about 5uV), let me know, and I'll publish the circuit of a shunt regulated supply with capacitance multiplier. I have only simulated this so far, but output noise is virtually unmeasurable. I don't really think that this is necessary, because even if all 5uV of hum from the circuit above found its way to the output (which it doesn't, by a wide margin), the signal to noise referred to 1V is 106dB (hum component only). This is unlikely to be achieved by the circuit itself, so the hum component can safely be ignored as completely inaudible.

As always, be very careful with the mains wiring. If possible, use a 16V AC power pack and mount it remotely from the preamp so that there is no possibility of inducing hum into the circuit from the transformer. Using the full wave voltage doubler as shown will provide about 40V minimum before regulation. This is more than enough to ensure that the preamp supply remains stable.

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright (c) 13 Nov 1999

METER SHUNTS & MULTIPLIERS

How do you turn a 0-1mA panel meter into one that reads 1A full scale? Or one that reads 100mA full scale? Or into a voltmeter that reads 0-100V? It's not hard to adapt almost any panel meter in this way, once you understand the concepts of current shunts and voltage multipliers.

Let's look at current shunts first. A current shunt is basically just a resistor that is connected directly across the meter movement you want to adapt — so that it *shunts* a significant proportion of the total current around it. This is the way to adapt a meter into one that measures **higher currents**.

Fig.1 shows the basic idea. M is our basic meter movement, while Rs is the shunt resistor. As you can see it's connected directly across the meter's terminals, so that all of the current the meter can't handle is able to pass through the shunt instead.

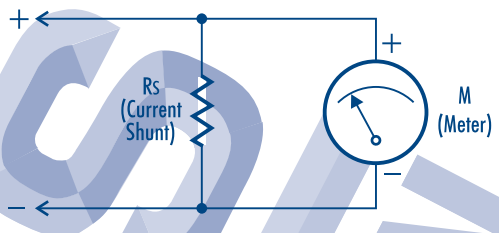


Fig.1: A current shunt is connected in parallel with the meter.

How do you work out the value of the shunt? It's easy once you understand exactly what a shunt does. Let's say we want to adapt our 0-1mA panel meter into one that measures 0-1A. That means that we want the meter to indicate full-scale deflection (FSD) when 1A is flowing — right? But the meter itself will read FSD when it has only 1mA flowing through its coil, of course. So the way we can make this happen is by arranging for the shunt to take **the remaining 999mA** of current. Or putting it in more general terms,

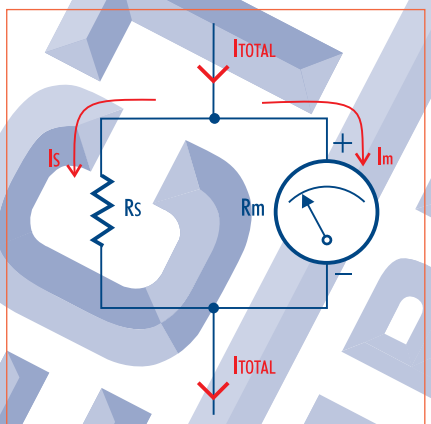


Fig.2: The shunt bypasses a large proportion of the total current around the meter, leaving it with only the current it can handle...

Is = ITOTAL - Im

where Is is the current taken by the shunt, ITOTAL is the total current equal to the new effective FSD we want to give the meter), and Im is the meter's own basic FSD.

So that's the easy way to work out how much current we want our shunt to carry, compared with the meter itself.

As another example, how much current would the shunt need to carry to turn our 0-1mA meter into one reading 100mA? That's right, 100mA minus 1mA, or 99mA.

Now what we have to do is work out the value of the shunt resistor in order to make it carry that proportion of the current. If you look at Fig.2, you see that because the meter and shunt are connected in parallel, they always have the same voltage across them. So how do we persuade say 999mA of the total current to go through the shunt, and only 1mA through the meter?

Right again — simply by making the resistance of the shunt 1/999th that of the meter. That's because in a parallel circuit, the current will always divide in proportions which are the reciprocal of the path resistances. Or in general terms here:

$$\frac{Is}{Im} = \frac{Rm}{Rs}$$

or in a more convenient form for our present purposes,

$$Rs = Rm \div \left(\frac{Is}{Im}\right)$$

In other words, the value of our shunt resistor Rs can be found by dividing the meter resistance Rm by the desired ratio of shunt current to meter current.

So in our example of turning the meter into one reading 0-100mA, the shunt would need to have a resistance of Rm divided by 99.

Meter resistance

Clearly, though, you need to know the resistance of the meter itself in order to work out the value of the shunt you're going to need.

Luckily the internal resistance of most meters is specified when you buy them new. For example the resistance of the MU-45 type 0-1mA meter movement sold by Electus (QP-5010) is specified as 200 ohms, while the more sensitive 0-50uA model (QP-5012) has a specified resistance of 3500 ohms. So it's not hard to work out the value of shunts needed to adapt either of these meters to read higher currents.

But what if you have a meter movement salvaged from a piece of equipment, and you don't know its resistance? Or even its current sensitivity? (It might already have an internal shunt, for example.) The simplest way is to measure it — but don't just slap a multimeter across it, because that could damage the meter by passing too much current through it. The best way is to use the simple test circuit shown in Fig.3.

Here the meter is connected to a 12V battery via a DMM (set to a suitable current range) and adjustable resistor RV1, which is carefully adjusted until the meter is just giving its full-scale reading. The DMM will then be showing the current needed to produce this current FSD — in other words, the meter's sensitivity.

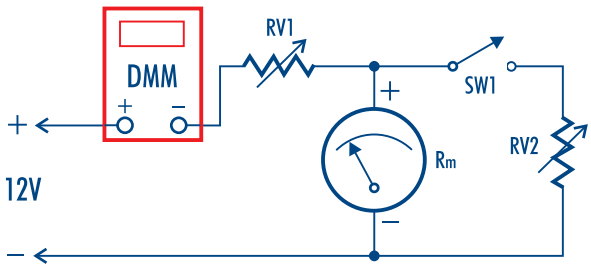


Fig.3: A simple circuit used to measure a meter's internal resistance.

ELECTUS DISTRIBUTION

Now switch SW1 is closed, to connect the second adjustable resistor RV2 directly across the meter, as a shunt. RV2 is then adjusted until the meter reading is reduced to **exactly half scale**, while monitoring the total current via the DMM, and if necessary adjusting RV1 to maintain it at the original FSD figure.

When the meter reading is exactly 50%, the value of RV2 will be equal to the meter's own resistance, because the current is being split equally through the two paths. So you can then disconnect the battery and use the DMM to measure the value of RV2 (by itself), to learn the meter's resistance.

Note that even if the meter should have a current shunt built inside it already, this test will still give you its *effective* sensitivity and resistance — **including** any internal shunt. So the figures you get will allow you to work out the value of any new shunt you need to adapt it for measuring higher currents, regardless of what may be inside the meter case.

Practical shunts

Because a current shunt is generally designed to take most of the current, leaving only a small amount to pass through the meter, most shunts need to have quite low resistance values. For example the shunt to adapt our 1mA/200Ω meter into one reading 0-1A needs a resistance of only 200Ω/999, or 0.20Ω. Similarly the shunt needed to adapt it for reading 0-100mA would need a resistance of only 200Ω/99, or 2.02Ω.

And these values need to be quite accurate, because the accuracy of the resulting 0-1A or 0-100mA meter depends directly on the shunt taking 999/1000ths or 99/100ths of the total current, respectively.

So practical shunts often have to have quite low values, yet be quite accurate. And of course they have to be able to carry the necessary current level (i.e., virtually all of the current being measured), without overheating.

For this reason, current shunts used to adapt meters for measuring very high currents are generally made from stout strips of nichrome or some similar metal with a very low temperature coefficient of resistance (tempco), carefully filed or machined until they reach the exact resistance needed.

This isn't very practical for hobbyist use. But if you only need moderately low shunt values, say for adapting a meter to read 0-10mA or 0-100mA, you can often achieve the value you need simply by using parallel combinations of standard low value resistors — preferably the metal film type, for stability.

For example to get a shunt of 2.02Ω, to adapt an Electus QP-5010 meter into one measuring 0-100mA, you could use a combination of six low value 1% metal film resistors in parallel: four of 10Ω value (RR-0524), one of 20Ω (RR-0531) and one of 22Ω (RR-0532). This combination will give a value very close to 2.02Ω.

You can use the same idea to make practical shunts for most common meters, to adapt them for measuring current up to about 500mA or so. It's really only when you want to make them measure significantly higher currents that you need to worry about pieces of nichrome wire or metal strip.

MULTIPLIERS

Now let's look at **multipliers**, which are used to adapt meters for measuring higher *voltages*. Unlike a current shunt, which as we've seen is wired in parallel with the meter, a multiplier is a resistor connected in *series* with the meter.

To understand how a multiplier works, first think about the meter itself. As we've already seen, a meter has two key parameters: its *sensitivity*, usually specified in terms of the current needed to indicate FSD, and its *internal resistance*.

Now using Ohm's law, it's easy to work out the voltage drop across the meter when it's carrying its FSD current:

$$V_m = I_m \times R_m$$

So when our 0-1mA meter with a resistance of 200Ω is reading FSD, for example, it will have a voltage drop of (1mA x 200Ω) or 200mV. Similarly a 50uA/3500Ω meter like the Electus QP-5012 will have a voltage drop of (50uA x 3500Ω) or 175mV.

These figures are sometimes called the **voltage sensitivity** of the meter movements, because they basically show the voltage that needs to be applied directly across the meter to produce its full-scale reading. As you can see the voltage sensitivity of most meter movements is typically quite low: 200mV or less.

To adapt such a meter for measuring much higher voltages — say 0-20V, or 0-50V — the simplest approach is to

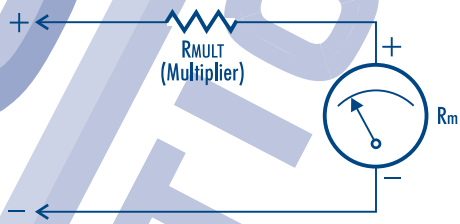


Fig.4: A multiplier is connected in series with the meter, not across it.

connect a resistor in **series** with it, to drop the excess voltage. This is exactly the idea of a *multiplier* resistor: it effectively 'multiplies' the voltage sensitivity of the basic meter to allow measuring higher voltages, without affecting its current sensitivity.

But how do we work out the value of the multiplier resistor? It's not hard.

First, work out the total resistance necessary to pass the FSD current of the meter, knowing the full-scale voltage we want our voltmeter to read:

$$R_T = V_{FSD} / I_m$$

where R_T is the total resistance, V_{FSD} is the FSD voltage we want for our voltmeter, and I_m is the meter's current sensitivity.

For example, let's say we want to turn our 0-1mA meter into a 0-50V voltmeter. R_T will need a value of 50V/1mA, or 50kΩ — right?

Similarly if we want to use a 0-50uA meter as a 0-20V voltmeter, the value of R_T will need to be 20V/50uA, or 400kΩ.

Now the meter itself has its own internal resistance, of course. And this will form part of the total resistance, whether we like it or not. So in order to find the required value of multiplier resistance as accurately as possible, we need to *subtract* the meter resistance from the total resistance we just worked out:

$$R_{MULT} = R_T - R_m$$

So because our 0-1mA meter has a resistance of 200Ω, the actual value of multiplier we'll need to turn it into a 0-50V voltmeter will be (50kΩ - 200Ω), or 49.8kΩ. Similarly the multiplier needed to turn our 50uA/3500Ω meter into a 0-20V voltmeter will be (400kΩ - 3.5kΩ) or 396.5kΩ. Get the idea?

ELECTUS DISTRIBUTION

Needless to say these multiplier values again need to be quite accurate, because the accuracy of the resulting voltmeter will depend on them quite heavily. Luckily because multiplier resistor values are much higher than those for current shunts, it's generally much easier to make up the value you need using standard 1% metal film resistors — this time series combinations.

For example to make up our 49.8kΩ multiplier, you'd use say a 47kΩ resistor (RR-0612) and a 2.7kΩ (RR-0582) in series, giving very close to the correct value. Similarly to make up the 396.5kΩ multiplier you'd use say two 180kΩ (RR-0626) resistors and a 36kΩ (RR-0609) in series, again giving very close to the correct value.

If you work out the exact figures, you'll find that they're within 0.2% or so of the correct figure. There's no point in striving for higher accuracy, because meter movements themselves are generally only rated to an accuracy of 2.5%, and metal film resistors have a tolerance of 1%.

When you want to make the meter into a voltmeter for high voltages, it's actually a good idea to split the multiplier into two or more physical resistors so they share the voltage drop. Metal film resistors are generally not designed to drop more than 200V or so, and they may break down if more than this is applied. So bear this in mind too when you're choosing the resistor values.

Choosing a meter

It's not difficult then, to turn almost any meter into a voltmeter, simply by connecting a suitable multiplier resistance in series. But does this mean that we can choose any meter we like for use as a voltmeter — is a 0-1mA meter just as good as a 50uA meter, given the appropriate multiplier?

No, for a very simple reason. In most circuits when we want to measure a voltage, we want do so while causing the absolute minimum disturbance to the circuit itself. And not surprisingly there will be less disturbance when less current needs to be drawn by the meter. So if a choice of meters is available, it's always better to use the one that needs less current: say the 50uA meter instead of the 1mA meter.

Remember when we worked out those multiplier values, for example. To turn the 1mA meter into a 0-50V voltmeter we ended up with a total resistance of 50kΩ, while for the 50uA we needed a total resistance of 400kΩ just to produce a 0-20V voltmeter. If we wanted to use the 50uA meter to read 0-50V, we'd end up with a total resistance of 1MΩ — much higher than 50kΩ, and therefore much less likely to disturb the circuit we're measuring in.

On the other hand for use as a **current** meter, it's generally better to use a less sensitive meter like a 1mA movement. This is because these movements have a lower internal resistance (like 200Ω instead of 3500Ω), so they will generally need a lower shunt resistance to adapt them for higher current measurement. And the lower the shunt resistance, the lower the voltage drop or 'voltage burden' that will be produced in the circuit when the meter and shunt are being used. Remember that to measure a current, the circuit must be generally be *broken* to connect the meter in series — so in this case, it's the voltage drop of the meter and shunt that will cause the disturbance.

Digital panel meters

Until now, we've been looking at what you need to adapt traditional moving-coil (or moving iron) meters so they will measure higher currents or voltages. But more and more, you might prefer to use a digital panel meter (DPM) instead. They're not much dearer nowadays, after all.

So what's different about DPMs? Can you use current shunts and multipliers with them, in the same way? Not quite, but there are similarities.

The main thing to realise is that a DPM is basically a digital voltmeter, not a current activated device like a meter. Most of them have a voltage sensitivity of 200mV, which is very similar to that of a meter movement, but on the other hand like a DMM they tend to have an internal resistance of 10MΩ or higher — much higher than a meter movement. As a result of this much higher resistance, we need to change the way we adapt a DPM for either current or voltage measurement. But often it makes things easier rather than harder.

When you want to measure a current, for example, you don't have to worry about the DPM's internal resistance any more. It's now so much higher than the shunt that you can ignore it. All you need to do is work out the value of 'shunt' resistor necessary to drop the DPM's full-scale reading voltage, when the desired full-scale current is flowing through it (Fig.5). So by Ohm's law:

$$R_S = V_{FSR} / I_{FSR}$$

Where R_S is the required shunt resistor value, V_{FSR} is the DPM's voltage sensitivity for full-scale reading, and I_{FSR} is the full-scale current reading we want to give it.

So to adapt either the Electus QP-5570 or QP-5580 DPMs (both of which have a voltage sensitivity of 200mV) to read say 0-100mA, for example, you'd need a shunt resistor of (200mV/100mA) or 2.00Ω. Similarly to make them read 0-10mA, you'd need a shunt of 20.0Ω. As you can see, it's even easier to work out the shunt value than it was before. The values are likely to be 'rounder numbers' too, and easier to achieve as a result.

There's even more reason to be accurate in achieving the right shunt value, though, because DPMs are much more accurate themselves than meter movements — typically 0.1%. So it's worth going to a bit more trouble to get the correct shunt value, to make sure that this higher accuracy isn't degraded.

When it comes to adapting your DPM for measuring higher voltages, though, things are a little more complex. In most cases we can't use a simple series multiplier resistor any

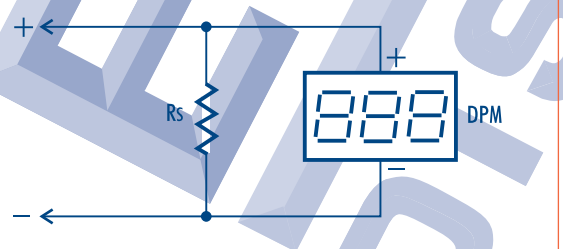


Fig.5: A shunt is also used to adapt a DPM for current measurement, but its value is easier to work out...

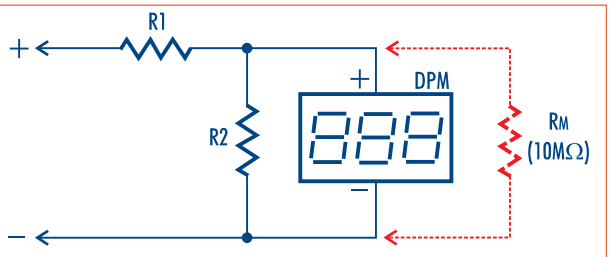


Fig.6: For voltage measurement, though, you need a resistive divider rather than a simple multiplier.

ELECTUS DISTRIBUTION

This table will save you having to work out common shunt and multiplier values for the meters and DPMs stocked by Electus.

ELECTUS METER	SHUNT VALUE (FOR CURRENT RANGE)					MULTIPLIER VALUE (FOR VOLTAGE RANGE)					
	10mA	20mA	50mA	100mA	500mA	5V	10V	20V	50V	100V	500V
QP-5010 (0-1mA/200Ω)	22.2Ω	10.53Ω	4.08Ω	2.02Ω	0.40Ω	4.80kΩ	9.80kΩ	19.8kΩ	49.8kΩ	100kΩ	500kΩ
QP-5012 (0-50uA/3500Ω)	17.6Ω	8.77Ω	3.50Ω	1.75Ω	0.35Ω	96.5kΩ	197kΩ	397kΩ	1MΩ	2MΩ	10MΩ
QP-5570 (LCD DPM. 200mV/10MΩ)	20.0Ω	10.0Ω	4.00Ω	2.00Ω	0.40Ω	R1= 1MΩ R2= 41.8kΩ	R1= 1MΩ R2= 20.5kΩ	R1= 1MΩ R2= 10.1kΩ	R1= 1MΩ R2= 4.00kΩ	R1= 1MΩ R2= 2.00kΩ	R1= 1.53MΩ R2= 612.2Ω
QP-5580 (LED DPM. 200mV/10MΩ)	20.0Ω	10.0Ω	4.00Ω	2.00Ω	0.40Ω	R1= 1MΩ R2= 41.8kΩ	R1= 1MΩ R2= 20.5kΩ	R1= 1MΩ R2= 10.1kΩ	R1= 1MΩ R2= 4.00kΩ	R1= 1MΩ R2= 2.00kΩ	R1= 1.53MΩ R2= 612.2Ω

more, because the DPM's own internal resistance is already so high (10MΩ). You'd need a whopping 40MΩ series resistor just to turn it into a 0-1V meter, for example. The circuit loading would be absolutely negligible, but finding or making up an accurate and stable 40MΩ multiplier wouldn't be easy.

So with a DPM, we need to use a **voltage divider** to adapt it for measuring higher voltages, rather than a simple multiplier. It's still fairly straightforward, though: see Fig.6.

If you like, you can think of resistor R1 as our original multiplier resistor, and R2 as an extra resistor added across the DPM to bring its 10MΩ internal resistance down to a suitably lower value. Or if you prefer, you can think of R1 and R2 simply as a voltage divider across the input voltage, and designed to 'break it down' to the 200mV needed by the DPM. To make the DPM read 0-20V you need a divider ratio of 20V/200mV or 100:1, for example.

The formula relating division ratio to R1 and R2 is:

$$V_R = (R1 + R2)/R2$$

where V_R is the divider ratio, expressed as a number, and R1 and R2 are the resistor values — although strictly speaking R2 here is really the parallel combination of resistor R2 and the DPM's own input resistance R_M.

The only thing you have to bear in mind is that to minimise the voltmeter's loading on the circuit you're measuring, the total resistance of the voltage divider should be as high as possible. But that may call for impractically high resistor values, especially for R1, and also mean that you *will* need to take the DPM's resistance R_M into account when you're working out the value for R2 (because of its shunting effect). So there's a bit more to it than simply picking a pair of resistors to give the right voltage division ratio.

Often, though, it *can* be that simple. If you want to make your DPM into a 50V meter, for example, it would probably be quite acceptable in terms of circuit loading to have the total divider resistance around 1MΩ. So since the divider ratio would be (50/0.2) or 250, you could pick a 1MΩ resistor for R1, and a value of 4.0kΩ for R2. Since R_M is 10MΩ, it'll have a very small shunting effect on R2 — about 0.04%. This can probably be ignored.

In fact with R1 fixed at 1MΩ, you can probably ignore the effect of R_M across R2 for any division ratio larger than 150 — corresponding to a 0-30V range or higher. You only need to take R_M into account for lower ratios; i.e., lower voltage ranges.

Finally, you again need to remember resistor voltage ratings when you're making a voltage divider to adapt a DPM into a voltmeter. So when you're building a divider for very high input voltages (say above 200V), it's again a good idea to split R1 into a number of series-connected resistors so they share the voltage drop.

Obviously most of the divider's voltage drop is across R1 (i.e., all but 200mV), and as we mentioned before metal film resistors generally aren't designed to withstand more than about 200V. So for a voltmeter designed to measure up to say 500V full scale, it would be a good idea to use say three 510kΩ resistors in series for R1, giving it a value of 1.53MΩ but with the three resistors able to share the load. Your corresponding value for R2 would be 612.2Ω, so the shunting effect of R_M would of course be negligible. Using a 620Ω resistor in parallel with one of 47kΩ would probably be close enough.

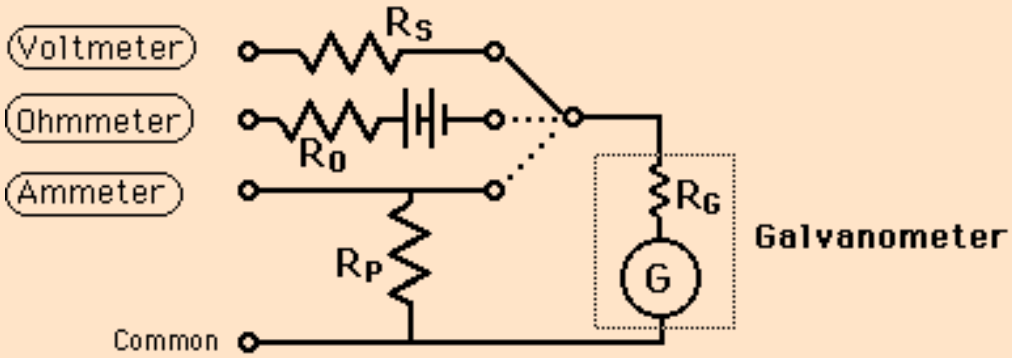
To save you time and effort, the table above shows the shunt and multiplier values for adapting Electus meters and DPMs for many common current and voltage ranges.

(Copyright © Electus Distribution, 2001)

ELECTUS DISTRIBUTION

Moving Coil Meters

The design of a voltmeter, ammeter or ohmmeter begins with a current-sensitive element. Though most modern meters have solid state digital readouts, the physics is more readily demonstrated with a moving coil current detector called a galvanometer. Since the modifications of the current sensor are compact, it is practical to have all three functions in a single instrument with multiple ranges of sensitivity. Schematically, a single range "multimeter" might be designed as illustrated.



Voltmeter design | Ammeter design | Ohmmeter design

Index

DC Circuits

HyperPhysics***** Electricity and Magnetism

R Nave

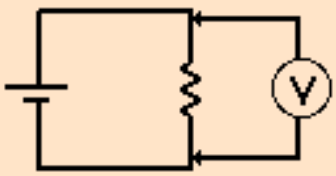
Go Back

Voltmeter

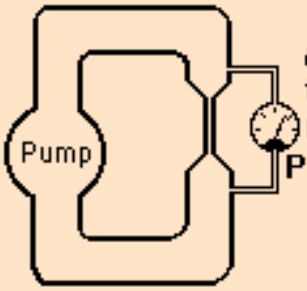
A voltmeter measures the change in voltage between two points in an electric circuit and therefore must be connected in parallel with the portion of the circuit on which the measurement is made. By contrast, an ammeter must be connected in series. In analogy with a water circuit, a voltmeter is like a meter designed to measure pressure difference. It is necessary for the voltmeter to have a very high resistance so that it does not have an appreciable affect on the current or voltage associated with the measured circuit. Modern solid-state meters have digital readouts, but the principles of operation can be better appreciated by examining the older moving coil meters based on galvanometer sensors.

Index

DC Circuits



A voltmeter is connected in parallel to measure the voltage change across a circuit element



A pressure gauge is connected in parallel to measure the pressure drop across the resistance.

Details of voltmeter design

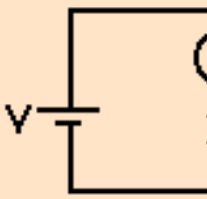
HyperPhysics***** Electricity and Magnetism *R Nave* Go Back

Ammeter

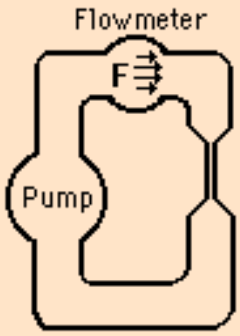
An ammeter is an instrument for measuring the electric current in amperes in a branch of an electric circuit. It must be placed in series with the measured branch, and must have very low resistance to avoid significant alteration of the current it is to measure. By contrast, an voltmeter must be connected in parallel. The analogy with an in-line flowmeter in a water circuit can help visualize why an ammeter must have a low resistance, and why connecting an ammeter in parallel can damage the meter. Modern solid-state meters have digital readouts, but the principles of operation can be better appreciated by examining the older moving coil meters based on galvanometer sensors.

Index

DC Circuits



Ammeter in series with resistor measures current through the resistor.



A meter for volume flowrate must be in series to measure the flow, but must not appreciably affect the flow.

Details about ammeter design

HyperPhysics***** Electricity and Magnetism *R Nave* Go Back

Ohmmeter

The standard way to measure resistance in ohms is to supply a constant voltage to the resistance and measure the current through it. That current is of course inversely proportional to the resistance according to Ohm's law, so that you have a non-linear scale. The current registered by the current sensing element is proportional to 1/R, so that a large current implies a small resistance. Modern solid-state meters have digital readouts, but the principles of operation can be better appreciated by examining the older moving coil meters based on galvanometer sensors.

Index

DC Circuits

HyperPhysics***** Electricity and Magnetism *R Nave*

Go Back

Voltmeter/Ammeter Measurements

The value of electrical resistance associated with a circuit element or appliance can be determined by measuring the voltage across it and the current through it and then dividing the measured voltage by the current. This method works even for non-ohmic resistances where the resistance might depend upon the current.

Index

DC Circuits

HyperPhysics***** Electricity and Magnetism	*R Nave*	Go Back

About me
Electronics
Freeware
Linux
SF & F
Info

Silvio Klaic's Home Page

This main page is loaded ... times since 10.Jan.1998.

To access one of my sections, click on left image of that section.

Short description of my sections:

- About me
  - In this section you can find some information about me.
- Electronics
  - Here you can find some programs and information about electronic devices which I made.
- Freeware
  - Here are some programs which I wrote, mostly for Linux and dos/windows.
- Linux
  - In this section are links to other Linux sites and programs.
- SF & F
  - Over here are links to some science fiction and fantasy pages which I found interesting.
- Info
  - Info section contains links and some information which are more or less related with my home page.

Home page is fully rebuild at 2. April 2001.
What? You expected dhtml, text animation, etc.?
No way, I try that and I didn't like it, it consumes too much resources and it's so large.
But this home page is nice, works fast, stable and you can view it with any browsers - textual or graphics, old or new.

Linux
Powered

How Many?

A Dictionary of Units of Measurement

For information on a specific unit, click on the first letter of its name:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Answers to the three most frequently asked questions:

How many micrograms (µg, ug or mcg) in a milligram (mg)?
1000 micrograms = 1 milligram, and 1000 milligrams = 1 gram.

How can I convert from international units (IU) to milligrams or micrograms?
Generally speaking, you can't. IU's measure the potency of a drug, not its mass or weight.

What countries besides the U.S. have not adopted the metric system?
Many U.S. teachers think the answer is "Liberia and Burma" (make that Myanmar). Let's give Liberia and Myanmar a break! *All* countries have adopted the metric system, including the U.S., and most countries (but not the U.S.) have taken steps to eliminate most uses of traditional measurements. However, in nearly all countries people still use traditional units sometimes, at least in colloquial expressions. Becoming metric is not a one-time event that has either happened or not. It is a process that happens over time. Every country is somewhere in this process of going metric, some much further along than others.

Selected by Science Educators SCILINKS from NSTA
Roadside Attractions
Seal of Approval Web Feet
Lightspan Study Web Academic Excellence Award
We're Featured In 300 Incredible Things to Learn on the Internet Available Now at Amazon.com
PSIgate

Commentary and Explanation

- Using the Dictionary
  - Using Numbers and Units
  - Roman and Arabic Numerals
  - Names of Large Numbers
  - Symbols and Abbreviations
- The International System (SI)
  - SI Base (Fundamental) Units
  - SI Derived Units
  - Metric Prefixes
  - SI Units for Clinical Data in Medicine
- The English Customary Systems
- The Metric System
  - CGS and MKS Units
  - Spelling of Metric Units
  - The Metric System in the United States
- ISO and IOC country codes
- **Links to Related Sites**
- Bibliography

What's New

- Nepalese land units: the anna and ropani
- a ground of land in south India
- the British unit of alcohol
- small angles measured in mgons
- measuring sand grains in phi units
- Strong-Cobb units for tablet hardness
- land area by the tarefa in Brazil
- large and small: *n*-plex and *n*-minex
- land area in Iraq: the olk and the dunum
- the use of mc- instead of μ- for micro-
- land by the soendre in Bhutan
- more on the origin of the ell
- revised Torino Impact Hazard Scale
- the talangwah, a Thai unit of area
- the ancient Hebrew hin
- the Dominican Republic's caballeria
- ancient Greek pous,

daktylos, plethron, and stadion

Index of Tables and Scales

- Apgar scoring (newborns)
- Beaufort scales (wind velocity)
- Bushel weights (U.S.)
- Cotton bale weights
- Danjon scale (lunar eclipse brightness)
- Drought severity (U.S.)
- Fujita scale (tornados, U.S.)
- Glasgow coma scale
- Grit sizes
- Hat sizes
- Mercalli scale (earthquakes)
- Nutritional daily values (U.S.)
- Paper sheet sizes (ISO)
- Paper sheet sizes (traditional)
- Paper sheet sizes (U.S. basic sizes)
- Radiocarbon year conversion
- Saffir-Simpson scale (hurricanes, U.S.)
- Sheet metal gauges
- Shot pellet sizes
- Shotgun gauges
- SI units for medical data
- Solar flare intensity
- Tennis racquet gauges
- Torino impact hazard scale
- Tropical cyclone categories (Australia)
- Viscosity grades (ISO)
- Volcanic Explosivity Index
- Wind chill chart (U.S.)
- Wire gauge table (U.S./U.K.)

Introduction

Americans probably use a greater variety of units of measurement than anyone else in the world. Caught in a slow-moving transition from customary to metric units, we employ a fascinating and sometimes frustrating mixture of units in talking about the same things. We measure the length of a race in meters, but the length of the long jump event in feet and inches. We speak of an engine's power in horsepower and its displacement in liters. In the same dispatch, we describe a hurricane's wind speed in knots and its central pressure in millibars.

Furthermore, our English customary units do not form a consistent system. Reflecting their diverse roots in Celtic, Roman, Saxon, and Norse cultures, they are often confusing and contradictory. There are two systems for land measurement (one based on the yard and the other on the rod) and a third system for distances at sea. There are two systems (avoirdupois and troy) for small weights and two more (based on the long and short tons) for large weights. Americans use two systems for volumes (one for dry

commodities and one for liquids) and the British use a third (British Imperial Measure).

Meanwhile, only a few Americans know that the legal definitions of the English customary units are actually based on metric units. The U. S. and British governments have agreed that a yard equals exactly 0.9144 meter and an avoirdupois pound equals exactly 0.453 592 37 kilograms. In this way, all the units of measurement Americans use every day are based on the standards of the metric system. Since 1875, in fact, the United States has subscribed to the International System of Weights and Measures, the official version of the metric system.

This dictionary began as a collection of notes describing the relationship between various English and metric units. It gradually grew until it finally became too large a word-processing document; I couldn't find my way around in it any more. So I turned it into a folder of html documents and added it to my Internet site. For many months, no one looked at the site except me and my students. Then, gradually, the dictionary began to attract users from around the world. Many users were kind enough to point out errors; others suggested additions and improvements. Questions about units began to appear in my email inbox. Sometimes I could answer the questions, sometimes not.

Today the dictionary has become a kind of interactive resource. It grows slowly and steadily, mostly through suggestions from readers and my efforts to answer questions posed by readers. You can participate in this process! Please let me know if you find any errors on the site, or if you can't find what you wanted to know, or if you know of units used in your field of study or in your part of the world that aren't included.

I hope you find the dictionary useful and informative.

Formalities

Written by:

Russ Rowlett, Director,
Center for Mathematics and Science Education
University of North Carolina at Chapel Hill.

You are welcome to email the author (rowlett@email.unc.edu) with comments and suggestions.

All material in this folder is copyright © 2005 by Russ Rowlett and the University of North Carolina at Chapel Hill. Permission is granted for personal use and for use by individual teachers in conducting their own classes. All other rights reserved. You are welcome to make links to this or any page of the dictionary. **Please do not copy the contents of any page of the dictionary to another site.** The material at this site is updated frequently as new information is added, so linking to the site rather than copying it is in the best interests of everyone.

The information contained in the dictionary is as accurate as I can make it; please notify me if you find any errors. **Neither the author nor the University of North Carolina assume any liability for uses made of the information presented by this web site.** The dictionary is not designed to promote any system of measurement. Its only purpose is to present information useful to anyone interested in the subject.

July 11, 2005

Field descriptions for TRAFO7 Excel spreadsheet ...

Field	Type	Description
Ud	Input	Si diode voltage at high current (change this voltage for using valve (tube) rectifiers, or special diodes)
Ubrss	Input	max. allowed typical peak-peak ripple voltage
Ia	Input	desired nom. DC output current per coil (ohmic load)
alpha	Input	current form factor (RMS > average!) 1.5 ... 2
Uaminn	Input	desired minimum nominal output voltage, to remember your goal
Uamin	Input	minimum nominal output voltage for computation
rho	constant	Copper thermal specific resistance
L	Input	line length from house mains entry to transformer,one way (metres)
A	Input	wire cross section area of this line with length *L*
Rk	computed	resulting line resistance for both ways
Unetz	Input	nominal effective mains voltage
fn	Input	mains frequency
tolup	Input	upper limit of mains voltage variation
tolow	input	lower limit of mains voltage variation
ü	computed	transformation ratio
ü2	computed	impedance transformation ratio
Rks	computed	line resistance transformed to secondary, usually very small
f	Input	actual used transformer loss factor
f	computed	heuristical determined loss factor, no large difference with actually used, unless reason for this is known!
U0eff	computed	nominal sine AC effective transformer voltage per coil with no load
Ineff	computed	nominal sine AC effective transformer current per coil
In	computed	nominal sine AC transformer peak current per coil, compare with *Imax* for a hint for *alpha* (value is doubled in case of Center circuit, since only one coil at a time is conducting in this rectifier circuit, while both are conducting with pure resistive load)
Imax	computed	secondary peak current per coil, also current through diodes, compare with *In* to avoid core saturation
Idm	computed	average diode current
Pa	computed	total power per winding delivered to load resistor
Pn	computed	total transformer power, including rectifier power, load power, all coils and factor alpha
Pns	computed	*Pn* divided by number of coils (1, 2)
Ua0+	computed	idle voltage on capacitor, including mains upper *Ua0+* tolerance

Ua0	computed	idle voltage on capacitor, nominal mains
Ua0-	computed	idle voltage on capacitor, including mains lower tolerance
Usper+	computed	idle voltage stress on diodes, including mains upper *usper+* tolerance
Ua8+	computed	infinite capacitor voltage with load and mains upper *Ua8+* tolerance, can be used to estimate e.g. amplifier losses
Ua-	computed	capacitor voltage (capacitance value *Cl*), including mains lower tolerance and line losses and ripple, lowest instantaneous value
Ua8	computed	infinite capacitor voltage with load, nominal
Ua8ber	computed	tryed infinite capacitor voltage with load
r	computed	ratio between *Ua8* and *Ua8ber* in %, error of iteration, minimise to 0.0!
Uneff	iteration input	nominal effective transformer secondary voltage, this value must be iterated to minimise *r*
Rv	computed	load resistor that would draw *Ia* with infinite capacitor
Ri	computed	transformer resistance for each coil
Ri +main	computed	effective transformer resistance for each coil, including line losses
Cl	computed	capacitor size for give ripple, frequency, load
Pd	computed	power loss per diode
Pg	computed	total power loss, sum for all diodes

Close

Microsoft Visual Studio

Visual Basic 6.0 SP5: Run-Time Redistribution Pack (vbrun60sp5.exe)

Brief Description

vbrun60sp5.exe is a self-extracting executable file that installs versions of the Microsoft Visual Basic run-time files required by all applications created with Visual Basic 6.0. The files include the fixes shipped with Visual Studio 6.0 Service Pack 5.

On This Page

Quick Details
System Requirements
Related Resources
Overview
Instructions
What Others Are Downloading

Download

Quick Details

File Name:	vbrun60sp5.exe
Version:	sp5
Date Published:	2/28/2001
Language:	English
Download Size:	1020 KB
Estimated Download Time:	3 min 56K

Overview

vbrun60sp5.exe is a self-extracting executable file that installs versions of the Microsoft Visual Basic run-time files required by all applications created with Visual Basic 6.0. The files include the fixes shipped with Visual Studio 6.0 Service Pack 5.

Top of page

System Requirements

- **Supported Operating Systems:** Windows 2000; Windows 95; Windows 98; Windows ME; Windows NT

Top of page

Instructions

1. Click the **Download** button on this page to start the download.
2. Do one of the following:
   - To start the installation immediately, click **Open** or **Run this program from its current location**.
   - To copy the download to your computer for installation at a later time, click **Save** or **Save this program to disk**.

Top of page

Related Resources

1. Product Documentation for Visual Basic 6.0
2. Code Samples for Visual Basic 6.0
3. Order the Visual Basic 6.0 Reference Library
4. Try Visual Studio 2005

Top of page

What Others Are Downloading

Others who downloaded **Visual Basic 6.0 SP5: Run-Time Redistribution Pack (vbrun60sp5.exe)** also downloaded:

1. Service Pack 6 for Visual Basic 6.0
2. Visual Basic 6.0 Run-time Files
3. Microsoft Visual Basic 6.0 Common Controls
4. Service Pack 6 for Visual Basic 6.0, Visual C++ 6.0 with Visual Source Safe 6.0d
5. Microsoft .NET Framework Version 2.0 Redistributable Package (x86)

Top of page

Manage Your Profile | Contact Us

© 2006 Microsoft Corporation. All rights reserved. Terms of Use | Trademarks | Privacy Statement

Skip to content

Search Go

Advanced search

Google search

- Home
- Download
- Support
- Contributing
- Projects
- My pages
- About

start:

- Native language
- New user & general Info

OpenOffice.org 2 Components:

- Base
- Calc
- Draw
- Math
- Impress
- Writer

General links
Donating Funds
Mailing Lists
Report a Bug
What's New
Project Guidelines
Licenses

Search

Go
Advanced search
Google search
How do I...
Sitemap
FAQ

free office suite

OpenOffice.org is a multiplatform and multilingual office suite and an open-source project. Compatible with all other major office suites, the product is free to download, use, and distribute. To help build the community, join us.

Get OpenOffice.org
Version: 2.0.2

In the media

25 May	Easing Migration to OOo
18 May	Interview: Carlos Dantas de Menezes
17 May	Goodbye Office, Hello OpenOffice.org

Get help?

16 May	Get Legal - Get OOo!
16 May	ISO Approval Unlikely for MS' OpenXML
16 May	IBM: ODF for Lotus
15 May	WSJ: Out the Window
15 May	Bridging Office/OpenOffice.org
14 May	Spotlight on OOo
09 May	Politics of Plug-Ins
08 May	Rotten Effort
Always	OOo Major Deployments
Always	The OpenOffice.org Newsletter

OOoCon 2006 Call for Papers

Almost Final Call : Just Days Left! Send in your abstract now.

We have issued our Call for Papers, and are inviting you, the community, to submit abstracts on a range of topics. The deadline is **1 June**.

OOoCon will be held in Lyon, France, from 11-13 September. All are welcome. Past conferences have been fun, important, memorable, and have dealt with technical as well as business and social concerns.

- OOoCon 2006 Call for Papers

New Winner, Developer Article Contest

S. Sevki Dincer has won April's Developer Article Contest; congratulations! Sevki's article "is a step by step guide on how to write a simple component (a Calc add-in)

for OpenOffice.org (OOo) in C++."

The Developer Article Contest continues each month. Winners receive USD $750 plus invaluable attention and the gratification of helping others. Read our wiki on the matter for more information and start writing! The next deadline is the end of this month.

- S. Dincer's Simple Calc Add-In
- Developer Article Contest

By any use of this Website, you agree to be bound by these Policies and Terms of Use

CollabNet is a trademark of CollabNet, Inc., Sun, Sun Microsystems, the Sun logo, Java, Solaris, StarOffice are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries.

.NET Home

Solutions for

Business Leaders

Industries

Developers

Resources

What Is .NET?

Business Articles

Momentum

.NET Home

What Is .NET?

Updated: February 1, 2005

.NET is the Microsoft Web services strategy to connect information, people, systems, and devices through software. Integrated across the Microsoft platform, .NET technology provides the ability to quickly build, deploy, manage, and use connected, security-enhanced solutions with Web services. . NET-connected solutions enable businesses to integrate their systems more rapidly and in a more agile manner and help them realize the promise of information anytime, anywhere, on any device.

The Microsoft platform includes everything a business needs to develop and deploy a Web service-connected IT architecture: servers to host Web services, development tools to create them, applications to use them, and a worldwide network of more than 35,000 Microsoft Certified Partner organizations to provide any help you need.

What Are Web Services?

If you ask a developer what Web services are, you'll hear something like, "self-describing software modules, semantically encapsulating discrete functionality, wrapped in and accessible via standard Internet communication protocols like XML and SOAP."

But if you ask a business leader who has implemented Web service-based solutions, you'll get a different kind of answer. You'll hear that Web services are an approach that helps the business connect with its customers, partners, and employees. They enable the business to extend existing services to new customers. They help the business work more efficiently with its partners and suppliers. They unlock information so it can flow to every employee who needs it. They reduce development time and expense for new projects. You'll hear less about what

Related Links

- Case Studies: .NET Helps Businesses Solve Tough Problems
- .NET Connected Directory Helps Businesses Find Web Services Solutions and Products
- See Why .NET is Gaining Momentum Around the World
- Maximize Your Applications with Smart Client Technology

More Resources

- .NET Glossary
- Frequently Asked Questions about .NET
- The ABC's of Web Services
- See What Web Services Can Do for You
- What's Next—the Future of Web Services

Web services are and more about what they enable the business to do.

Top of page

Benefits of Web Services

By enabling applications to share data across different hardware platforms and operating systems, Web services provide many benefits, including:

- Opening the door to new business opportunities by making it easy to connect with partners.
- Delivering dramatically more personal, integrated experiences to users through the new breed of smart devices—including PCs.
- Saving time and money by cutting development time.
- Increasing revenue streams by enabling businesses to easily make their own Web services available to others.

Top of page

Connecting Applications Through Web Services

Web services are revolutionizing how applications talk to other applications—or, more broadly, how computers talk to other computers—by providing a universal data format that lets data be easily adapted or transformed. Based on XML, the universal language of Internet data exchange, Web services can communicate across platforms and operating systems, regardless of the programming language in which the applications are written.

Each Web service is a discrete unit of code that handles a limited set of tasks. However, although Web services remain independent of each other, they can loosely link themselves into a collaborating group that performs a particular task.

Top of page

Example: Your Inventory System

Say you have a stand-alone inventory system. If you don't connect it to anything else, it's not as valuable as it could be. The system can track inventory, but not much more. You may have to enter inventory information twice—once in your accounting system and once in your customer relationship management system. The inventory system may be unable to automatically place orders to suppliers. The benefits of such an inventory system are diminished by high overhead costs.

However, if you connect your inventory system to your accounting system, it gets more interesting. Now, whenever you buy or sell something, the implications for your inventory and your cash flow can be tracked in one step. If you go further, and connect your warehouse management system, customer ordering system, supplier ordering systems, and your shipping company, suddenly that inventory management system is worth a lot. You can do end-to-end management of your business while dealing with each transaction only once, instead of once for every system it affects. That's a lot less work—and a lot less opportunity for errors.

These connections can be made easily using Web services. Web services allow the applications to share information through the Internet, regardless of the operating system or back-end software that the application is using.

Top of page

Web Services Use Industry-Standard Protocols

Web services also make it possible for developers to choose between building all pieces of their applications, or consuming (using) Web services created by others. This means that an individual company doesn't have to supply every piece for a complete solution. The ability to expose (announce and offer) your own Web services creates new revenue streams for your company.

Web services are invoked over the Internet by means of industry-standard protocols including SOAP; XML; and Universal Description, Discovery, and Integration (UDDI). They are defined through public standards organizations such as the World Wide Web Consortium (W3C).

SOAP is an XML-based messaging technology standardized by the W3C, which specifies all the necessary rules for locating Web services, integrating them into applications, and communicating between them. UDDI is a public registry, offered at no cost, where one can publish and inquire about Web services.

Top of page

Manage Your Profile | Contact us

© 2006 Microsoft Corporation. All rights reserved. Terms of Use | Trademarks | Privacy Statement

Microsoft

SDKs, Redistributables & Service Packs

Stay up-to-date with these downloadable service packs, updates, samples, and related components for the .NET Framework.

Order the .NET Framework on DVD

Need the .NET Framework on DVD? We'll be shipping the Framework to customers in the US on DVD (International customers must contact their local Microsoft subsidiary.) You can order the disks today.

.NET Framework Version 2.0 Redistributable Package

Download the most recent update to the Microsoft .NET Framework and get everything you need to run applications developed using the .NET Framework.

- Download x86 version
- Download x64 version
- Download IA64 version

.NET Framework Version 2.0 Software Development Kit

This SDK includes everything you need to write, build, test, and deploy .NET Framework applications—documentation, samples, and command-line tools and compilers. You must install the .NET Framework version 2.0 Redistributable Package prior to installing the SDK.

- Download x86 version
- Download x64 version
- Download IA64 version

Related Redistributables

- .NET Compact Framework Version 2.0 End-User Redistributable
- Microsoft Visual J# Version 2.0 Redistributable Package
- Microsoft Visual Studio 2005 Tools for the

- Microsoft Office System (VSTO 2005) Redistributable Package (x86)
- Microsoft Report Viewer Redistributable 2005
- Microsoft Data Access Components (MDAC)
- SQL Server 2005 Express Edition
- SQL Native Client
- Windows Installer 3.1
- Windows Installer 2.0 for Windows 2000 and Windows NT 4.0
- Windows Installer 2.0 for Windows 95, 98, and Me

Related SDKs

- DirectX
- Microsoft Component Installer 2.0
- Windows Installer XML (WiX)

.NET Framework 2.0 Language Packs

.NET Framework language packs contain translated text, such as error messages, for languages other than English. Without the installation of a Language Pack, these messages are displayed in English by default. Multiple language packs, each for a different language, can be installed on a single computer.

- Download x86 version
- Download x64 version
- Download IA64 version

Top of Page

Manage Your Profile | Legal | Contact Us | MSDN Flash Newsletter

© 2006 Microsoft Corporation. All rights reserved. Terms of Use | Trademarks | Privacy Statement

Microsoft

Audio Millivoltmeter

Rod Elliott - ESP (Original Design)

Introduction

When performing any tests on an audio system, some form of measuring device is essential. Digital multimeters are not useful, since they will not give the true picture of what is happening, and most have a fairly limited frequency range. An oscilloscope is the ideal tool, but not all hobbyists can afford the outlay for a scope, and would find justifying the not inconsiderable cost a tad difficult.

An AC millivoltmeter - calibrated in dB - with a range of 30V down to 3mV full scale (80dB range) would be extremely useful. Attach a microphone (electret mic capsules are quite good), and you have a relative sound level meter, even better if you have some way of calibration.

The meter presented here has a very wide frequency range, and uses a switched attenuator for range adjustment. The attenuator uses the 30-10-3 sequence, which provides 10dB steps between ranges. The standard attenuator provides an input impedance of over 2M Ohms, but is a nuisance because with such high impedances stray capacitance causes havoc with the calibration, so a parallel capacitive attenuator is also needed. If you expect to work with valve amplifiers, you will want the high impedance, but otherwise the low impedance attenuator should do nicely.

The Attenuators

The two attenuator networks are shown in Figures 1 and 2, and as you can see the Hi-Z version requires all those capacitors. They must be accurate, too. Otherwise high frequency performance will be all over the place, so you need a capacitance meter or a

source of close tolerance caps. The resistors are standard E24 series 1% metal film types, and the caps (if used) should ideally be polystyrene or polyester, but if ceramic is all you can get, then ceramics are what you use. If you do have to use ceramic caps, make sure that they have low thermal drift - NPO or similar.

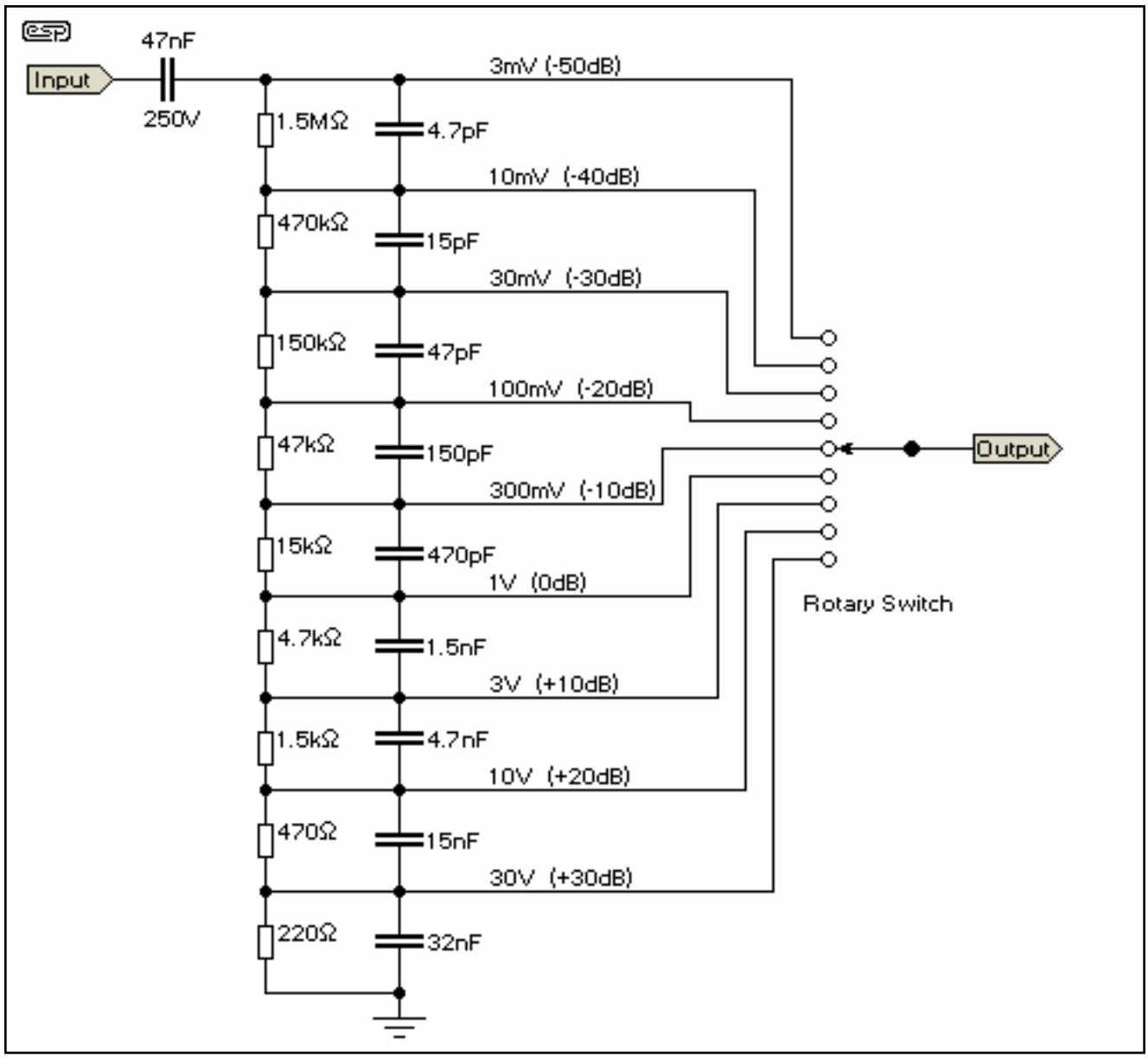


Figure 1 - High Impedance Attenuator (> 2M Ohm)

Note that the 32nF capacitor really is 32nF, and will have to be built up from smaller caps or selected from a batch. The capacitor tolerance ideally should be the same as that for the resistors - 1% is suggested to get acceptable accuracy.

With the Lo-Z attenuator, performance can be expected to be quite linear up to at least 80kHz before stray capacitance starts to influence the measurement. Without the paralleled capacitive attenuator, the Hi-Z version will start to show incorrect readings above 10kHz, which is unacceptable. The stray capacitance comes from the switch

contacts and the proximity of the resistors to each other, and only a few pF will cause havoc at high frequencies.

To minimise capacitance, mount all resistors (and capacitors) directly off the rotary switch, and keep them as separated from each other and the remainder of the circuitry as possible. Do not be tempted to try to make the arrangement nice and neat (with all the components nicely aligned with each other), for this will increase the capacitance of the circuit and ruin the high frequency performance. All component leads must be as short as possible.

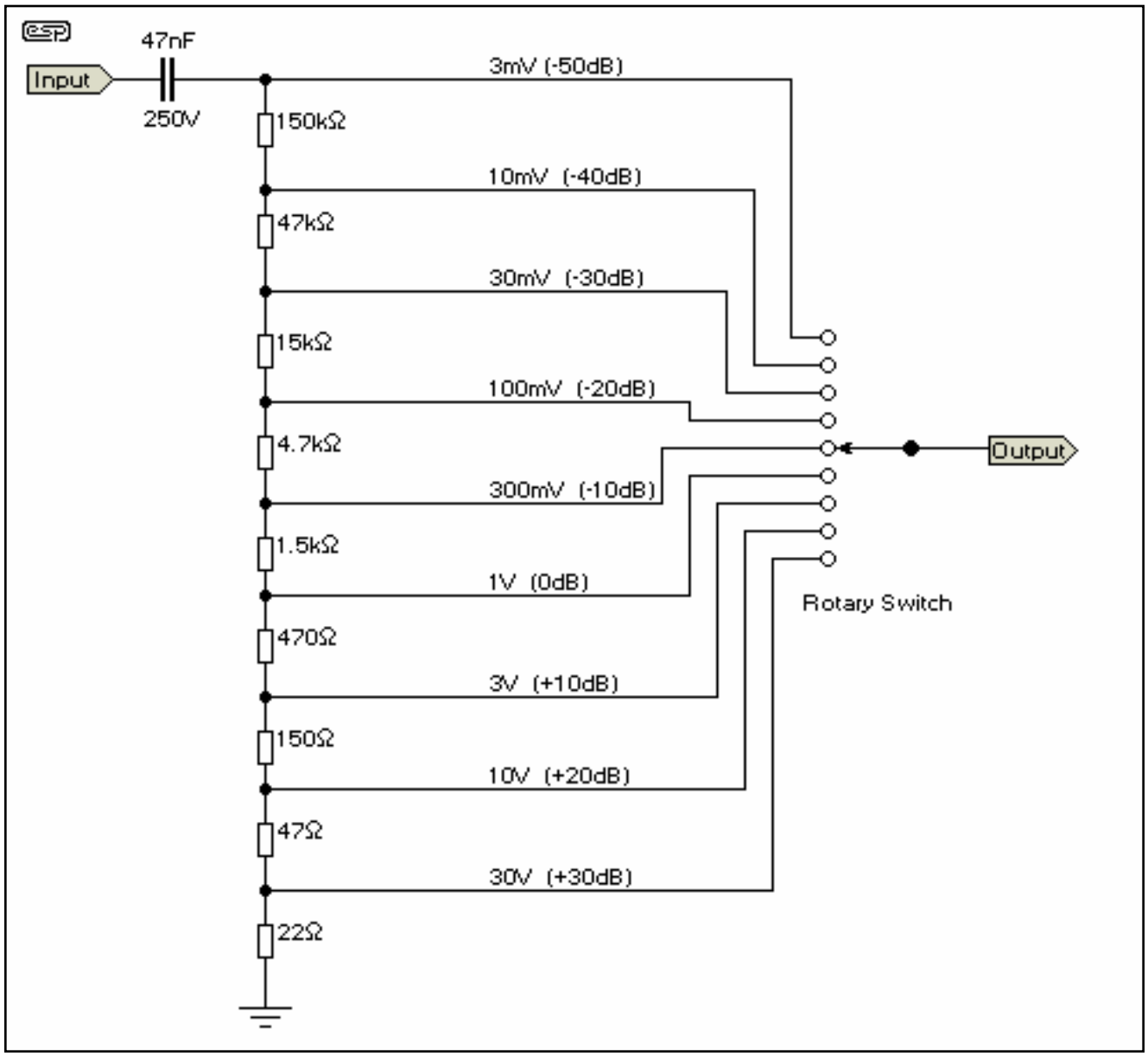


Figure 2 - Low Impedance Attenuator (> 200k Ohm)

The low-Z version is less susceptible to stray capacitance, but even at 150k (the highest value resistor in the circuit), only a few pF is needed to start to have an adverse influence at 100kHz or so. Again, do not strive for neatness, as this will only degrade

performance.

As can be seen, both attenuators have the same ranges - from 3mV to 30V in 10dB steps. Because 10dB is a ratio of 3.16, the scale must be calibrated with two voltage scales, 0-1 and 0-3. If you are very fortunate, it may be possible to find a meter with these ranges already (I have used about 3 of them in various projects), but they are now rather scarce. We can blame digital technology for this, but some analogue multimeters might have the scales you need - such a meter will not be cheap, however.

Amplifier

The amplifier(s) used in such a project are critical - we need wide bandwidth and low noise, coupled with low current drain, since we want to be able to run the meter on a 9V battery. The meter amplifier also requires high input impedance - especially for the high impedance attenuator version.

Consequently, a discrete opamp is the device of choice, since it can be built to satisfy all the desirable features we need. The input sensitivity of the meter amp is 3mV for full scale deflection on the meter, so it requires a fair bit of gain. In the final circuit, a JFET is used to provide the necessary high impedance input, and has the added benefit of supplying extra gain - this helps to reduce the demands on the opamp, since a typical 2N5459 JFET in the circuit shown (Figure 4) provides a gain of about 15, raising the input voltage to the opamp circuit to about 45mV for 3mV input.

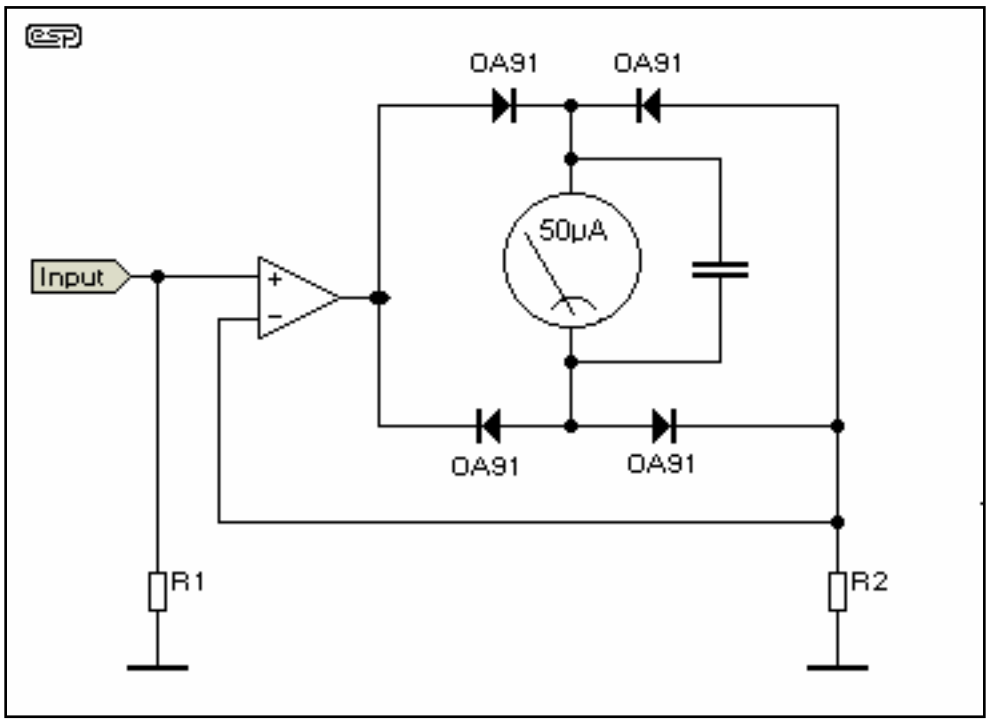


Figure 3 - Basic Meter Amplifier Scheme

Another requirement is simplicity and good linearity. The basic meter amplifier shown in Figure 3 satisfies all our requirements, but as you can see uses germanium diodes. Although these are harder to get and more expensive than silicon (and have higher leakage current), they also have very wide bandwidth and significantly less voltage drop than silicon, which reduces the requirement of the opamp to have an infinite slew rate.

This basic design has been around for many years, and is still one of the easiest to make, having the minimum of parts. The voltage across R2 must be the same as the input voltage (basic law of opamps) for the amp to be stable, so all losses in the meter and diodes are 'restored' by the opamp. The capacitor will need to be selected for the meter movement you use, since different meters have different damping. Initially this can be left out, but if excessive meter swings cause problems (or oscillation at low frequencies), then the capacitor will be needed. A value of 10uF is always a good starting point.

The input sensitivity is simply set by changing the value of R2 (In Figure 3), with lower values providing higher sensitivity and vice versa. Typically, R2 will be a multi-turn trimpot to allow for calibration.

Complete Meter Amplifier

The entire circuit can be built easily on a piece of perforated board (Veroboard or similar is good for this type of circuit), and a printed circuit board is quite unnecessary. Lay the physical circuit out following the schematic layout as closely as possible. This nearly always works well with discrete circuits, and makes it easy to follow 10 years later when you need to fix it. (I have had mine for nearly 20 years, and have not had to fix it yet.)

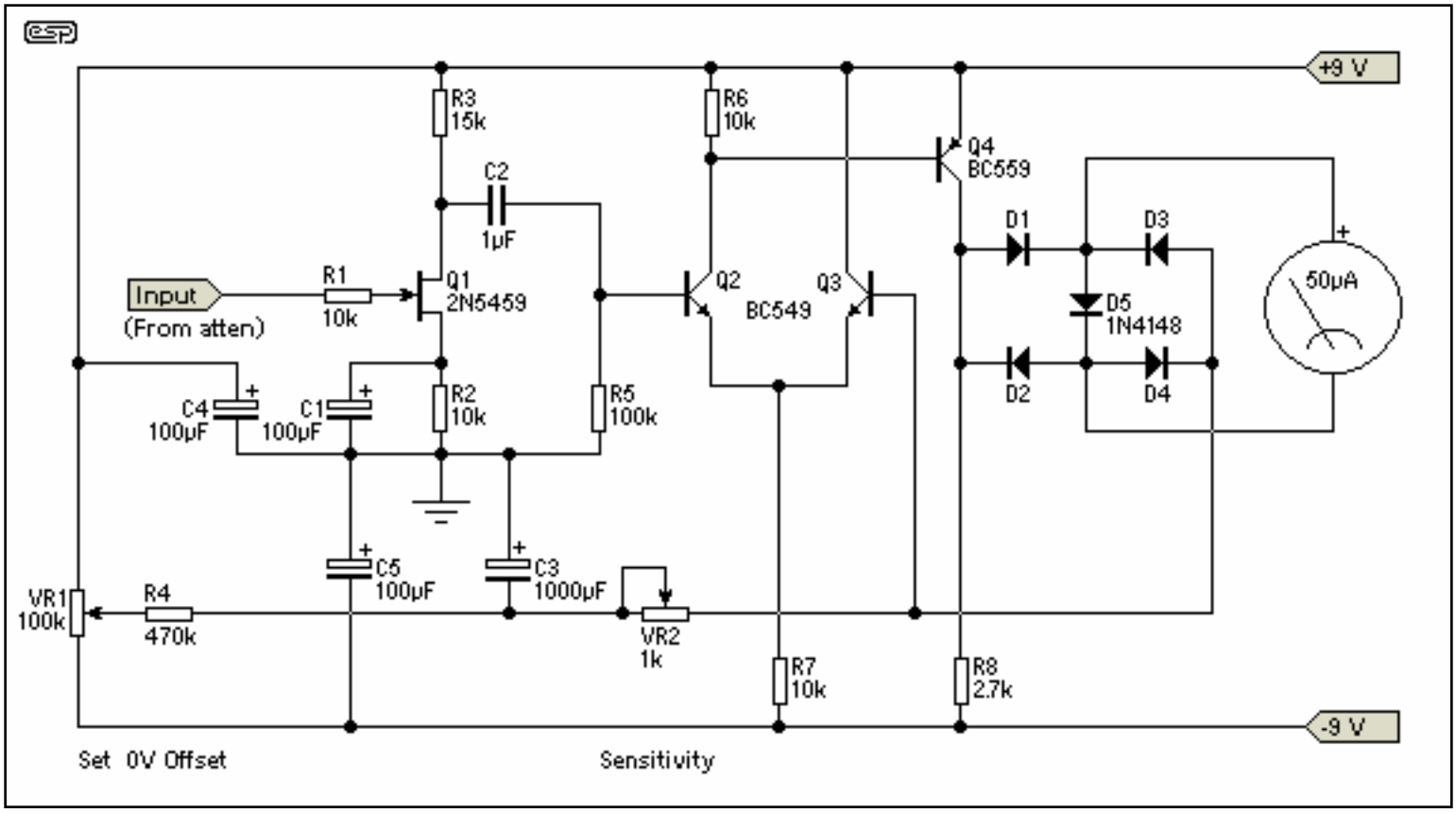


Figure 4 - The Complete Meter Amplifier

Figure 4 shows the complete meter amplifier, and uses a 2N5459 J-FET for the input. This provides a very high input impedance as well as some useful gain (about 15), allowing the entire unit to use a single discrete opamp - a discrete design is used to obtain optimum performance (frequency response) not readily obtainable with standard opamps for this application. The distortion characteristics are relatively unimportant, but the requirement for wide bandwidth and high gain over the entire frequency range excludes most conventional opamps. Use OA91 or similar germanium diodes for D1-D4 for best results. D5 *must* be silicon.

Note that the FET has no gate resistor, but relies on the voltage divider (attenuator) for its earth reference. This is not generally considered good practice, but will cause only minor 'flicks' of the needle when changing ranges. Also be aware that there is minimal input protection, so if you have the meter set to the 3mV range and connect it to the speaker output of an amplifier, you will probably cause damage. A resistor (10k) in series with the gate is be used, but will not offer full protection. A resistor which would offer full protection at any voltage will also cause problems at high frequencies because of its high value. Likewise, the capacitance of protection diodes will also adversely affect the high frequency performance (which is why there aren't any).

The discrete opamp is only a simple design, but manages a frequency response to well

over 100kHz (-1dB) with the 50uA meter load, and will operate satisfactorily with battery voltages down to 8V (the lower limit for a 9V battery before its internal impedance rises significantly). 100uF capacitors are used to ensure that the battery supply is bypassed, to help counteract the impedance rise as the battery ages.

I used BC549 (NPN) and BC559 (PNP) transistors, but any high gain, reasonably wide bandwidth, low current device will (should) work fine. As always, all resistors should be metal film, and the two pots should be multi-turn to allow accurate setting.

For those who find the 2N5459 JFET difficult to obtain, you can use a BF244, which a reader tells me works very well. During assembly, it is extremely important to keep stray capacitance to the minimum. The amplifier has very high gain and wide bandwidth, and oscillation will (not might) occur if you are not careful. In particular, keep the leads of C2 short, and make certain that the output (meter) leads are kept well separated from inputs and the attenuator.

If you use Veroboard, make sure that the track strips are cut at each end, so that they cannot act as antennas at high frequencies - this will cause oscillation or poor high frequency response, and neither will add to the instrument's usefulness.

Test and Calibration

The initial test involves connecting the meter amp to the attenuator, and applying power. All wiring must be carefully checked before you do this - the 9V batteries can supply enough current to damage the transistors, but batteries are more expensive than the transistors, and a wiring mistake may place a heavy discharge on the batteries rendering them dead before their time. Normally, batteries should last for quite a while, since the current drain is only about 4.5mA.

The meter is provided with protection by D5, a 1N914 (or 1N4148) which will conduct if too much voltage (or current) is applied - the meter will be hard against the stops though, and may still be damaged if the condition is allowed to persist. When power is applied, the meter should flick to full scale, then quickly settle back to near 0 volts. If it remains at full scale, you have made a mistake, so remove power immediately and locate (and fix) the error.

Calibration involves first setting the Set 0V Offset pot to its midway position, then carefully adjusting until the meter reading is as close as possible to the zero voltage mark. Any remaining offset must be removed using the meter's mechanical zero

adjustment - this is a little crude, but there is not much choice with this type of circuit. You will find that the meter reading will drop to some minimum value then start to climb again - this is because of the full-wave rectifier in the meter and feedback circuit.

Next, an accurate voltage at somewhere between 100 to 2,000 Hz is used to calibrate the meter. Select a suitable range on the attenuator, then adjust the Sensitivity control until the meter shows a reading that is identical to the applied voltage. Where possible, this should be done with the millivoltmeter at full scale on the 1V range. Remember that the scales are different for 1V and 3V ranges. The sensitivity pot will have more than enough range to allow the unit to be calibrated, provided no wiring mistakes have been made.

If by some chance your version decides it wants to oscillate at a MHz or three, you will need to add a small capacitance between collector and base of Q4 - I would not expect that more than 10pF should be needed, and even this will reduce the high frequency response slightly. Mine has a -1dB upper frequency limit of about 250kHz, a frequency which is more than adequate for audio use (by nearly an order of magnitude), but this is without a frequency limiting (miller or dominant pole) capacitor.

Meter Movement

It is important to obtain the best meter movement you can find, or the unit will be hard to read and possibly inaccurate. You will need to make a new scale for the meter, showing the two ranges and a dB scale. One possible reproduction is shown below, and there are links to a couple of others, one of which should suit whatever passes for a fairly standard 50uA movement - these should be available wherever you are, but you might need to look around a bit.

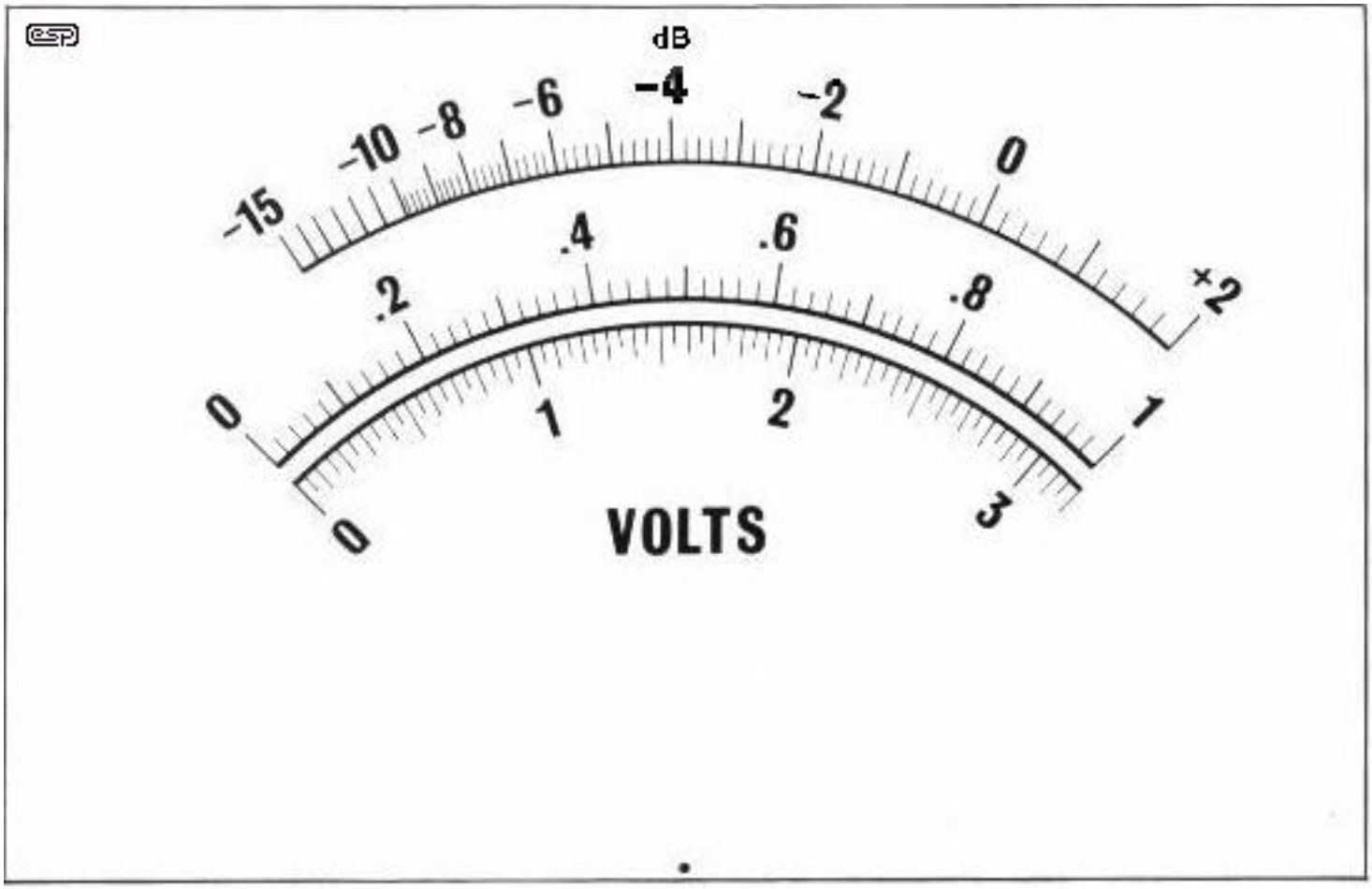


Figure 5 - Meter Face

The alternate meter faces are Meter Face 1 and Meter Face 3 - One of these should be able to be resized to suit the movement you use, but some experimentation is needed. You will notice that #3 appears to be hand-drawn in some areas - that's because it was. This is a scan from my own millivoltmeter, and when it was built, all these new-fangled scanners and computer thingies were a bit less common than today (you can gather from this that I have had the meter for a while).

To see what you are trying to find, check out this link to an Australian company called Jaycar Electronics. This is a link to the company website, and you will have to search for the movement (I used to have a link to the picture of the meter itself, but Jaycar has changed their website and the link will not work). Those available are not great meter movements, but are similar to the one I am using, and work quite well.

A typical 50uA movement will have a resistance of around 2k to 3k Ohms, and on average, expect to pay about AU$20 (or about US$15 or so) for a passably good movement.

When completed, the meter can be calibrated against a known accurate digital

multimeter, using a frequency of about 100Hz (most digitals will give an accurate reading at this frequency).

Optional Battery Check

If you would like to be able to measure the battery voltage without dismantling the instrument (I feel that this is a worthwhile addition), the switching shown in Figure 6 can be added. Note that both terminals of the meter must be switched, and the average of the two 9V batteries can be read on the 1V scale (so 0.9V would indicate 9V for each battery.

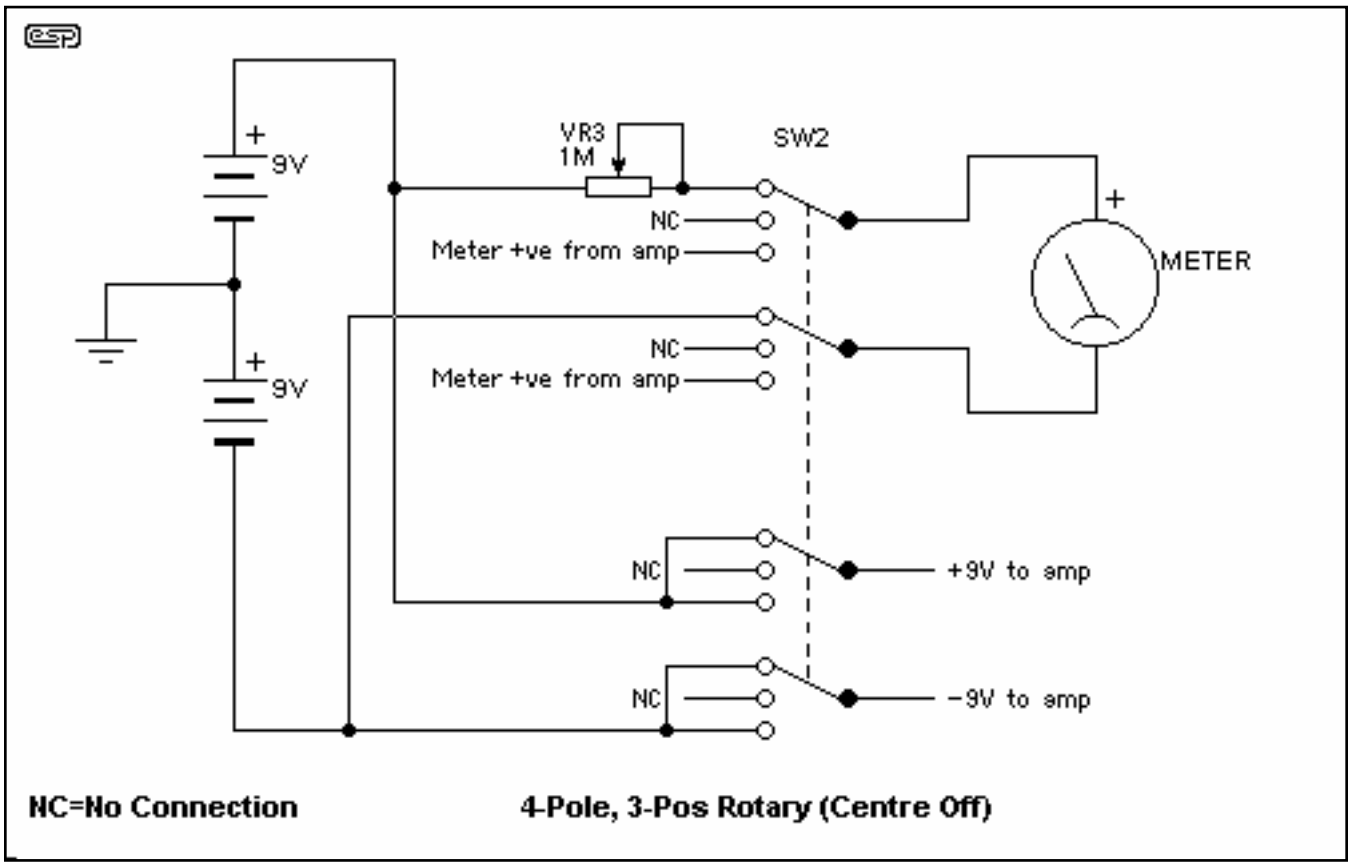


Figure 6 - Optional Battery Check Circuit

Use of a trimpot - preferably multi-turn - allows the voltmeter to be calibrated against an accurate multimeter, and the voltage shown is with the meter electronics switched on, so it will read the real loaded voltage. Both batteries are measured in series, so the nominal voltage read by the meter is 18V (R=18/50uA=360k, so a 1M trimpot should be near the middle of its travel). This option requires that the power switch be a 3-position 4-pole rotary, so it will cost a little more than a simple DPST mini-toggle.

WARNING: Make sure that the trimpot is set to approximately 1/2 way before attempting

calibration of the voltmeter. 2 x 9V batteries will destroy the coil of the meter *instantly* if the pot is set at minimum!

Enclosure

The case used must be all metal, since the attenuator and meter amplifier needs good shielding against noise pickup. This can be made from sheet aluminium or other metal (steel, brass, etc) if you have the tools to work with it, otherwise a suitable case may be obtainable from your normal parts supplier.

Another alternative is to use un-etched copper clad printed circuit board. Cut the panels to size, and solder together from the inside, filing off the outsides so the panels are all flush, and finally finishing the unit with a suitable coat or two of paint. There are many different finishes available in spray cans, so take your pick.

Cases built in this way can look surprisingly good if you take the time to finish them off well.

Construction

Make sure that the 0V line (the junction of the batteries, bottom of the attenuator string and earth input terminal are all tied to a common point on the front panel, and that the remainder of the case is in good electrical contact. If the case is not earthed properly, this is worse than using a non-shielded case!

It may also be necessary to add shielding between the FET stage and the main meter, and a small cap (10nF should be connected across the meter output, as close as possible to the diodes. Keep all leads short, and ensure that the output leads are kept well away from the input.

The meter amp is wide band, and has a full scale sensitivity of 3mV. It *will* oscillate if there is *any* feedback from O/P to I/P or between stages.

Front Panel

Drill all holes first for the two rotary switches, the meter and its mounting bolts and the input connectors. I always use a BNC connector, but for audio an RCA connector might

be better. One can also use "banana" sockets, so you can use ordinary multimeter leads, but being unshielded they will pick up noise - especially on the lower voltage ranges.

Make sure that all panel components fit properly, and de-burr the panel on both sides.

Mark the switch positions for each setting very carefully, since markings that do not line up with the pointer on the switch knobs look tacky, and can be confusing when you use the instrument.

Once you have the exact switch positions marked out, the front panel can be labelled any way you see fit. One method that works well is to use a graphics drawing package to create the label, and print it out onto ordinary paper. Carefully stick clear 'contact' film (as used for covering school books, etc) on both sides, ensuring that there are no air bubbles trapped under the film. Trim to the exact size of the front panel.

Use spray adhesive on both the panel and the label, leave for a few minutes then very carefully apply the label. You have to get this right first time - once stuck you will damage the label trying to move it ! The hole cutouts should then be very carefully removed using a hobby knife, scalpel or other suitable (sharp) instrument.

That's it - you are now the proud owner of a very useful piece of test equipment.

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

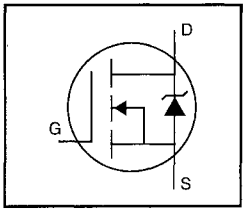
Updated 03 Oct 01 - added some additional construction hints and alternative FET details./ 15 Jul 01 - corrected mistake in attenuators./ 21 Feb 2006 - reformatted drawings, clarified text.

International IOR Rectifier

IRFP240

HEXFET® Power MOSFET

- Dynamic dv/dt Rating
- Repetitive Avalanche Rated
- Isolated Central Mounting Hole
- Fast Switching
- Ease of Paralleling
- Simple Drive Requirements



V_{DSS} = 200V

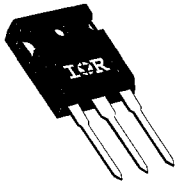
$R_{DS(on)}$ = 0.18Ω

I_D = 20A

Description

Third Generation HEXFETs from International Rectifier provide the designer with the best combination of fast switching, ruggedized device design, low on-resistance and cost-effectiveness.

The TO-247 package is preferred for commercial–industrial applications where higher power levels preclude the use of TO-220 devices. The TO-247 is similar but superior to the earlier TO-218 package because of its isolated mounting hole. It also provides greater creepage distance between pins to meet the requirements of most safety specifications.



TO-247AC

DATA SHEETS

Absolute Maximum Ratings

	Parameter	Max.	Units
I_D @ T_C = 25°C	Continuous Drain Current, V_{GS} @ 10 V	20	A
I_D @ T_C = 100°C	Continuous Drain Current, V_{GS} @ 10 V	12	
I_{DM}	Pulsed Drain Current ①	80	
P_D @ T_C = 25°C	Power Dissipation	150	W
	Linear Derating Factor	1.2	W/°C
V_{GS}	Gate-to-Source Voltage	±20	V
E_{AS}	Single Pulse Avalanche Energy ②	510	mJ
I_{AR}	Avalanche Current ①	20	A
E_{AR}	Repetitive Avalanche Energy ①	15	mJ
dv/dt	Peak Diode Recovery dv/dt ③	5.0	V/ns
T_J T_{STG}	Operating Junction and Storage Temperature Range	-55 to +150	°C
	Soldering Temperature, for 10 seconds	300 (1.6mm from case)	
	Mounting Torque, 6-32 or M3 screw	10 lbf•in (1.1 N•m)	

Thermal Resistance

	Parameter	Min.	Typ.	Max.	Units
$R_{\theta JC}$	Junction-to-Case	—	—	0.83	°C/W
$R_{\theta CS}$	Case-to-Sink, Flat, Greased Surface	—	0.24	—	
$R_{\theta JA}$	Junction-to-Ambient	—	—	40	

Electrical Characteristics @ T_J = 25°C (unless otherwise specified)

	Parameter	Min.	Typ.	Max.	Units	Test Conditions
$V_{(BR)DSS}$	Drain-to-Source Breakdown Voltage	200	—	—	V	V_{GS}=0V, I_D= 250μA
$\Delta V_{(BR)DSS}/\Delta T_J$	Breakdown Voltage Temp. Coefficient	—	0.29	—	V/°C	Reference to 25°C, I_D= 1mA
$R_{DS(on)}$	Static Drain-to-Source On-Resistance	—	—	0.18	Ω	V_{GS}=10V, I_D=12A ④
$V_{GS(th)}$	Gate Threshold Voltage	2.0	—	4.0	V	V_{DS}=V_{GS}, I_D= 250μA
g_{fs}	Forward Transconductance	6.9	—	—	S	V_{DS}=50V, I_D=12A ④
I_{DSS}	Drain-to-Source Leakage Current	—	—	25	μA	V_{DS}=200V, V_{GS}=0V
		—	—	250		V_{DS}=160V, V_{GS}=0V, T_J=125°C
I_{GSS}	Gate-to-Source Forward Leakage	—	—	100	nA	V_{GS}=20V
	Gate-to-Source Reverse Leakage	—	—	-100		V_{GS}=-20V
Q_g	Total Gate Charge	—	—	70	nC	I_D=18A
Q_{gs}	Gate-to-Source Charge	—	—	13		V_{DS}=160V
Q_{gd}	Gate-to-Drain ("Miller") Charge	—	—	39		V_{GS}=10V See Fig. 6 and 13 ④
$t_{d(on)}$	Turn-On Delay Time	—	14	—	ns	V_{DD}=100V
t_r	Rise Time	—	51	—		I_D=18A
$t_{d(off)}$	Turn-Off Delay Time	—	45	—		R_G=9.1Ω
t_f	Fall Time	—	36	—		R_D=5.4Ω See Figure 10 ④
L_D	Internal Drain Inductance	—	5.0	—	nH	Between lead, 6 mm (0.25in.) from package and center of die contact
L_S	Internal Source Inductance	—	13	—		
C_{iss}	Input Capacitance	—	1300	—	pF	V_{GS}=0V
C_{oss}	Output Capacitance	—	400	—		V_{DS}= 25V
C_{rss}	Reverse Transfer Capacitance	—	130	—		f=1.0MHz See Figure 5

Source-Drain Ratings and Characteristics

	Parameter	Min.	Typ.	Max.	Units	Test Conditions
I_S	Continuous Source Current (Body Diode)	—	—	20	A	MOSFET symbol showing the integral reverse p-n junction diode.
I_{SM}	Pulsed Source Current (Body Diode) ①	—	—	80		
V_{SD}	Diode Forward Voltage	—	—	2.0	V	T_J=25°C, I_S=20A, V_{GS}=0V ④
t_{rr}	Reverse Recovery Time	—	300	610	ns	T_J=25°C, I_F=18A
Q_{rr}	Reverse Recovery Charge	—	3.4	7.1	μC	di/dt=100A/μs ④
t_{on}	Forward Turn-On Time	Intrinsic turn-on time is neglegible (turn-on is dominated by L_S+L_D)				

Notes:

① Repetitive rating; pulse width limited by max. junction temperature (See Figure 11)

② V_{DD}=50V, starting T_J=25°C, L=1.9mH R_G=25Ω, I_{AS}=20A (See Figure 12)

③ I_{SD}≤18A, di/dt≤150A/μs, V_{DD}≤$V_{(BR)DSS}$, T_J≤150°C

④ Pulse width ≤ 300 μs; duty cycle ≤2%.

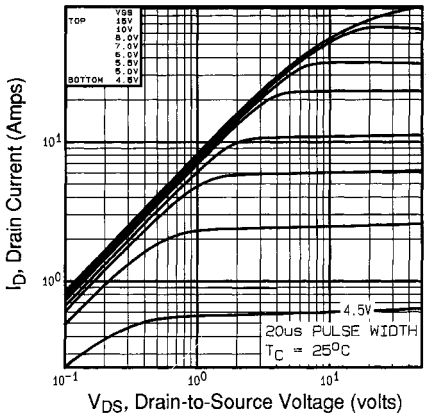


Fig 1. Typical Output Characteristics, T_C=25°C

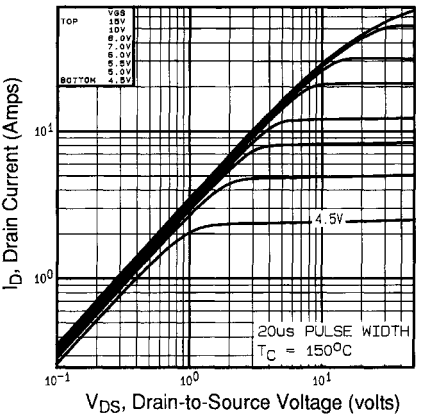


Fig 2. Typical Output Characteristics, T_C=150°C

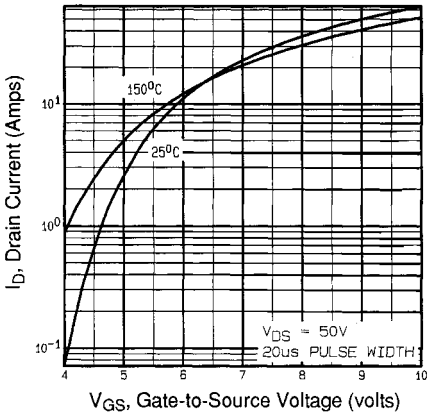


Fig 3. Typical Transfer Characteristics

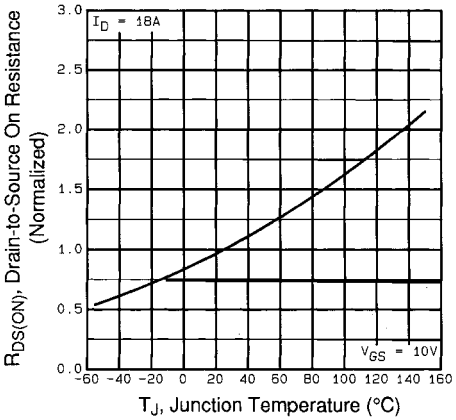


Fig 4. Normalized On-Resistance Vs. Temperature

DATA SHEETS

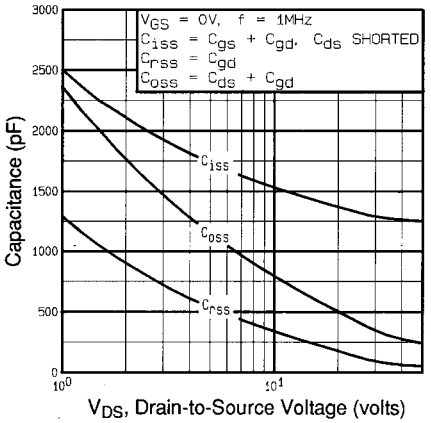


Fig 5. Typical Capacitance Vs. Drain-to-Source Voltage

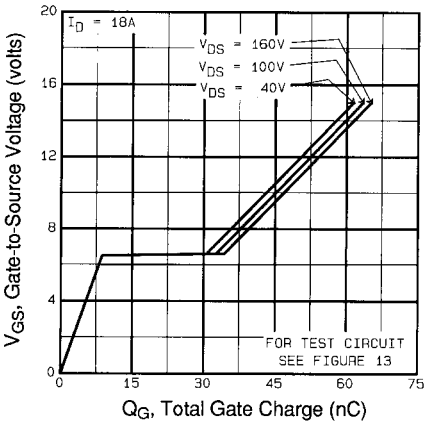


Fig 6. Typical Gate Charge Vs. Gate-to-Source Voltage

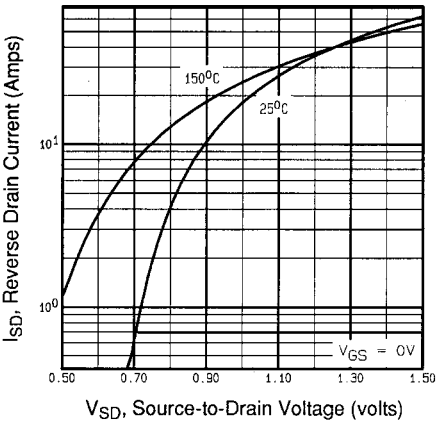


Fig 7. Typical Source-Drain Diode Forward Voltage

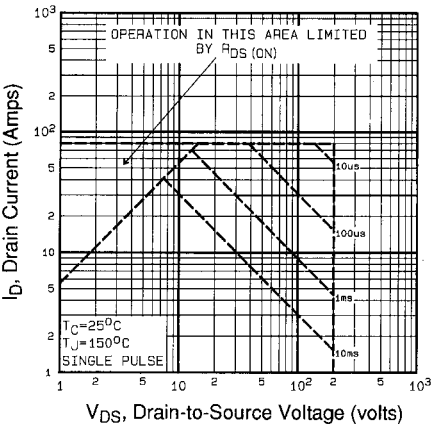


Fig 8. Maximum Safe Operating Area

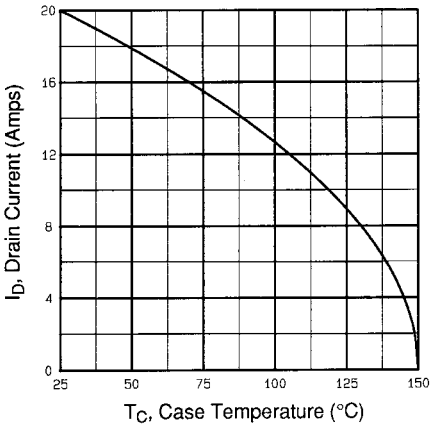


Fig 9. Maximum Drain Current Vs. Case Temperature

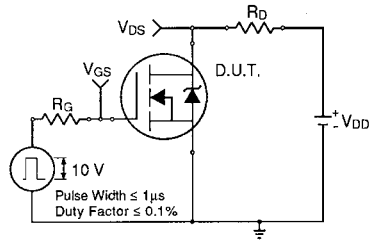


Fig 10a. Switching Time Test Circuit

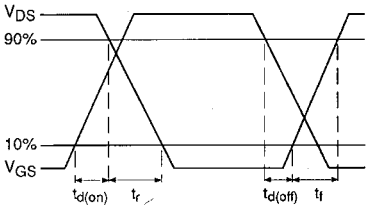


Fig 10b. Switching Time Waveforms

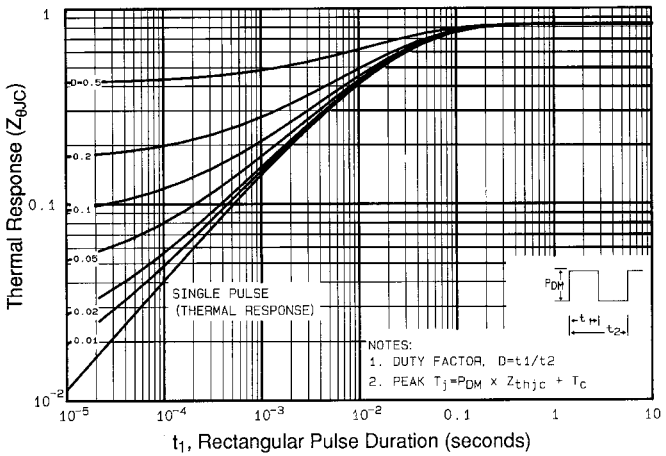


Fig 11. Maximum Effective Transient Thermal Impedance, Junction-to-Case

DATA SHEETS

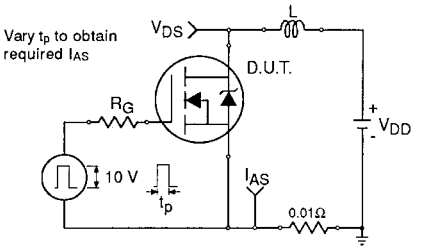


Fig 12a. Unclamped Inductive Test Circuit

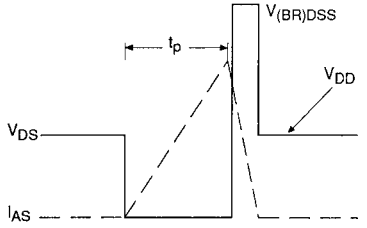


Fig 12b. Unclamped Inductive Waveforms

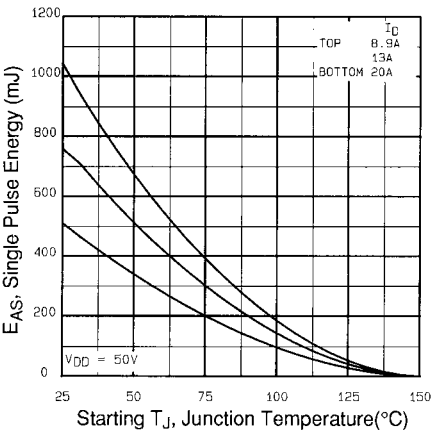


Fig 12c. Maximum Avalanche Energy Vs. Drain Current

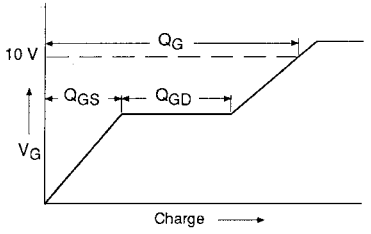


Fig 13a. Basic Gate Charge Waveform

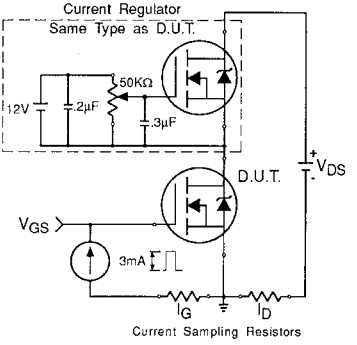


Fig 13b. Gate Charge Test Circuit

Appendix A: Figure 14, Peak Diode Recovery dv/dt Test Circuit – See page 1505

Appendix B: Package Outline Mechanical Drawing – See page 1511

Appendix C: Part Marking Information – See page 1517

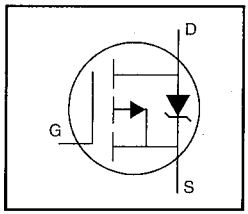
International
IOR Rectifier

International IOR Rectifier

IRFP9240

HEXFET® Power MOSFET

- Dynamic dv/dt Rating
- Repetitive Avalanche Rated
- P-Channel
- Isolated Central Mounting Hole
- Fast Switching
- Ease of Paralleling
- Simple Drive Requirements



V_{DSS} = -200V

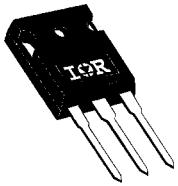
$R_{DS(on)}$ = 0.50Ω

I_D = -12A

Description

Third Generation HEXFETs from International Rectifier provide the designer with the best combination of fast switching, ruggedized device design, low on-resistance and cost-effectiveness.

The TO-247 package is preferred for commercial–industrial applications where higher power levels preclude the use of TO-220 devices. The TO-247 is similar but superior to the earlier TO-218 package because of its isolated mounting hole. It also provides greater creepage distance between pins to meet the requirements of most safety specifications.



TO-247AC

DATA SHEETS

Absolute Maximum Ratings

	Parameter	Max.	Units
I_D @ T_C = 25°C	Continuous Drain Current, V_{GS} @ -10 V	-12	A
I_D @ T_C = 100°C	Continuous Drain Current, V_{GS} @ -10 V	-7.5	
I_{DM}	Pulsed Drain Current ①	-48	
P_D @ T_C = 25°C	Power Dissipation	150	W
	Linear Derating Factor	1.2	W/°C
V_{GS}	Gate-to-Source Voltage	±20	V
E_{AS}	Single Pulse Avalanche Energy ②	790	mJ
I_{AR}	Avalanche Current ①	-12	A
E_{AR}	Repetitive Avalanche Energy ①	15	mJ
dv/dt	Peak Diode Recovery dv/dt ③	-5.0	V/ns
T_J T_{STG}	Operating Junction and Storage Temperature Range	-55 to +150	°C
	Soldering Temperature, for 10 seconds	300 (1.6mm from case)	
	Mounting Torque, 6-32 or M3 screw	10 lbf•in (1.1 N•m)	

Thermal Resistance

	Parameter	Min.	Typ.	Max.	Units
$R_{\theta JC}$	Junction-to-Case	—	—	0.83	°C/W
$R_{\theta CS}$	Case-to-Sink, Flat, Greased Surface	—	0.24	—	
$R_{\theta JA}$	Junction-to-Ambient	—	—	40	

Electrical Characteristics @ T_J = 25°C (unless otherwise specified)

	Parameter	Min.	Typ.	Max.	Units	Test Conditions
$V_{(BR)DSS}$	Drain-to-Source Breakdown Voltage	-200	—	—	V	V_{GS}=0V, I_D=-250μA
$\Delta V_{(BR)DSS}/\Delta T_J$	Breakdown Voltage Temp. Coefficient	—	-0.20	—	V/°C	Reference to 25°C, I_D=-1mA
$R_{DS(on)}$	Static Drain-to-Source On-Resistance	—	—	0.50	Ω	V_{GS}=-10V, I_D=-7.2A ④
$V_{GS(th)}$	Gate Threshold Voltage	-2.0	—	-4.0	V	V_{DS}=V_{GS}, I_D=-250μA
g_{fs}	Forward Transconductance	4.2	—	—	S	V_{DS}=-50V, I_D=-7.2A ④
I_{DSS}	Drain-to-Source Leakage Current	—	—	-100	μA	V_{DS}=-200V, V_{GS}=0V
		—	—	-500		V_{DS}=-160V, V_{GS}=0V, T_J=125°C
I_{GSS}	Gate-to-Source Forward Leakage	—	—	-100	nA	V_{GS}=-20V
	Gate-to-Source Reverse Leakage	—	—	100		V_{GS}=20V
Q_g	Total Gate Charge	—	—	44	nC	I_D=-11A
Q_{gs}	Gate-to-Source Charge	—	—	7.1		V_{DS}=-160V
Q_{gd}	Gate-to-Drain ("Miller") Charge	—	—	27		V_{GS}=-10V See Fig. 6 and 13 ④
$t_{d(on)}$	Turn-On Delay Time	—	14	—	ns	V_{DD}=-100V
t_r	Rise Time	—	43	—		I_D=-11A
$t_{d(off)}$	Turn-Off Delay Time	—	39	—		R_G=9.1Ω
t_f	Fall Time	—	38	—		R_D=8.6Ω See Figure 10 ④
L_D	Internal Drain Inductance	—	5.0	—	nH	Between lead, 6 mm (0.25in.) from package and center of die contact
L_S	Internal Source Inductance	—	13	—		
C_{iss}	Input Capacitance	—	1200	—	pF	V_{GS}=0V
C_{oss}	Output Capacitance	—	370	—		V_{DS}=-25V
C_{rss}	Reverse Transfer Capacitance	—	81	—		f=1.0MHz See Figure 5

Source-Drain Ratings and Characteristics

	Parameter	Min.	Typ.	Max.	Units	Test Conditions
I_S	Continuous Source Current (Body Diode)	—	—	-12	A	MOSFET symbol showing the integral reverse p-n junction diode.
I_{SM}	Pulsed Source Current (Body Diode) ①	—	—	-48		
V_{SD}	Diode Forward Voltage	—	—	-5.0	V	T_J=25°C, I_S=-12A, V_{GS}=0V ④
t_{rr}	Reverse Recovery Time	—	250	300	ns	T_J=25°C, I_F=-11A
Q_{rr}	Reverse Recovery Charge	—	2.9	3.6	μC	di/dt=100A/μs ④
t_{on}	Forward Turn-On Time	Intrinsic turn-on time is neglegible (turn-on is dominated by L_S+L_D)				

Notes:

① Repetitive rating; pulse width limited by max. junction temperature (See Figure 11)

② V_{DD}=-50V, starting T_J=25°C, L=8.2mH R_G=25Ω, I_{AS}=-12A (See Figure 12)

③ I_{SD}≤-12A, di/dt≤150A/μs, V_{DD}≤$V_{(BR)DSS}$, T_J≤150°C

④ Pulse width ≤ 300 μs; duty cycle ≤2%.

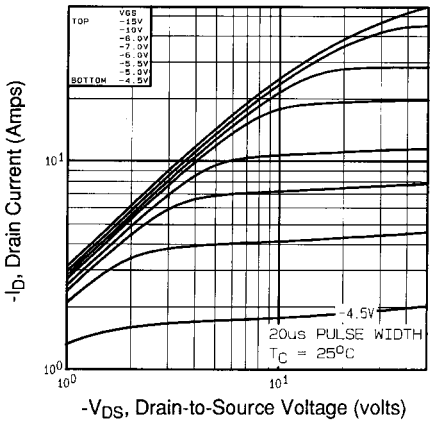


Fig 1. Typical Output Characteristics, T_C=25°C

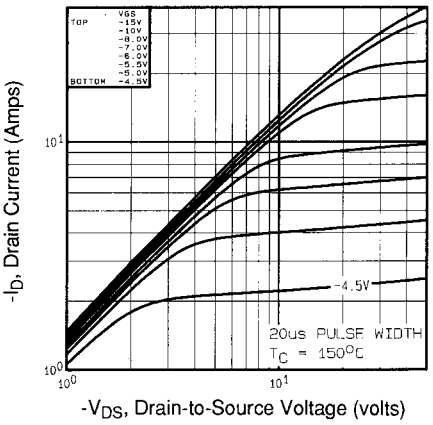


Fig 2. Typical Output Characteristics, T_C=150°C

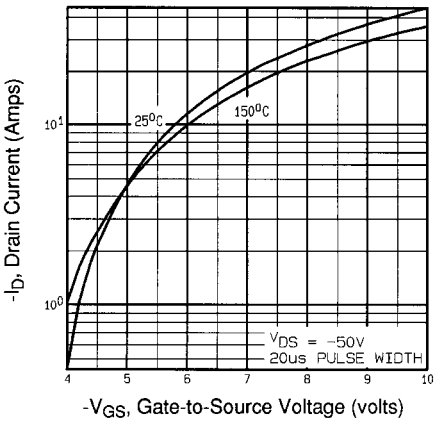


Fig 3. Typical Transfer Characteristics

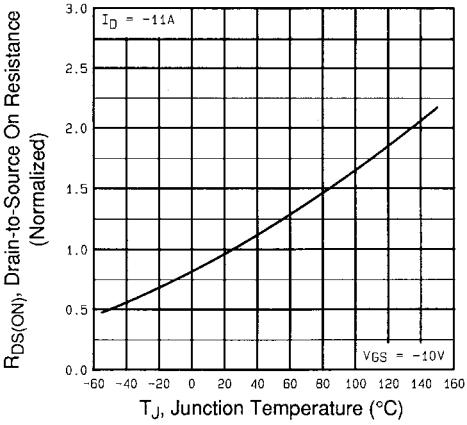


Fig 4. Normalized On-Resistance Vs. Temperature

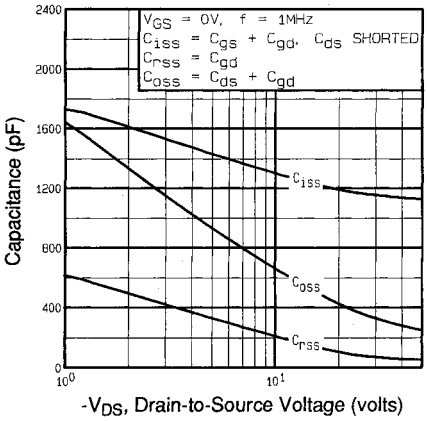


Fig 5. Typical Capacitance Vs. Drain-to-Source Voltage

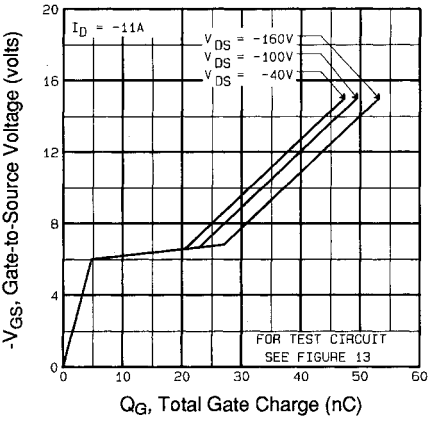


Fig 6. Typical Gate Charge Vs. Gate-to-Source Voltage

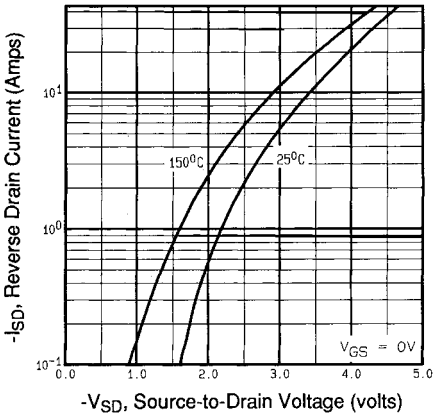


Fig 7. Typical Source-Drain Diode Forward Voltage

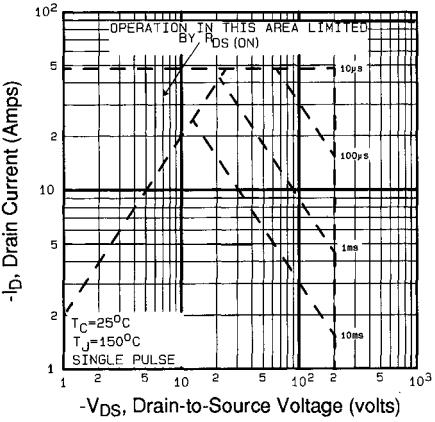


Fig 8. Maximum Safe Operating Area

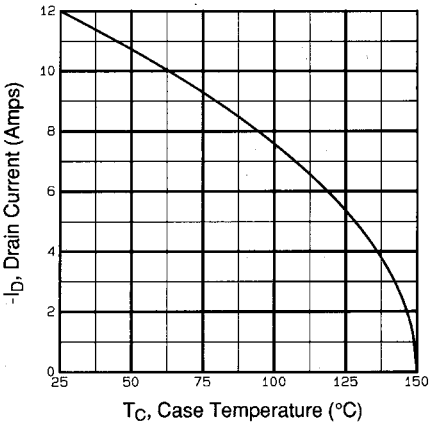


Fig 9. Maximum Drain Current Vs. Case Temperature

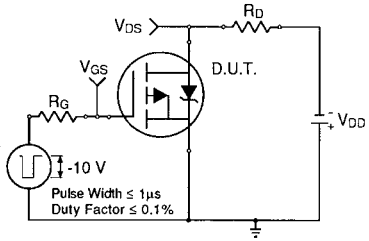


Fig 10a. Switching Time Test Circuit

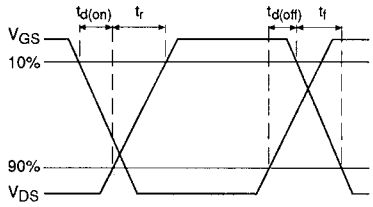


Fig 10b. Switching Time Waveforms

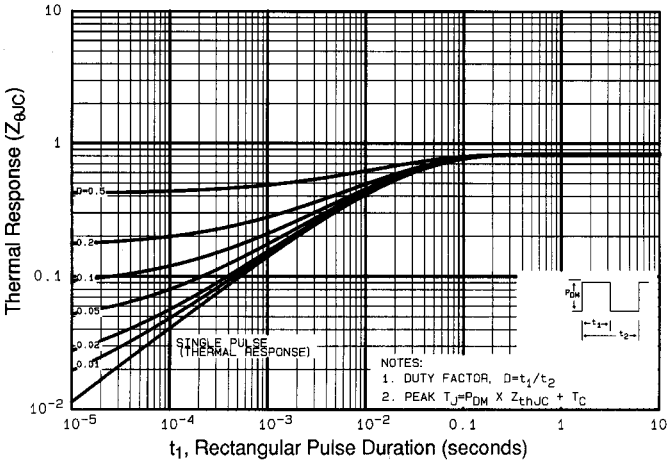


Fig 11. Maximum Effective Transient Thermal Impedance, Junction-to-Case

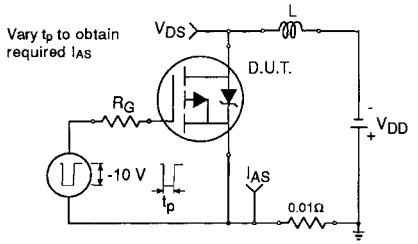


Fig 12a. Unclamped Inductive Test Circuit

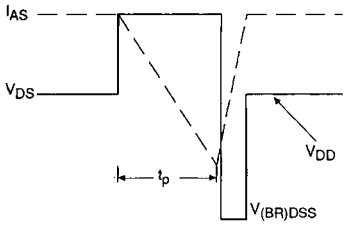


Fig 12b. Unclamped Inductive Waveforms

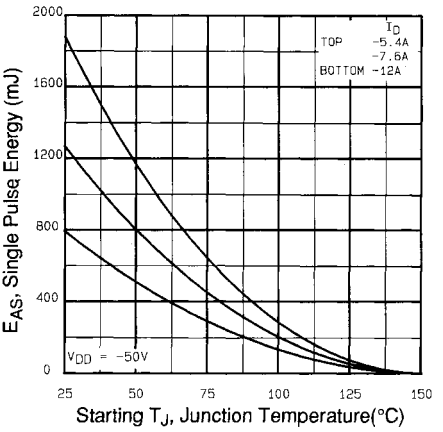


Fig 12c. Maximum Avalanche Energy Vs. Drain Current

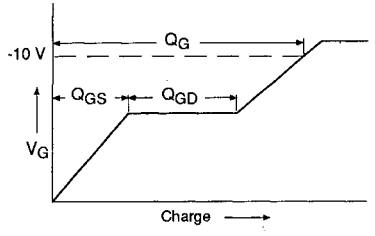


Fig 13a. Basic Gate Charge Waveform

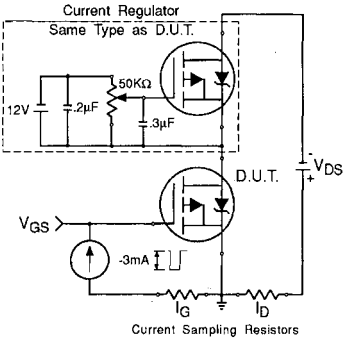


Fig 13b. Gate Charge Test Circuit

Appendix A: Figure 14, Peak Diode Recovery dv/dt Test Circuit – See page 1506

Appendix B: Package Outline Mechanical Drawing – See page 1511

Appendix C: Part Marking Information – See page 1517

International IOR Rectifier

International
IOR Rectifier

APPLICATION NOTE AN-936

International Rectifier • 233 Kansas Street, El Segundo, CA 90245 USA

The Do's and Don'ts of Using MOS-Gated Transistors

In this application note, some of the most common do's and don'ts of using HEXFET® power MOSFETs are described. The objective is to help the user get the most out of these remarkable devices, while reducing "on the job" learning time to a minimum.

1. Be Mindful of The Reverse Blocking Characteristics of The Device

IGBTs have a limited reverse blocking capability of approximately 20-30 V, with high leakage. This is characterized in IR's data sheets with a Reverse Avalanche Energy (EARV). This rating is useful to absorb energy spikes due to the stray inductance in series with the anti-parallel diode. This is a significant advantage over bipolar transistors and power darlingtons. A feature of power MOSFETs is that they inherently have built into them an integral reverse body-drain diode. The existence of this diode is explained by reference to Figure 1. When the source terminal is made positive with respect to the drain, current can flow through the middle of the source cell, across a forward biased P-N junction. In the "reverse" direction, the HEXFET® power MOSFET thus behaves like a P-N junction rectifier. The integral body-drain diode is a real circuit element, and its current handling capability is typically as high as that of the transistor itself. Some circuits require an "inverse" rectifier to be connected across the switching device, and in these circuits it will often be possible to utilize the body-drain diode of the HEXFET power MOSFET provided the proper precautions are taken.

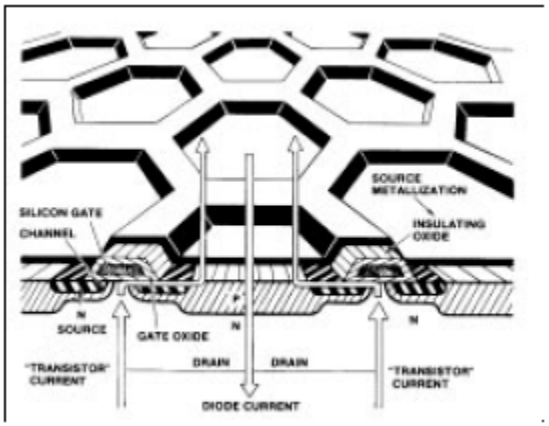


Figure 1. Basic HEXFET MOSFET Structure

2. Be Careful When Handling & Testing HEXFET® Power MOSFETs

The user's first "contact" with a MOS-gated transistor could be a package of parts arriving on his desk. Even at this stage, it behooves one to be knowledgeable about some elementary precautions. Being MOS devices, HEXFET power MOSFETs can be damaged by static charge when handling, testing or installing into a circuit. Power Devices have large input capacitance, and are able to absorb static charge without excessive buildup of voltage. In order to avoid possible problems, however, the following procedures should be followed as a matter of good practice, wherever possible:

- MOS-gated transistors should be left in their anti-static shipping bags, or conductive foam, or they should be placed in metal containers or conductive tote bins, until required for testing or connection into a circuit. The person handling the device should ideally be grounded through a suitable wrist strap, though in reality this added precaution is seldom essential.
- Devices should be handled by the package, not by the leads. When checking the electrical characteristics of the MOS-gated transistors on a curve tracer, or in a test circuit, the following precautions should be observed:
- Test stations should use electrically conductive floor and table mats that are grounded. Suitable mats are available commercially.

HEXFET® is the registered trademark for International Rectifier Power MOSFETs

- When inserting the device in a curve tracer or a test circuit, voltage should not be applied until all terminals are solidly connected into the circuit.
- When using a curve tracer, a resistor should be connected in series with the gate to damp spurious oscillations that can otherwise occur on the trace. A suitable value of resistance is 100 ohms.
- For repeated testing, it is convenient to build this resistor into the test fixture.
- When switching from one test range to another, voltage and current settings should be reduced to zero, to avoid the generation of potentially destructive voltage surges during switching.

The next step is to connect the device into an actual circuit. The following simple precautions should be observed:

- Work stations should use electrically grounded table and floor mats.
- Soldering irons *should be grounded.*

Now that the device has been connected into its circuit, it is ready for the power to be applied. From here on, success in applying the device becomes a matter of the integrity of the circuit design, and of what circuit precautions have been taken to guard against unintentional abuse of its ratings.

The following are the interrelated device and circuit considerations that lead to reliable, trouble-free design.

3. BEWARE OF UNEXPECTED GATE-TO-SOURCE VOLTAGE SPIKES

Excessive voltage will punch through the gate-source oxide layer and result in permanent damage. This seems obvious enough, but it is not so obvious that transient gate-to-source overvoltages can be generated that are quite unrelated to, and well in excess of, the amplitude of the applied drive signal. The problem is illustrated by reference to Figure 2.
If we assume that the impedance, Z, of the drive source is high, then any positive-going change of voltage applied across the drain and source terminals (caused, for example, by the switching of another device in the circuit) will be reflected as a positive-going voltage transient across the source and the drain terminals, in the approximate ratio of:

$$\frac{1}{1+\frac{C_{gs}}{C_{dg}}}$$

The above ratio is typically about 1 to 6. This means that a change of drain-to-source voltage of 300V, for example, could produce a voltage transient approaching 50V between the gate and source terminals. In practice this “aiming” voltage will not appear on the gate if the dv/dt is positive because the MOS-gated device goes in conduction at approximately Vgs = 4V, thereby clamping the dv/dt at the expense of a current transient and increased power dissipation. However, a negative-going dv/dt will not be clamped. This calculation is based upon the worst case assumption that the transient impedance of the drive circuit is high by comparison with the gate-to-source capacitance of the device. This situation can, in fact, be quite easily approximated if the gate drive circuit contains inductance—for example the leakage inductance of an isolating drive transformer. This inductance exhibits a high impedance for short transients, and effectively decouples the gate from its drive circuit for the duration of the transient.

The negative-going gate-to-source voltage transient produced under the above circumstances may exceed the gate voltage rating of the device, causing permanent damage. It is, of course, true that since the applied drain transient results in a voltage at the gate which tends to turn the device ON, the overall effect is to an extent self-limiting so far as the gate voltage transient is concerned. Whether this self-limiting action will prevent the voltage transient at the gate from exceeding the gate-source voltage rating of the device depends upon the impedance of the external circuit. Spurious turn-on is of itself undesirable, of course, though in practical terms one may grudgingly be able to accept this circuit operating imperfection, provided the safe operating area of the device is not violated.

Notice that a voltage clamp (a conventional zener diode is suitable for this purpose) to prevent the gate-source voltage rating from being exceeded will not prevent the dv/dt induced turn-on, as the gate will not reach the zener voltage. In many instances the zener is responsible for generating oscillations in the gate circuit, particularly when a significant amount of stray inductance is present. A more fundamental solution, of course, is to make the impedance of the gate circuit low enough that not only is the gate-source voltage rating not exceeded, but also the voltage transient at the gate is contained to a level at which spurious turn-on does not occur.

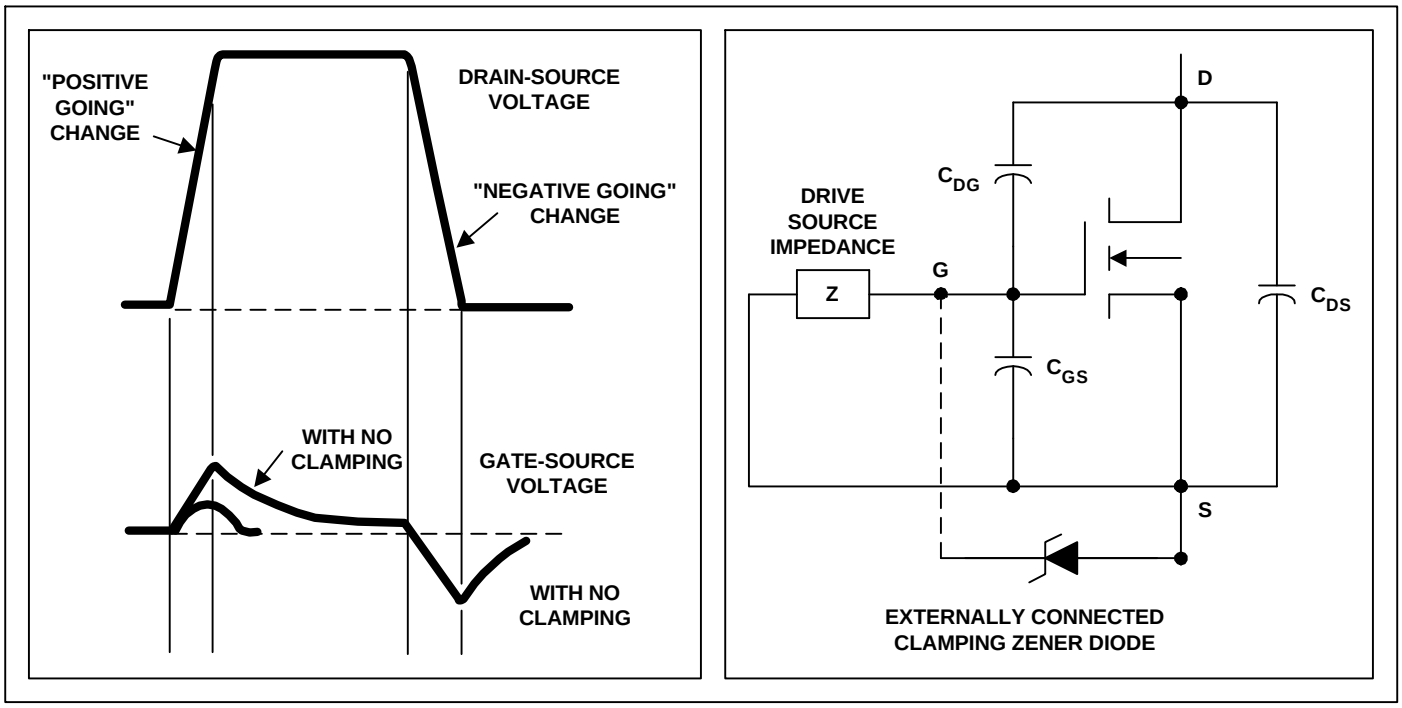


Figure 2. A Rapidly Changing Applied Drain-Source Voltage will Produce Gate-Source Transients

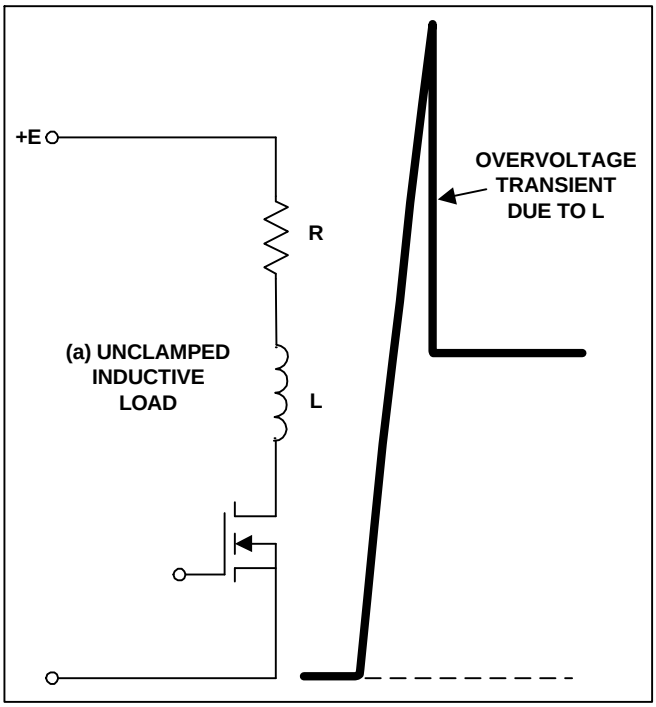


Figure 3. Drain-Source Overvoltage Transient when Switching Off with Unclamped Inductive Load

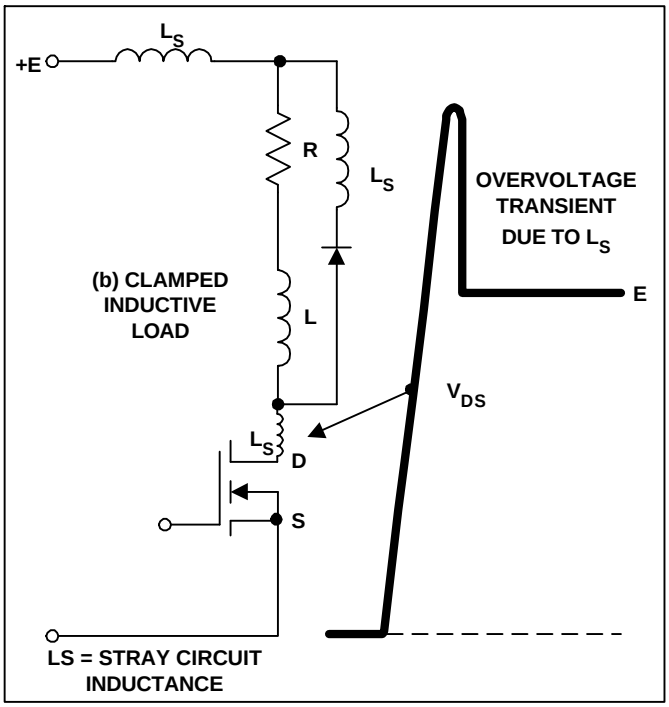


Figure 4. Drain-Source Overvoltage Transient Produced by Stray Circuit Inductance When Switching Off with Clamped, Inductive Load

It should be remembered that a collapse of voltage across the device (i.e., a negative-going dv/dt) will produce a transient negative voltage spike across the gate-source terminals. In this case, of course, there will be no tendency for the device to turn ON, and hence no tendency for the effect to be self-limiting. A zener diode connected to clamp positive transients will automatically clamp negative-going transients, limiting them to the forward conduction voltage drop of the zener.

4. BEWARE OF DRAIN OR COLLECTOR VOLTAGE SPIKES INDUCED BY SWITCHING

The uninitiated designer is often not aware that self-inflicted overvoltage transients can be produced when the device is switched OFF, even though the DC supply voltage for the drain circuit is well below the V_{DS} rating of the transistor.
Figure 3 shows how a voltage spike is produced when switching the device OFF, as a result of inductance in the circuit. The faster the device is switched, the higher the overvoltage will be.
Inductance is always present to some extent in a practical circuit, and therefore, there is always danger of inducing overvoltage transients when switching OFF. Usually, of course, the main inductive component of the load will be "clamped", as shown in Figure 4. Stray circuit inductance still exists, however, and overvoltage transients will still be produced as a result—to say nothing of the fact that the clamping diode may not provide an instantaneous clamping action, due to its "forward recovery" characteristic.

The first approach to this problem is to minimize stray circuit inductance, by means of careful attention to circuit layout, to the point that whatever residual inductance is left in the circuit can be tolerated. HEXFET®s have an inductive energy rating that makes capable of withstanding these inductive spikes, assuming that the data sheet limits for energy and temperature are not violated. IGBTs, however, do not have an avalanche rating, and a clamping device should be connected, physically as close as possible to the drain and source terminals, as shown in Figure 5. A conventional zener diode, or a "transorb" clamping device, are satisfactory for this purpose. An alternative clamping circuit is shown in Figure 6, depending on the voltage and current rating of the circuit.
The capacitor C is a reservoir capacitor and charges to a substantially constant voltage, while the resistor R is sized to dissipate the "clamping energy" while maintaining the desired voltage across the capacitor. The diode D must be chosen so that its forward recovery characteristic does not significantly spoil the transient clamping action of the circuit. A simple RC snubber can also be used, as shown in Figure 7. Note, however, that an RC snubber not only limits the peak voltage, it also slows down the effective switching speed. In so doing, it absorbs energy during the whole of the switching period, not just at the end of it, as does a voltage clamp. A snubber is therefore less efficient than a true voltage clamping device.

Note that the highest voltage transient occurs when switching the highest level of current. The waveform of the voltage across the device should be checked with a high-speed oscilloscope at the full load condition to ensure that switching voltage transients are within safe limits.

5. DO NOT EXCEED THE PEAK CURRENT RATING

All power transistors have a specified maximum peak current rating. This is conservatively set at a level that guarantees reliable operation and it should not be exceeded. It is often overlooked that, in a practical circuit, peak transient currents can be obtained that are well in excess of the expected normal operating current, unless proper precautions are taken. Heating, lighting and motor loads, for example, consume high in-rush currents if not properly controlled. A technique that ensures that the peak current does not exceed the capability of the device is to use a current sensing control that switches it OFF whenever the current instantaneously reaches a preset limit.

Unexpectedly high transient current can also be obtained as a result of rectifier reverse recovery, when a transistor is switched ON rapidly into a conducting rectifier. This is illustrated in Figure 8. The solution is to use a faster rectifier, or to slow down the switching of the transistor to limit the peak reverse recovery current of the rectifier.

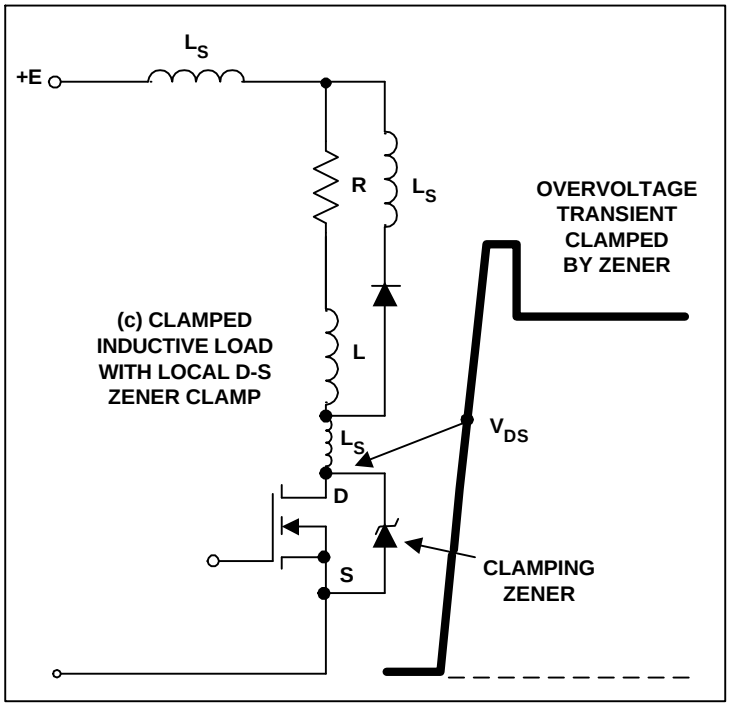


Figure 5. Overvoltage Transient at Switch-Off Clamped by Local Drain-Source Zener

6. STAY WITHIN THE THERMAL LIMITS

Power transistors are thermally limited. They must be mounted on a heatsink that is adequate to keep the junction temperature within the rated under the "worst case" condition of maximum power dissipation and maximum ambient temperature.

It must be remembered that in a switching application, the total power is due to the conduction losses and the switching loss. Switching time and switching losses of HEXFET®s are essentially independent of temperature, but the conduction losses increase with increasing temperature, because $R_{DS(on)}$ increases with temperature. IGBTs, on the contrary, have switching losses that highly dependent of temperature, while conduction losses are not. This must be taken into account when sizing the heatsink. The required thermal resistance of the heatsink can be calculated as follows:

The transistor conduction power, P_T, is given approximately by *P_T = On-state Voltage x Drain or Collector current*

The switching energy depends upon the voltage and current being switched and the type of load. The total switching loss, P_S, is the total switching energy, ε_T, multiplied by the operating frequency, f. e_T is the sum of the energies due to the individual switchings that take place in each fundamental operating cycle:

$$P_S = \varepsilon_T f$$

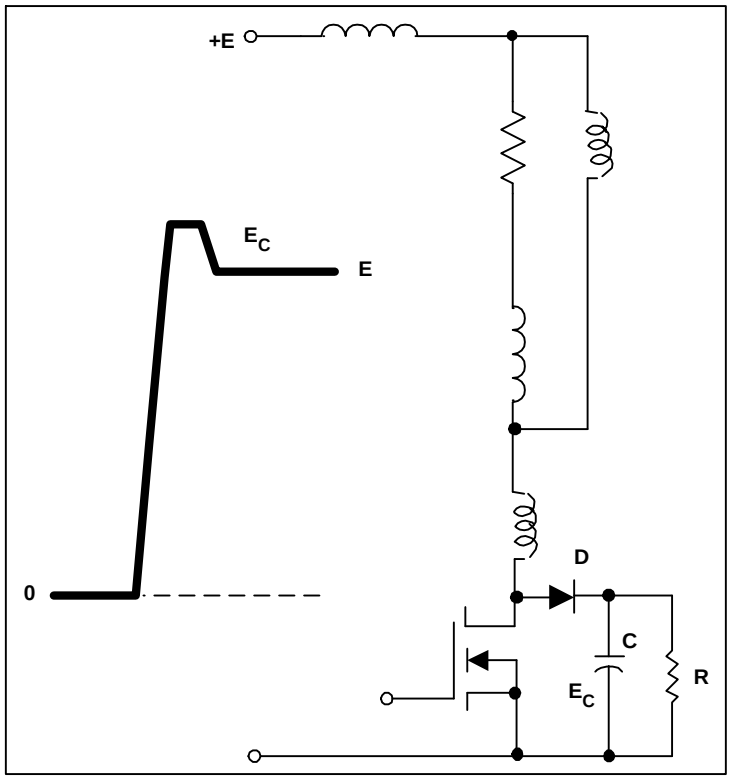


Figure 6. Overvoltage Transient at Switch-Off Limited by Local Clamp

The total power dissipation is the sum of the conduction power, P_T, and the switching power, P_S.

$$P = P_T + P_S$$

Since:

$$\Delta T_{JA} = PR_{th}$$

where:

R_{th} = junction-to-ambient thermal resistance

The junction-to-ambient thermal resistance, R_{JA}, is made up of the internal junction-to-case thermal resistance, R_{JC}, plus the case-to-heatsink thermal resistance, R_{CS}, plus the sink-to-ambient thermal resistance, R_{SA}. The first two terms are fixed for the device, and the required thermal resistance of the heatsink, R_{S-A}, for a given junction temperature rise DT_{J-A}, can be calculated from:

$$R_{S-A} = R_{J-A} - (R_{JC} + R_{C-S})$$

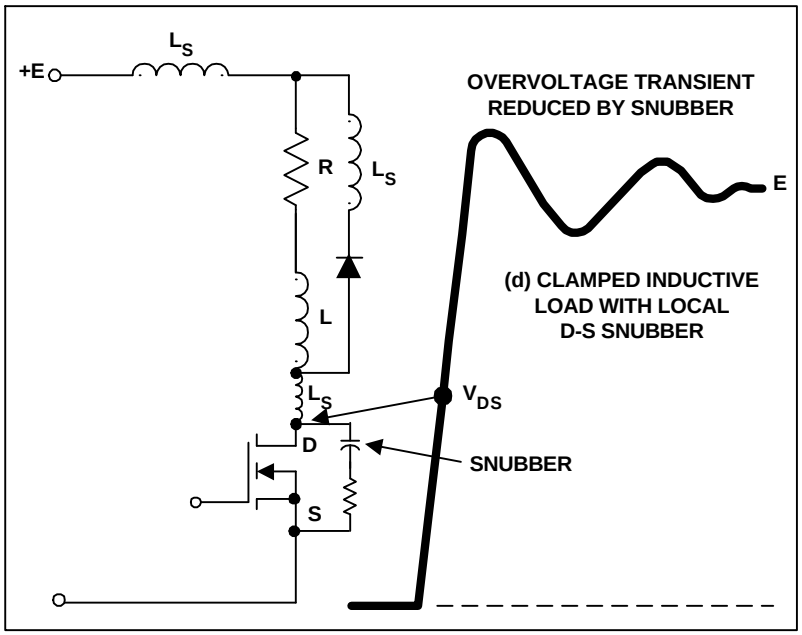


Figure 7. Overvoltage Transient at Switch-Off Limited by Local Capacitor-Resistor Snubber

7. PAY ATTENTION TO CIRCUIT LAYOUT

Stray inductance in the circuit can cause overvoltage transients, slowing down of the switching speed, unexpected unbalance of current between parallel connected devices, and unwanted oscillations.

In order to minimize these effects, stray circuit inductance must be minimized. This is done by keeping conduction paths as short as possible, by minimizing the area of current loops, by using twisted pairs of leads, and by using ground plane construction. Local decoupling capacitors alleviate the affects of any residual circuit inductance, once these measures have been taken. Circuit layout should be kept as symmetrical as possible in order to maintain balanced currents in parallel connected HEXFET®s or IGBTs. The gates of parallel connected devices should be decoupled by small ferrite beads placed over the gate connections, or by individual resistors in series with each gate. These measures prevent parasitic oscillations.

8. BE CAREFUL WHEN USING THE INTEGRAL BODY-DRAIN DIODE

The HEXFET®'s integral body-drain diode exhibits minority carrier reverse recovery. Reverse recovery presents a potential problem when switching any rectifier off; the slower the rectifier, the greater the problem. By comparison with the HEXFET® itself, the switching speed of the integral reverse rectifier is quite slow. The switching speed of a circuit which utilizes the body-drain diode of the HEXFET® may therefore be limited by the rectifier. Whether this will be so depends upon the circuit and the operating conditions.

Regardless of the overall circuit configuration, or the particular application, the "local" circuit operating situation that is troublesome occurs when the freewheeling current from an inductive load is commutated from the integral rectifier of one HEXFET® to the transistor of an "opposite" HEXFET®, the two devices forming a tandem series connected pair across a low impedance voltage source, as shown in Figure 8. This “local” circuit configuration occurs in most chopper and inverter schemes.

If the incoming HEXFET® switches ON too rapidly, the peak reverse recovery current of the integral body-drain diode of the opposite HEXFET® will rise too rapidly, the peak reverse recovery current rating will be exceeded, and the device may possibly be destroyed.

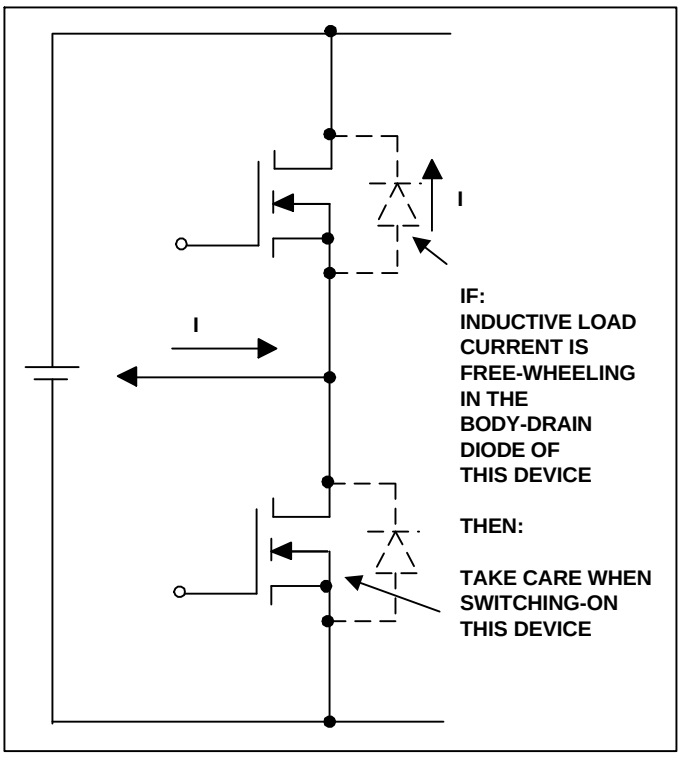


Figure 8. Local Circuit Configuration and Operating Condition Requiring Special Care When Using the HEXFET's Integral Body-Drain Diode.

The peak reverse recovery current of the rectifier can be reduced by slowing down the rate of change of current during the commutation process. The rate of change of current can be controlled by purposefully slowing down the rate of rise of the gate driving pulse. Using this technique, the peak current can be reduced to almost any desired extent, at the expense of prolonging the high dissipation switching period.

The oscillograms in Figure 9 illustrate the effect. By slowing the total switch-ON time from 300ns to 1.8ms, the peak current of the IRF330 has been decreased from 20A to 10A. The energy dissipation associated with the “unrestrained” switch-ON in Figure 9(a) is 0.9mJ, whereas it is 2.7mJ for the controlled switch-ON of Figure 9(b).

Note also that it is not necessary to slow the switching-OFF of the HEXFET®, hence the energy dissipation at switch-OFF will be relatively small by comparison with that at switch-ON. For operation at frequencies up to a few kHz, where ultra-fast switching is not mandatory, slowing the applied gate drive signal to reduce the peak reverse recovery current of the "opposite" rectifier offers a good practical solution.

9. BE ON YOUR GUARD WHEN COMPARING CURRENT RATINGS

The user can be forgiven if he assumes that the continuous drain current rating, that appears on the data sheet represents the current at which the device can actually be operated continuously in a practical system. To be sure, that's what it should represent; unfortunately it often does not.

Frequently a "continuous" current rating is assigned to the device which in practical terms cannot be used, because the resulting conduction power dissipation would be so large as to require a heatsink with an impractically low thermal resistance, and/or an impractically low ambient operating temperature. The best advice to the user is to compare different types *on the basis of high temperature conduction and switching losses*, and not of current rating. For MOSFETS, it is sufficient to compare $R_{DS(on)}$ at 25° C, and this provides a common basis for comparison. This parameter, taken in conjunction with the junction-case thermal resistance, is a much better indication of the power MOSFET true current handling capability.

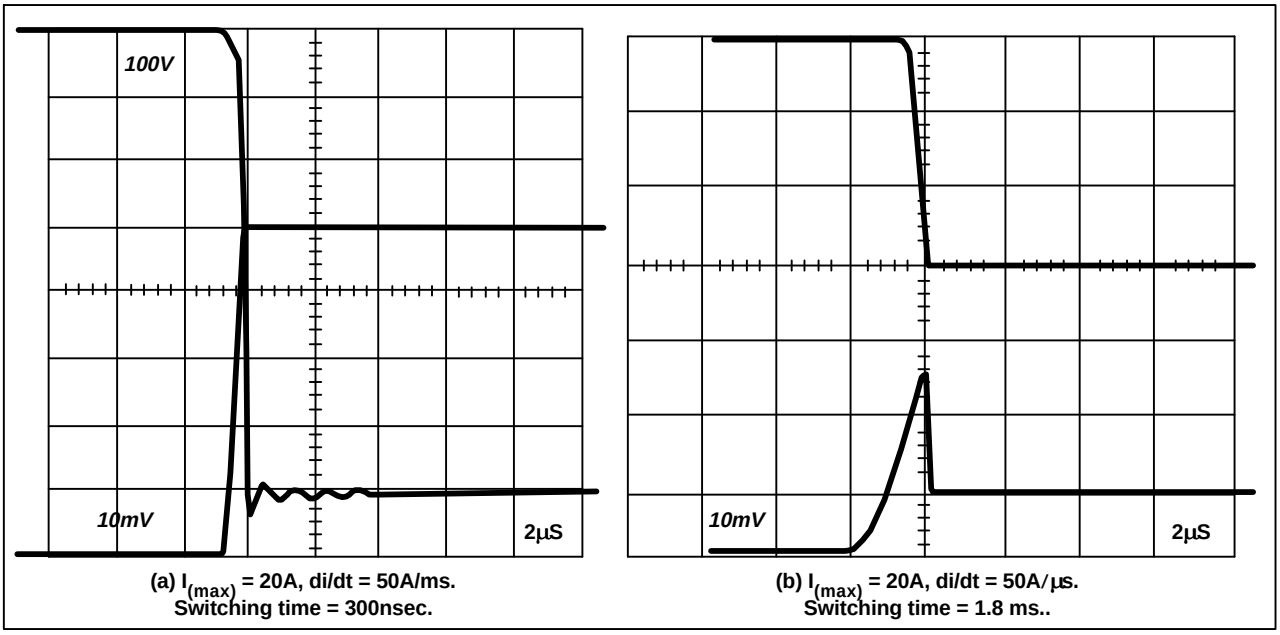


Figure 9. Oscillograms of IRF330 Switching into Reverse Rectifier of Another IRF330 with Freewheeling Current of 4A.
Top Trace: Voltage 100V/div.
Bottom Trace: Current 4A/div.
Time Scale: 2ms/div.

International
IOR Rectifier

Application Note AN-937

Gate Drive Characteristics and Requirements for HEXFET® Power MOSFETs

Table of Contents

Page

1. Gate Drive Vs Base Drive 1

2. Gate Voltage Limitations 2

3. The Impedance of the Gate Circuit 2

4. Driving Standard HEXFET® MOSFETs from TTL 5

5. Driving Standard HEXFET® MOSFETs from CMOS 5

6. Driving HEXFET® Power MOSFETs from Linear Circuits 6

7. Drive Circuits Not Referenced to Ground 7

8. Drive Requirements and Switching Characteristics of Logic Level HEXFET® MOSFETs 14

9. Simple and Inexpensive Methods to Generate Isolated Gate Drive Supplies 19

10. Photovoltaic Generators as Gate Drivers 20

11. Resonant Gate Drive Techniques 21

Gate Drive Characteristics and Requirements for HEXFET®s

Topics covered:

Gate drive vs base drive
Enhancement vs Depletion
N vs P-Channel
Max gate voltage
Zener diodes on gate?
The most important factor in gate drive: the impedance of the gate drive circuit
Switching 101 or Understanding the waveforms
What happens if gate drive impedance is high? dv/dt induced turn-on
Can a TTL gate drive a standard HEXFET®?
The universal buffer
Power dissipation of the gate drive circuit is seldom a problem
Can a C-MOS gate drive a standard HEXFET®?
Driving HEXFET®s from linear circuits
Drive circuits not referenced to ground
Gate drivers with optocouplers
Gate drive supply developed from the drain of the power device
Gate drivers with pulse transformers
Gate drivers with choppers
Drive requirements of Logic Level HEXFET®s
How fast is a Logic Level HEXFET®driven by a logic circuit?
Simple and inexpensive isolated gate drive supplies
A well-kept secret: Photovoltaic generators as gate drivers
Driving in the MHz? Use resonant gate drivers
Related topics

(Note: Most of the gate drive considerations and circuits are equally applicable to IGBTs. Only MOSFETs are mentioned for the sake of simplicity. Special considerations for IGBTs are contained in INT-990)

1. GATE DRIVE VS BASE DRIVE

The conventional bipolar transistor is a current-driven device. As illustrated in Figure 1(a). a current must be applied between the base and emitter terminals to produce a flow of current in the collector. The amount of a drive required to produce a given output depends upon the gain, but invariably a current must be made to flow into the base terminal to produce a flow of current in the collector.

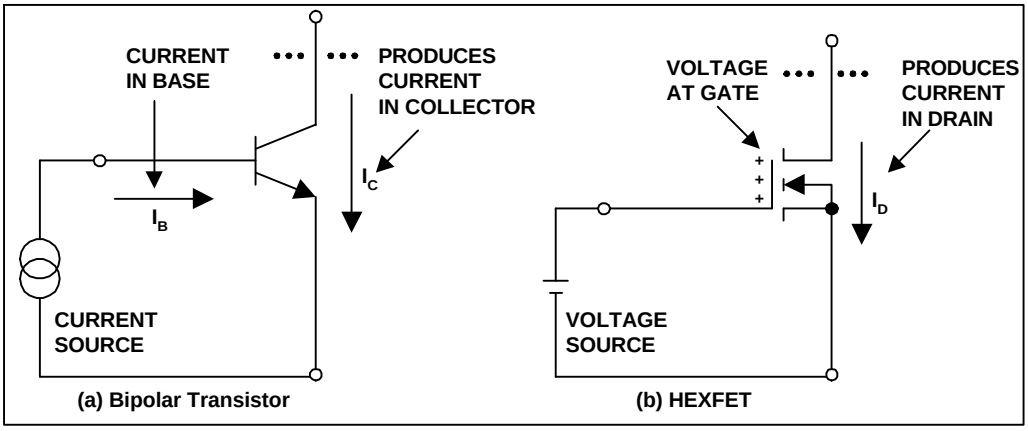


Figure 1. Bipolar Transistor is Current Driven, HEXFET is Voltage Driven

The HEXFET®is fundamentally different: it is a voltage-controlled power MOSFET device. A voltage must be applied between the gate and source terminals to produce a flow of current in the drain (see Figure 1b). The gate is isolated electrically from the source by a layer of silicon dioxide. Theoretically, therefore, no current flows into the gate when a DC voltage is applied to it - though in practice there will be an extremely small current, in the order of nanoamperes. With no voltage applied between the gate and source electrodes, the impedance between the drain and source terminals is very high, and only the leakage current flows in the drain.

When a voltage is applied between the gate and source terminals, an electric field is set up within the HEXFET®. This field "inverts" the channel (Figure 2) from P to N, so that a current can flow from drain to source in an uninterrupted sequence of N-type silicon (drain-channel-source). Field-effect transistors can be of two types: enhancement mode and depletion mode. Enhancement-mode devices need a gate voltage of the same sign as the drain voltage in order to pass current.

Depletion-mode devices are naturally on and are turned off by a gate voltage of the same polarity as the drain voltage. All HEXFET®s are enhancement-mode devices.

All MOSFET voltages are referenced to the source terminal. An N-Channel device, like an NPN transistor, has a drain voltage that is positive with respect to the source. Being enhancement-mode devices, they will be turned on by a positive voltage on the gate. The opposite is true for P-Channel devices, that are similar to PNP transistors.

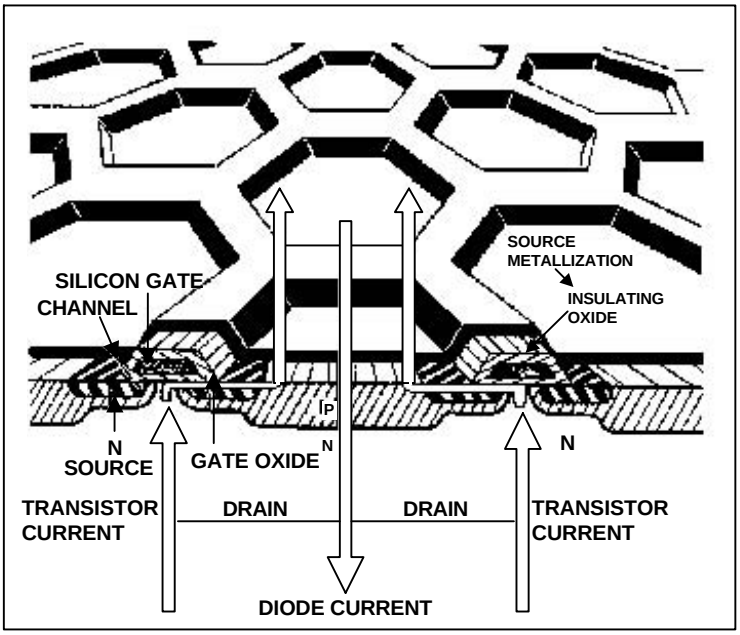


Figure 2. Basic HEXFET Structure

Although it is common knowledge that HEXFET®transistors are more easily driven than bipolars, a few basic considerations have to be kept in mind in order to avoid a loss in performance or outright device failure.

2. GATE VOLTAGE LIMITATIONS

Figure 2 shows the basic HEXFET®structure. The silicon oxide layer between the gate and the source regions can be punctured by exceeding its dielectric strength. The data sheet rating for the gate-to-source voltage is between 10 and 30 V for most HEXFET®s.

Care should be exercised not to exceed the gate-to-source maximum voltage rating. Even if the applied gate voltage is kept below the maximum rated gate voltage, the stray inductance of the gate connection, coupled with the gate capacitance, may generate ringing voltages that could lead to the destruction of the oxide layer. Overvoltages can also be coupled through the drain-gate self-capacitance due to transients in the drain circuit. A gate drive circuit with very low impedance insures that the gate voltage is not exceeded in normal operation. This is explained in more detail in the next section.

Zeners are frequently used "to protect the gate from transients". Unfortunately they also contribute to oscillations and have been known to cause device failures. A transient can get to the gate from the drive side or from the drain side. In either case, it would be an indication of a more fundamental problem: a high impedance drive circuit. A zener would compound this problem, rather than solving it. Sometimes a zener is added to reduce the ringing generated by the leakage of a gate drive transformer, in combination with the input capacitance of the MOSFET. If this is necessary, it is advisable to insert a small series resistor (5-10 Ohms) between the zener and the gate, to prevent oscillations.

3. THE IMPEDANCE OF THE GATE CIRCUIT

To turn on a power MOSFET a certain charge has to be supplied to the gate to raise it to the desired voltage, whether in the linear region, or in the "saturation" (fully enhanced) region. The best way to achieve this is by means of a voltage source, capable of supplying any amount of current in the shortest possible time. If the device is operated as a switch, a large transient current capability of the drive circuit reduces the time spent in the linear region, thereby reducing the switching losses.

On the other hand, if the device is operated in the linear mode, a large current from the gate drive circuit minimizes the relevance of the Miller effect, improving the bandwidth of the stage and reducing the harmonic distortion. This can be better understood by analyzing the basic switching waveforms at turn-on and turn-off for a clamped inductive load, as shown in Figures

3 and 5. Figure 3 shows the waveforms of the drain current, drain-to-source voltage and gate voltage during the turn-on interval. For the sake of simplicity, the equivalent impedance of the drive circuit has been assumed as purely resistive.

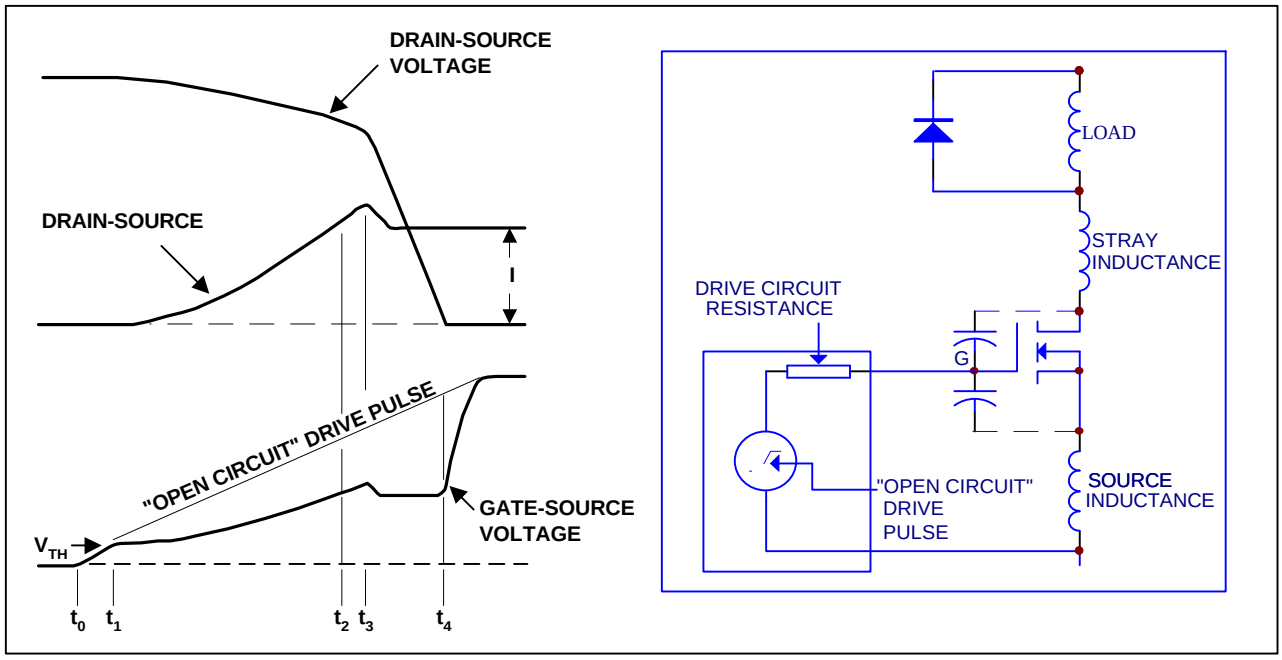


Figure 3. Waveforms at Turn-On

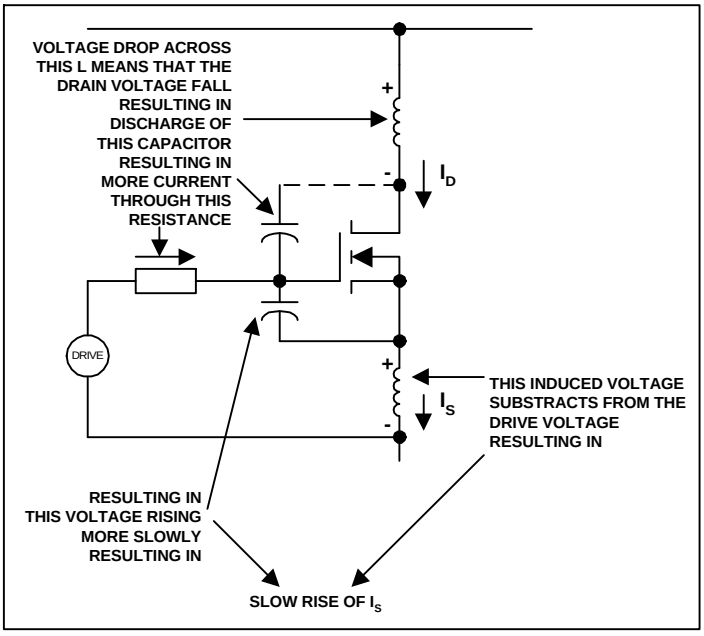


Figure 4. Diagrammatic Representation of Effects When Switching-ON

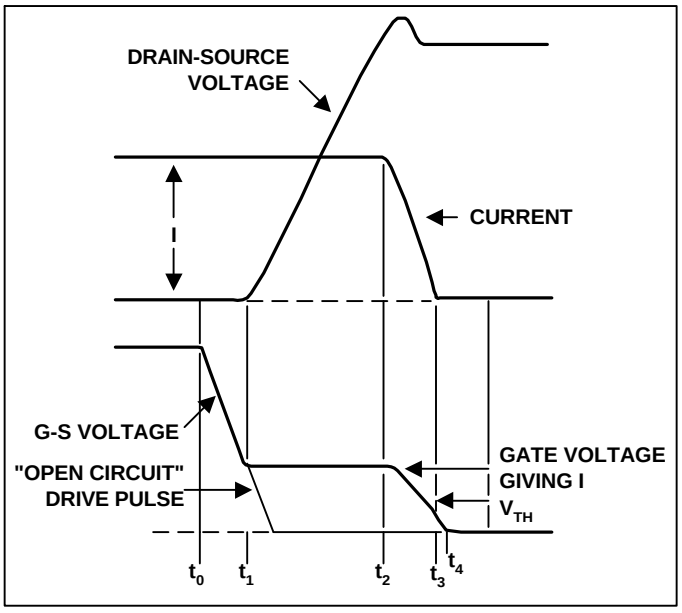


Figure 5. Waveforms at Turn-OFF

At time, t_0, the drive pulse starts to rise. At t_0 it reaches the threshold voltage of the HEXFET®s and the drain current starts to increase. At this point, two things happen which make the gate-source voltage waveform deviate from its original “path”. First, inductance in series with the source which is common to the gate circuit (“common source inductance”) develops an induced voltage as a result of the increasing source current. This voltage counteracts the applied gate drive voltage, and slows down the rate of rise of voltage appearing directly across the gate and source terminals; this in turn slows down the rate of rise of the source current. This is a negative feedback effect: increasing current in the source produces a counteractive voltage at the gate, which tends to resist the change of current.

The second factor that influences the gate-source voltage is the so called “Miller” effect. During the period t_1 to t_2 some voltage is dropped across “unclamped” stray circuit inductance in series with the drain, and the drain-source voltage starts to fall. The

decreasing drain-source voltage is reflected across the drain-gate capacitance, pulling a discharge current through it, and increasing the effective capacitive load on the drive circuit.
This in turn increases the voltage drop across the source impedance of the drive circuit, and decreases the rate of rise of voltage appearing between the gate and source terminals. Obviously, the lower the impedance of the gate drive circuit, the less this effect will be. This also is a negative feedback effect; increasing current in the drain results in a fall of drain-to-source voltage, which in turn slows down the rise of gate-source voltage, and tends to resist the increase of drain current. These effects are illustrated diagramatically in Figure 4. This state of affairs continues throughout the period t_1 to t_2, as the current in the HEXFET®rises to the level of the current, I_M, already flowing in the freewheeling rectifier, and it continues into the next period, t_2 to t_3, when the freewheeling rectifier goes into reverse recovery.

Finally, at time t_3 the freewheeling rectifier starts to support voltage and drain current and voltage start to fall. The rate of fall of drain voltage is now governed almost exclusively by the Miller effect, and an equilibrium condition is reached, under which the drain voltage falls at just the rate necessary for the voltage between gate and source terminals to satisfy the level of drain current estab-lished by the load. This is why the gate-to-source voltage falls as the recovery current of the freewheeling rectifier falls, then stays constant at a level corresponding to the drain current, while the drain voltage falls. Obviously, the lower the impedance of the gate-drive circuit, the higher the discharge current through the drain-gate self-capacitance, the faster will be the fall time of the drain voltage and the switching losses.

Finally, at time t_4, the HEXFET®is switched fully on, and the gate-to-source voltage rises rapidly towards the applied “open circuit” value.

Similar considerations apply to the turn-off interval. Figure 5 shows theoretical waveforms for the HEXFET®in the circuit of Figure 4 during the turn-off interval. At t_o the gate drive starts to fall until, at t_l , the gate voltage reaches a level that just sustains the drain current and the device enters the linear mode of operation. The drain-to-source voltage now starts to rise. The Miller effect governs the rate-of-rise of drain voltage and holds the gate-to-source voltage at a level corresponding to the constant drain current. Once again, the lower the impedance of the drive circuit, the greater the charging current into the drain-gate capacitance, and the faster will be the rise time of the drain voltage. At t_3 the rise of drain voltage is complete, and the gate voltage and drain current start to fall at a rate determined by the gate-source circuit impedance.

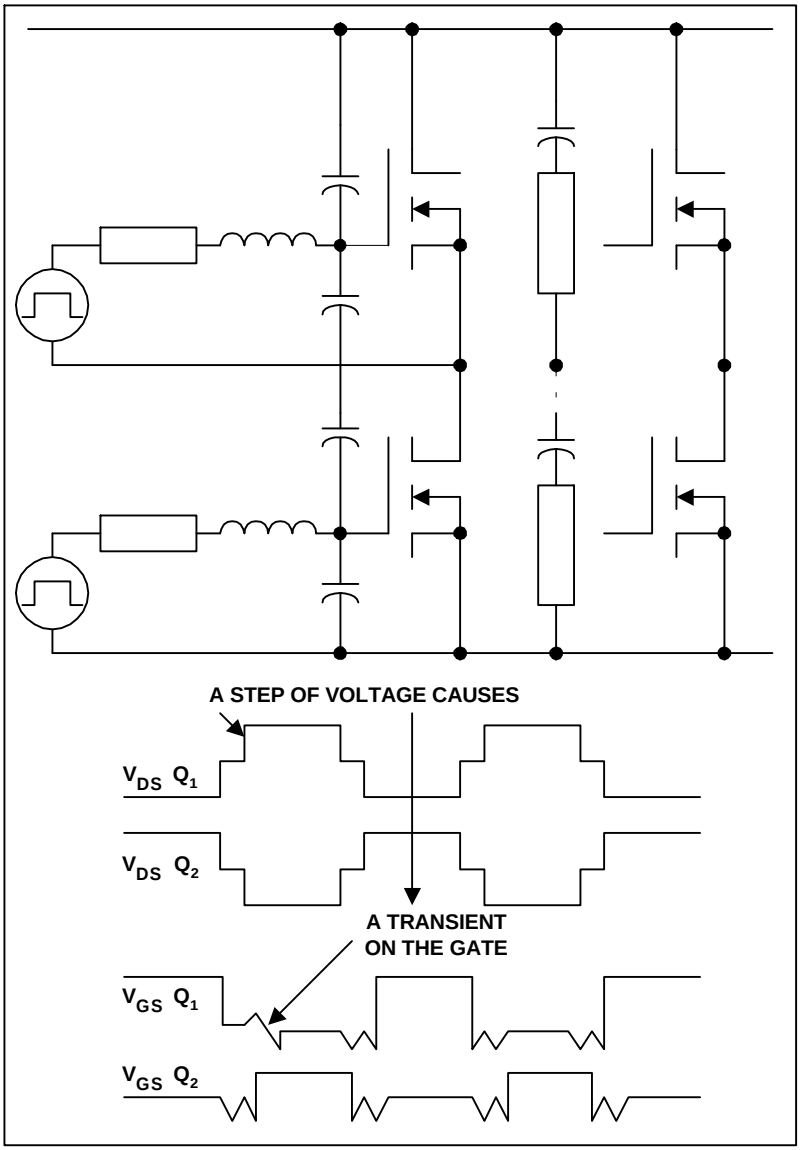


Figure 6. Transients of Voltage Induced on the Gate by Rapid Changes on the Drain-to-Source Voltage

We have seen how and why a low gate drive impedance is important to achieve high switching performance. However, even when switching performance is of no great concern, it is important to minimize the impedance in the gate drive circuit to clamp unwanted voltage transients on the gate. With reference to Figure 6, when one HEXFET®is turned on or off, a step of voltage is applied between drain and source of the other device on the same leg. This step of voltage is coupled to the gate through the gate-to-drain capacitance, and it can be large enough to turn the device on for a short instant (“dv/dt induced turn-on”). A low gate drive impedance would keep the voltage coupled to the gate below the threshold.

In summary: MOS-gated transistors should be driven from low impedance (voltage) sources, not only to reduce switching losses, but to avoid dv/dt induced turn-on and reduce the susceptibility to noise.

4. DRIVING STANDARD HEXFET®S FROM TTL

Table 1 shows the guaranteed sourcing and sinking currents for different TTL families at their respective voltages. From this table, taking as an example of the 74LS series, it is apparent that, even with a sourcing current as low as 0.4 mA, the guaranteed logic one voltage is 2.4V (2.7 for 74LS and 74S). This is lower than the possible threshold of a HEXFET®. The use of a pull-up resistor in the output, as shown in Figure 7, takes the drive voltage up to 5 V, as necessary to drive the gate of Logic Level HEXFET®s, but is not sufficient to fully enhance standard HEXFET®s. Section 8 covers the drive characteristics of the logic level devices in detail.

Logic Conditions	54 / 74	54H / 74H	(54L) / 74L	(54LS) / 74LS	74S
Logic Zero Min. sink current for V_{OL}	16mA ≤ 0.4V	20mA ≤ (0.4V) /	20mA ≤ (0.3V) / 0.4V	(4) / 8 ≤ (0.4V) / 0.5V	20mA 0.5V
Logic One Max. source current for V_{OH}	-0.4mA ≥2.4V	-0.5mA ≥2.4V	-0.2mA ≥2.4V	-0.4mA ≥ (2.5) / 2.7V	-1.0mA ≥2.7V
Typical Gate Propagation Delay	10ns	7ns	50ns	12ns	4ns

Table 1. Driving HEXFET®s from TTL (Totem Pole Outputs)

Open collector buffers, like the 7406, 7407, etc., possibly with several drivers connected in parallel as shown in Figure 9, give enough voltage to drive standard devices into "full enhancement", i.e. data sheet on-resistance. The impedance of this drive circuit, however, gives relative long switching times. Whenever better switching performance is required, interface circuits should be added to provide fast current sourcing and sinking to the gate capacitances. One simple interface circuit is the complementary source-follower stage shown in Figure 9. To drive a MOSFET with a gate charge of 60 nC in 60 ns an average gate current of 1 A has to be supplied by the gate drive circuit, as indicated in INT-944. The on-resistance of the gate drive MOSFETs has to be low enough to support the desired switching times.

With a gate charge of 60 nC and at a switching frequency is 100kHz, the power lost in the gate drive circuit is approximately:

$P = V_{GS} \times Q_G \times f = 12 \times 60 \times 10^{-9} \times 100 \times 10^{3} = 72mW$

The driver devices must be capable of supplying 1A without significant voltage drop, but hardly any power is dissipated in them.

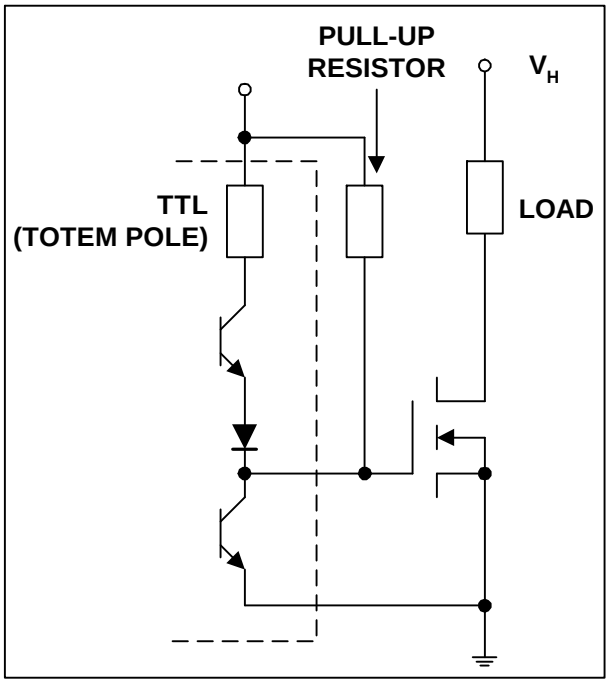


Figure 7. Direct Drive from TTL Output

5. DRIVING STANDARD HEXFET®S FROM C-MOS

While the same general considerations presented above for TTL would also apply to C-MOS, there are three substantial differences that should be kept in mind:

1. C-MOS has a more balanced source/sink characteristic that, on a first approximation, can be thought of as a 500 ohm resistance for operation over 8V and a 1k ohm for operation under 8V (Table 2).

2. C-MOS can operate from higher supply voltages than 5V so that HEXFET® saturation can be guaranteed.
3. Switching times are longer than those for TTL (Table 2).

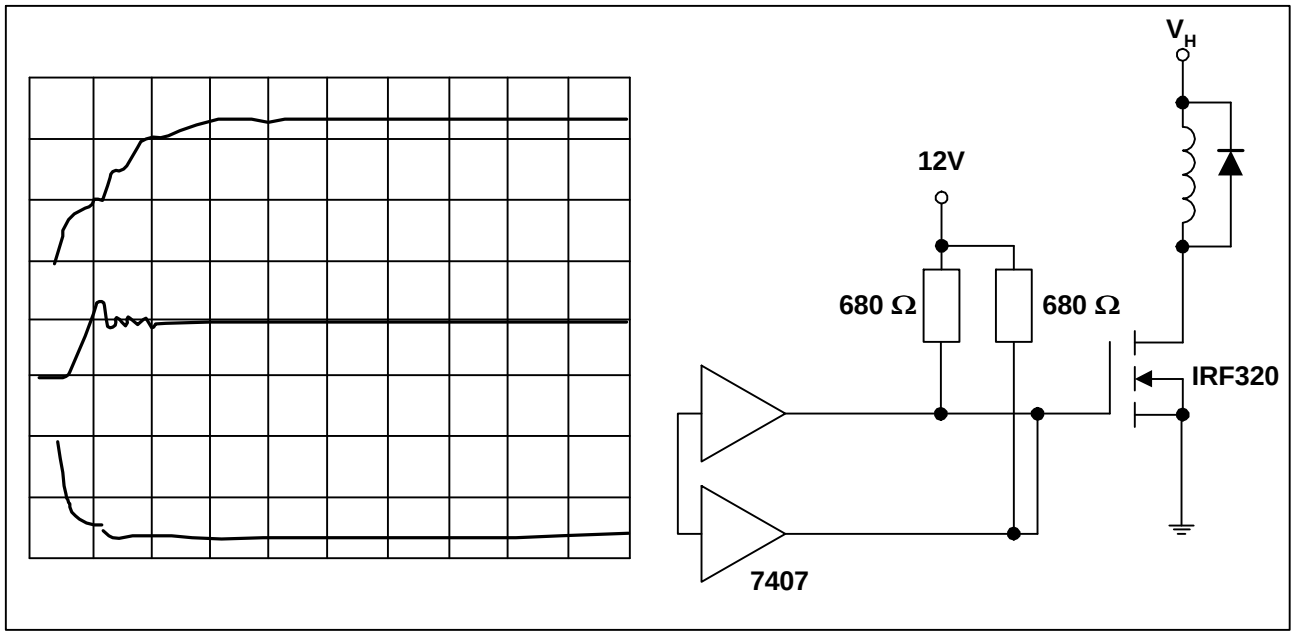


Figure 8. High Voltage TTL driver and its waveforms

When C-MOS outputs are directly coupled to the gate of a HEXFET®, the dominant limitation to performance is not the switching time, but the internal impedance (assuming that C-MOS are operated from a 10V or higher voltage supply). It will certainly not be able to turn OFF the HEXFET® as fast as the TTL, while the turn-ON waveform will be slightly better than what can be achieved with a 7407 with a 680 ohm pull-up resistor. Of course, gates can be paralleled in any number to lower the impedance and this makes C-MOS a very simple and convenient means of driving HEXFET®s. Drivers can also be used, like the 4049 and 4050 which have a much higher current sinking capability (Table 2), but they do not yield any significant improvement in current sourcing.

For better switching speeds, buffer circuits, like the one shown in Figure 9, should be considered, not only to provide better current sourcing and sinking capability, but also to improve over the switching times of the C-MOS output itself and the dv/dt noise immunity.

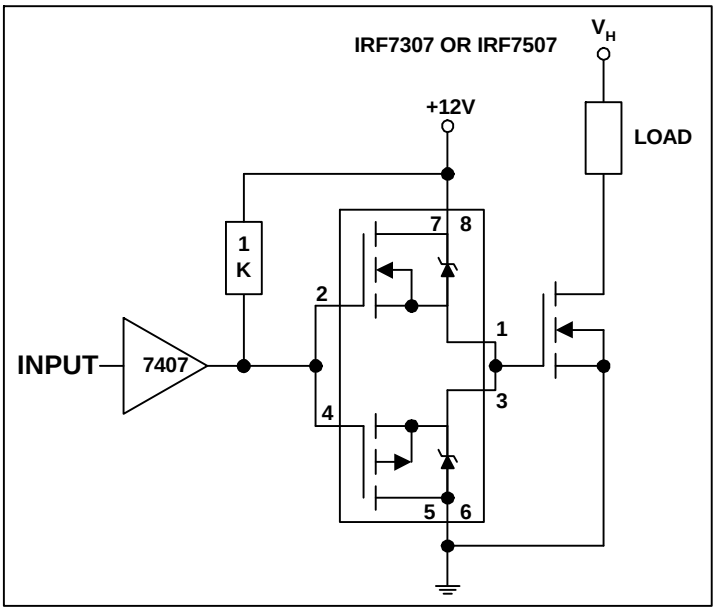


Figure 9. Simple Interface to Drive HEXFETs from TTL

6. DRIVING HEXFET®S FROM LINEAR CIRCUITS

The complementary source follower configuration of Figure 9 can also be used in linear applications to improve drive capability from an opamp or other analog source.

Most operational amplifiers have a very limited slew rate, in the order of few V/microsec. This would limit the bandwidth to less than 25kHz. A larger bandwidth can be obtained with better operational amplifiers followed by a current booster, like the ones shown in Figures 10 or 11. For a system bandwidth of 1MHz, the opamp bandwidth must be significantly higher than 1MHz and its slew rate at least 30V/μs.

Logic Supply Voltage / Logic Conditions	Standard Buffered Outputs			4049 / 4050 Drivers		
	5V	10V	15V	5V	10V	15V
Logic Zero: Approximate sink current for $V_{OL} \leq 1.5V$	1.5mA	3.5mA	4mA	20mA	40mA	40mA
Logic One: Minimum source current for V_{OH}	-0.5mA $\geq$ 4.6V	-13mA $\geq$ 9.5V	-3.4mA $\geq$ 13.5V	-1.25mA $\geq$ 2.5V	-1.25mA $\geq$ 9.5V	-3.75mA $\geq$ 13.5V
Typical switching times of logic drive signals:						
RISE	100ns	50ns	40ns	100ns	50ns	40ns
FALL	100ns	50ns	40ns	40ns	20ns	15ns

Table 2. Driving HEXFET®s from C-MOS (Buffered)

When analog signals determine the switching frequency or duty cycle of a HEXFET®, as in PWM applications, a voltage comparator is normally used to command the switching. Here, too, the limiting factors are the slew rate of the comparator and its current drive capability. Response times under 40ns can be obtained at the price of low output voltage swing (TTL compatible). Once again, the use of output buffers like the ones shown in Figures 9, may be necessary to improve drive capability and dv/dt immunity. If better switching speeds are desired. a fast op-amp should be used.

In many applications, when the HEXFET®is turned on, current transfers from a freewheeling diode into the HEXFET®. If the switching speed is high and the stray inductances in the diode path are small, this transfer can occur in such a short time as to cause a reverse recovery current in the diode high enough to short out the dc bus. For this reason, it may be necessary to slow down the turn-on of the HEXFET®while leaving the turn-off as fast as practical. Low impedance pulse shaping circuits can be used for this purpose, like the ones in Figures 12 and 13.

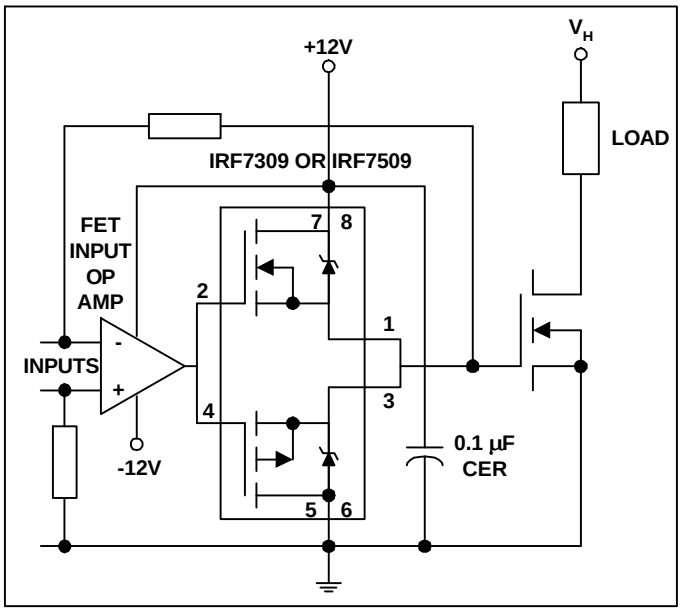


Figure 10. Dual Supply Op-Amp Drive Circuit

7. DRIVE CIRCUITS NOT REFERENCED TO GROUND

To drive a HEXFET®into saturation, an appropriate voltage must be applied between the gate and source. If the load is connected between source and ground, and the drive voltage is applied between gate and ground, the effective voltage between gate and source decreases as the device turns on. An equilibrium point is reached in which the amount of current flowing in the load is such that the voltage between gate and source maintains that amount of drain current and no more. Under these conditions the voltage drop across the MOSFET is certainly higher than the threshold voltage and the power dissipation can be very high. For this reason, the gate drive circuit is normally referenced to the source rather than to the ground. There are

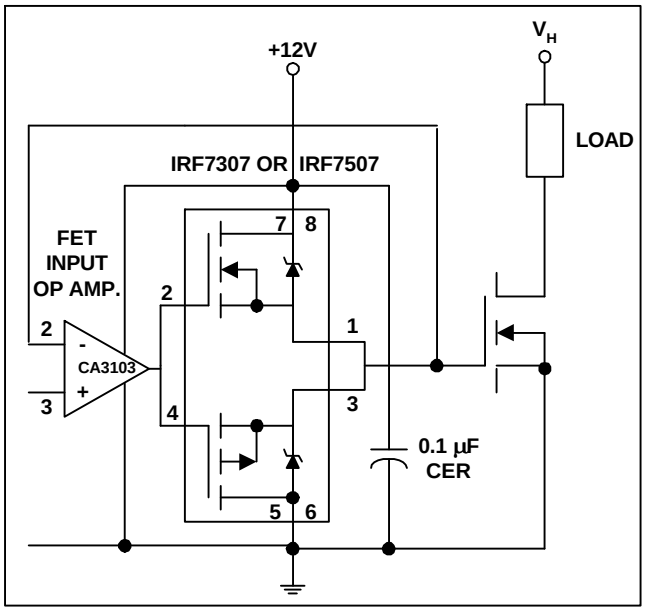


Figure 11. Single Supply Op-Amp Drive Circuit (Voltage Follower)

basically three ways of developing a gate drive signal that is referenced to a floating point:

1. By means of optically coupled isolators.
2. By means of pulse transformers.

By means of DC to DC chopper circuits with transformer isolation.

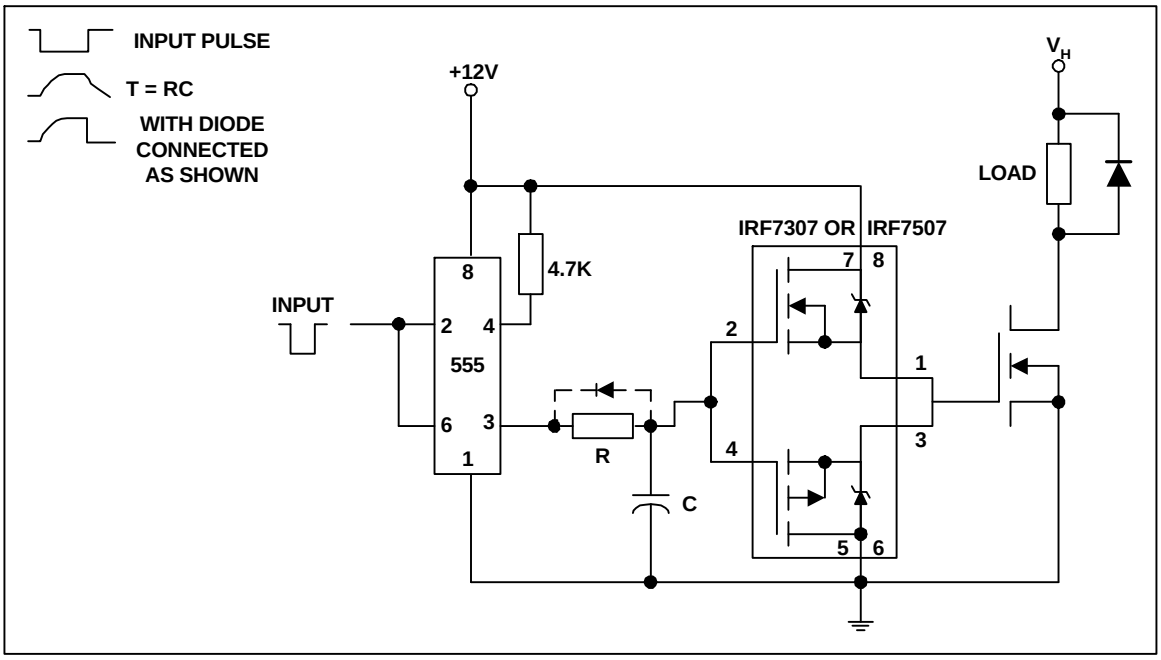


Figure 12. A pulse shaper. The 555 is used as an illustration of a Schmitt Trigger pulse shaper

7.1 MGDs with optocouplers

Most optocouplers require a separate supply grounded to the source on the receiving end of the optical link and a booster stage at the output, as shown in Figure 14a. One of the major difficulties encountered in the use of optocouplers is their susceptibility to noise. This is of particular relevance in applications where high currents are being switched rapidly. Because of the dv/dt seen by the V_{EE} pin, the optocoupler needs to be rated for high dv/dt, in the order of 10 V/ns.

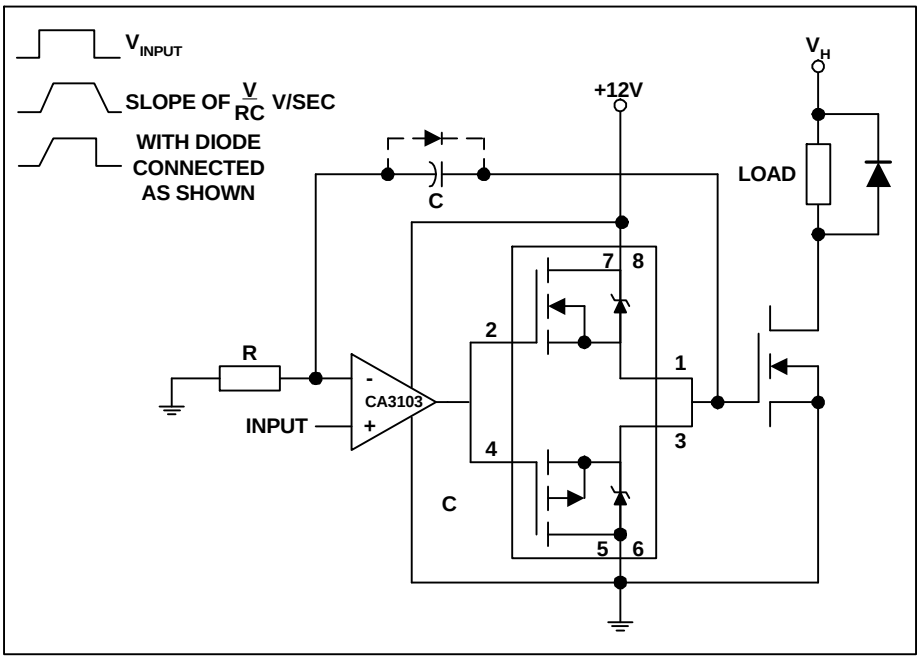


Figure 13. Pulse shaper implemented with an integrator

Figure 15a shows an MGD with under-voltage lockout and negative gate bias. When powered with a 19 V floating source, the gate drive voltage swings between +15V and -3.9V. D1 and R2 offset the emitter voltage by 3.9V. The switching waveforms shown in Figure 15b are similar to those in Figure 14b except for the negative bias. Q3, D2 and R5 form the under voltage lockout circuit.

The LED D2 is used as low voltage, low current reference diode. Q3 turns on when the voltage at the anode of D2 exceeds the sum of the forward voltage of LED and the base-emitter voltage of Q3. This enables the operation of the optocoupler. The tripping point of the under voltage lock-out circuit is 17.5V. The start-up wave forms are shown in Figure 16.

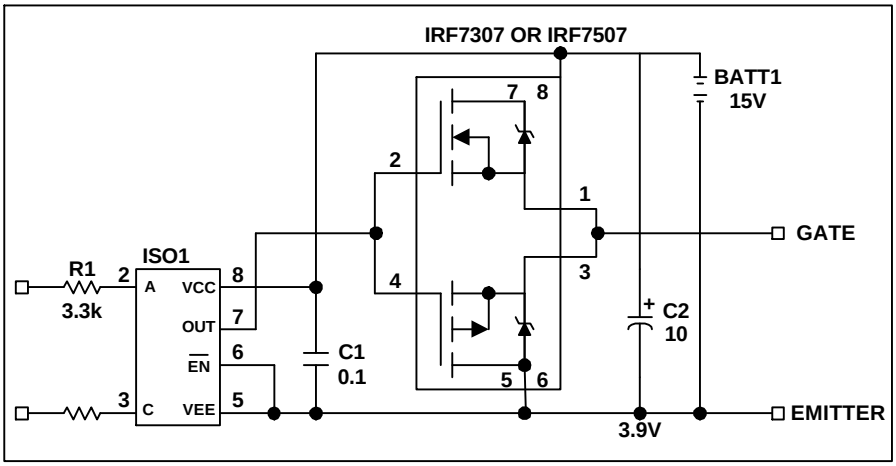


Figure 14a. Simple high current optoisolated driver

The auxiliary supply for the optocoupler and its associated circuitry can be developed from the drain voltage of the MOSFET itself, as shown in Figure 17, 18 and 19. This supply can be used in conjunction with the UV-lockout shown in Figure 15 to provide a simple high-quality optoisolated drive.

The circuit in Figure 17a can be modified to provide higher output current. By changing C1 to 680pF and R3 to 5.6k, its performance changes to what is shown in Figures 20, 21 and 22. Other methods of developing isolated supplies are discussed in Section 9.

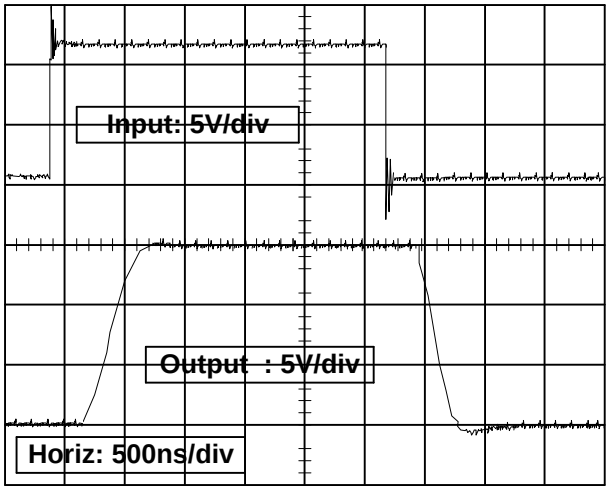


Figure 14b: Waveforms associated with the circuit of Figure 14a when loaded with 100nF

7.2 Pulse transformers

A pulse transformer is, in principle, a simple, reliable and highly noise-immune method of providing isolated gate drive. Unfortunately it has many limitations that must be overcome with additional components. A transformer can only transfer to the secondary the AC component of the input signal. Consequently, their output voltage swings from negative to positive by an amount that changes with the duty cycle, as shown in Figure 23. As a stand-alone component they can be used for duty cycles between 35 and 65%.

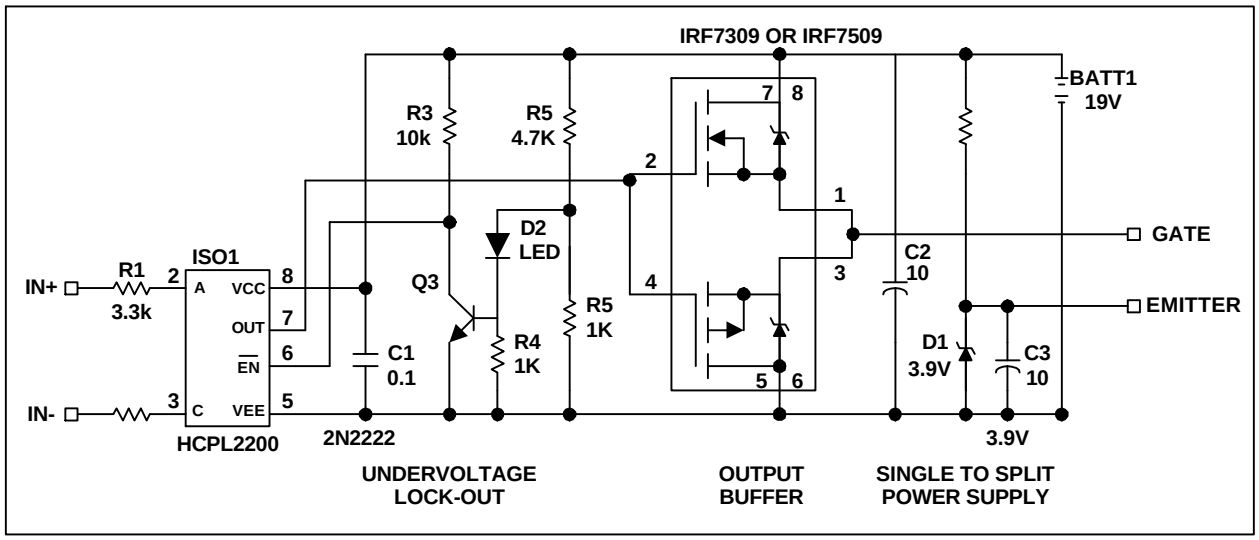


Figure 15a: Optoisolated driver with UV lockout and negative gate bias

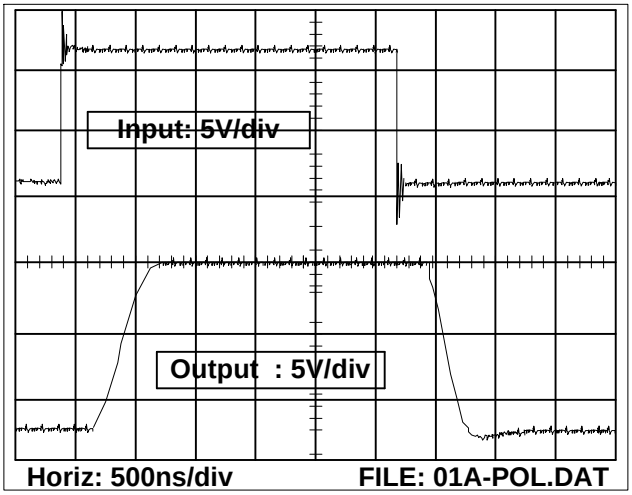


Figure 15b: Waveforms of the circuit in Figure 21a when loaded with 100nF

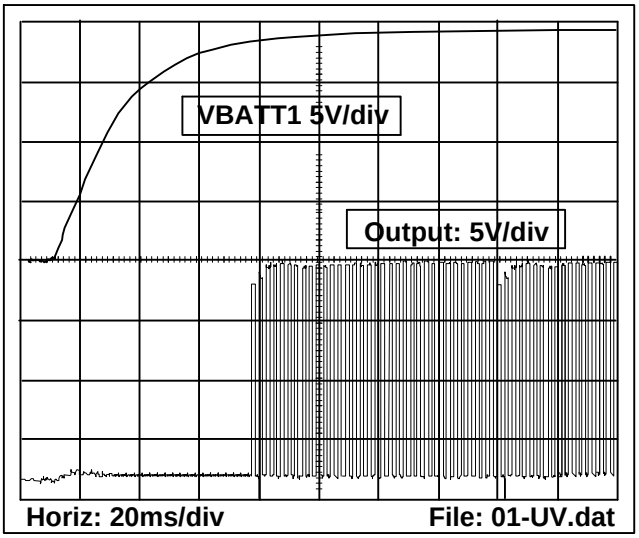


Figure 16. Start-up waveforms for the circuit of Figure 15a.

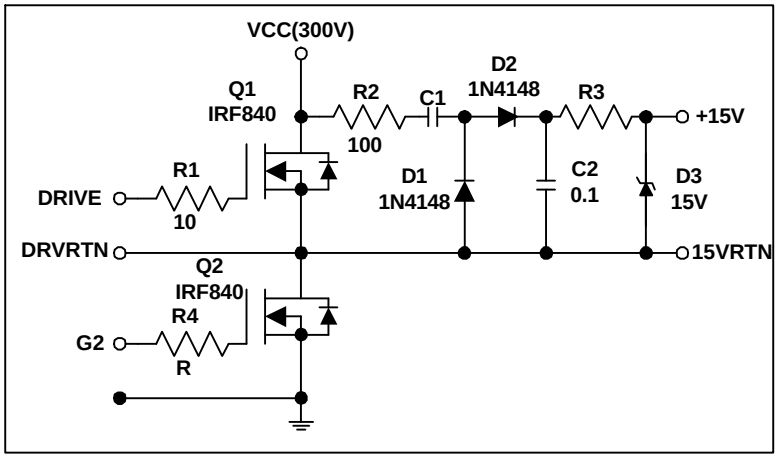


Figure 17a. Drive supply developed from the drain voltage

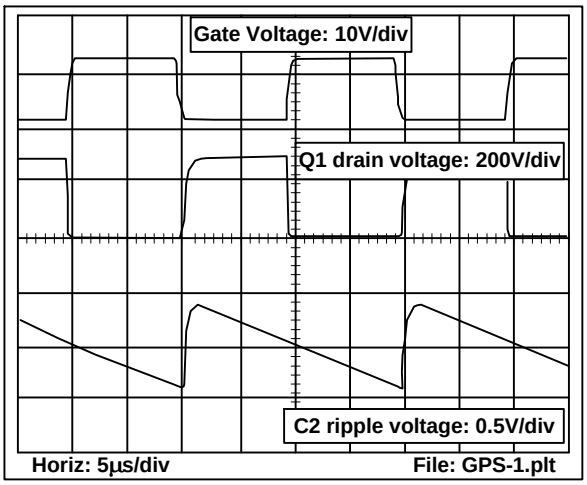


Figure 17b. Waveforms of the circuit in Figure 23a.C1 = 100 pF, R3 = 5.6 k, f = 50 kHz

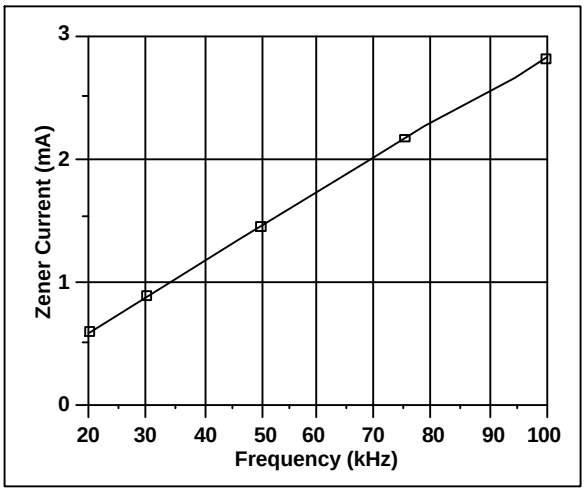


Figure 18. Zener current (max output current) for the circuit in Figure 23a.

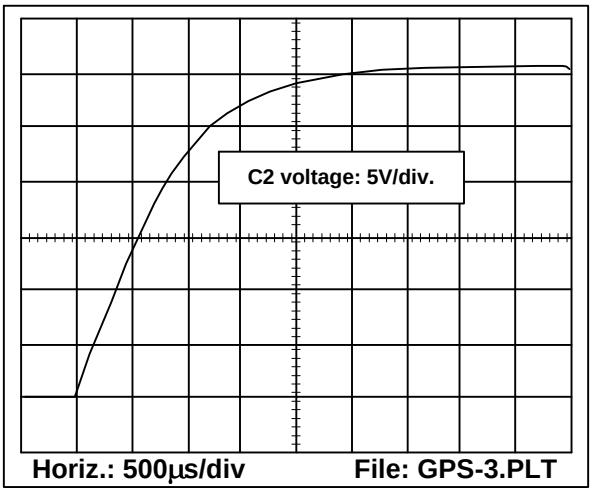


Figure 19. Start-up voltage at 50 kHz for the circuit in Figure 23a.

They have the additional advantage of providing a negative gate bias. One additional limitation of pulse transformers is the fact that the gate drive impedance is seriously degraded by the leakage inductance of the transformer. Best results are normally obtained with a few turns of twisted AWG30 wire-wrap wire on a small ferrite core.

Lower gate drive impedance and a wider duty-cycle range can be obtained with the circuit in Figure 24a. In this circuit, Q1 and Q2 (a single Micro-8 package) are used to buffer the input and drive the primary of the transformer. The complementary MOS output stage insures low output impedance and performs wave shaping. The output stage is fed by a dc restorer made by C2 and D1 that references the signal to the positive rail. D1 and D2 are also used to generate the gate drive voltage.

The input and output wave form with 1nF load capacitance are shown in Figure 24b. The turn-on and turn-off delays are 50ns. The rise and fall times are determined by the 10 Ohm resistor and the capacitive load. This circuit will operate reliably between 20 and 500 kHz, with on/off times from 0.5 to 15 microsecs.

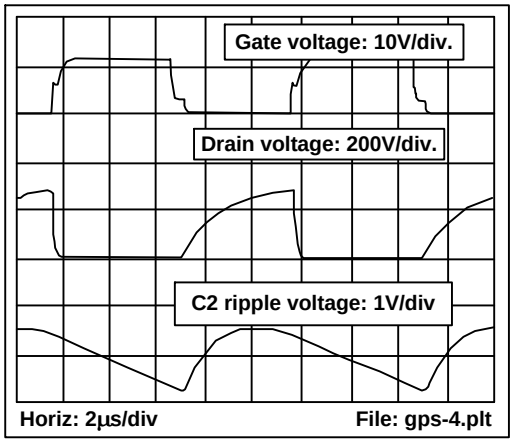


Figure 20. Waveforms of the circuit in Figure 23a. with C1=680pF, R3=1k, f=100kHz.

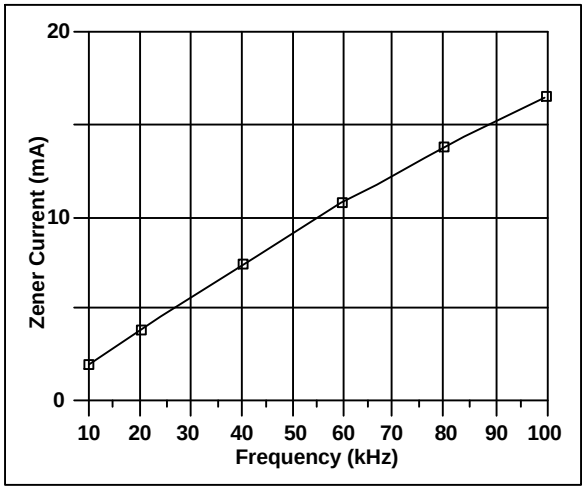


Figure 21. Zener current (max output current) for the circuit in Figure 23a. with C1 = 680pF, R3 = 1k

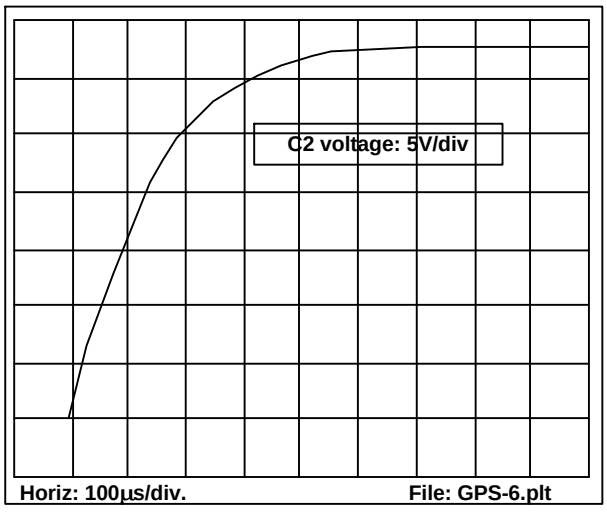


Figure 22. Start-up voltage at 100 kHz for the circuit in Figure 23a. with C1=680pF, R3=R3=1k

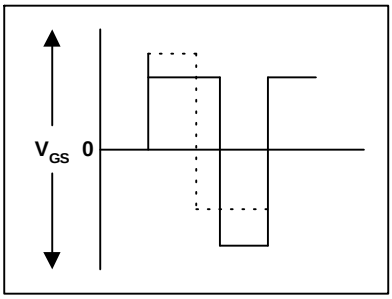


Figure 23. Volt-seconds across winding must balance

Due to the lack of an under voltage lock-out feature, the power-up and power down behavior of the circuit is important. Intentionally C1 and C2 are much bigger in value then C3 so that the voltage across C3 rises to an adequate level during the first incoming pulse. The power-up wave forms at 50kHz switching frequency and 50% duty cycle are shown in Figure 25. During the first pulse, the output voltage is 10V only, and drops back below 10V at the fifth pulse.

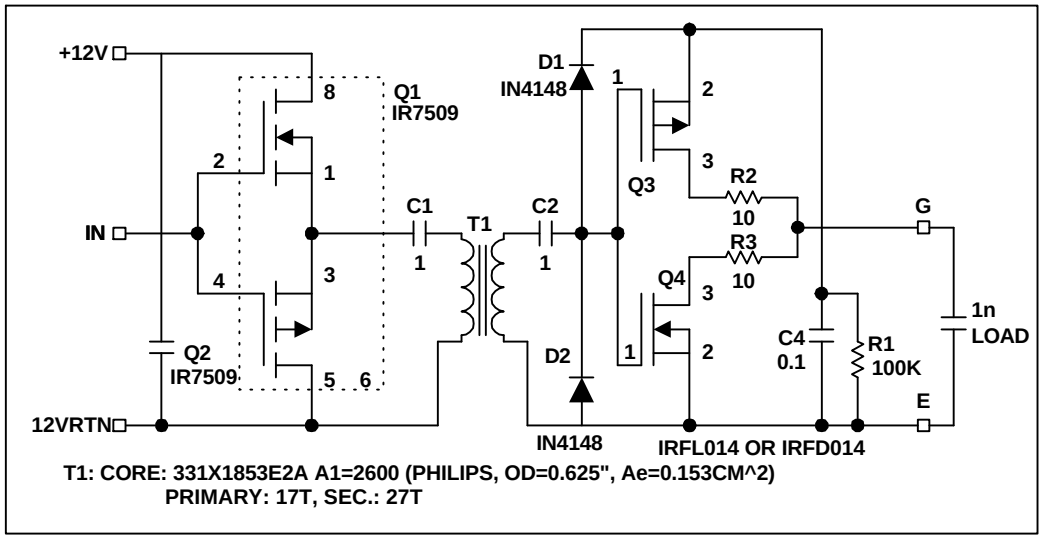


Figure 24a. Improving the performance of a gate drive transformer

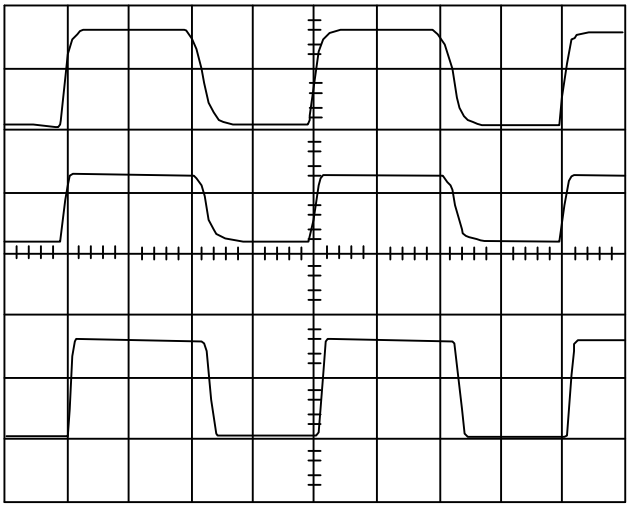


Figure 24b. Waveforms associated with the circuit of Figure 24a

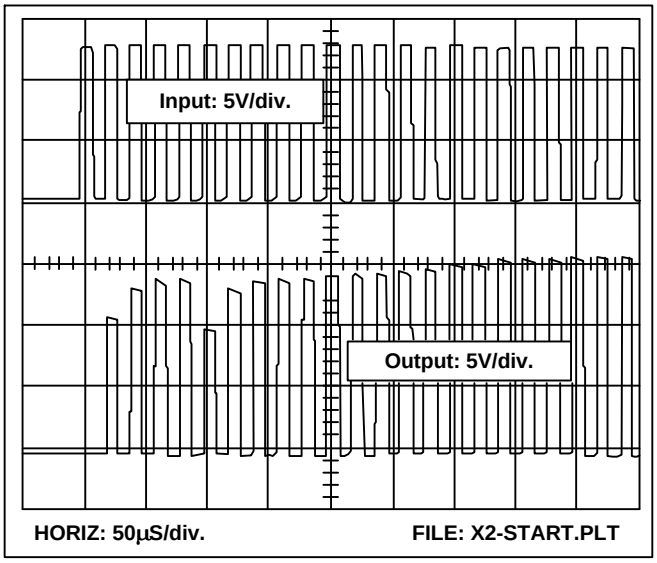


Figure 25. Waveforms during start-up for the circuit in Figure 24a.

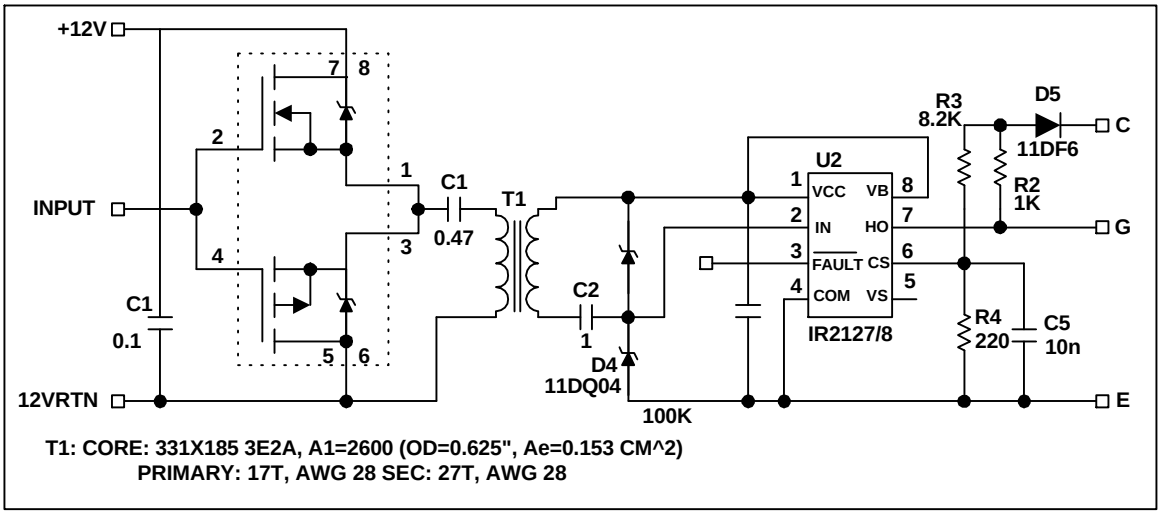


Figure 26a. Transformer-coupled MGD with UV lockout and short-circuit protection

The power down of the circuit is smooth and free from voltage spikes. When the pulse train is interrupted at the input, the C2 capacitor keeps the input of the CMOS inverter high and R1 discharges C3. By the time the input to the CMOS inverter drops below the threshold voltage of Q4, C3 is completely discharged the output remains low.

The addition of a MOS-Gate Driver IC improves the performance of the circuit in Figure 24a, at the expenses of prop delay. The circuit shown in Figure 26a has the following features:

- No secondary supply required
- Propagation delay ~500ns (CL= 10nF)
- Duty cycle range 5% to 85%
- Nominal operating frequency 50kHz (20kHz to 100kHz)
- Short circuit protection with Vce sensing. Threshold Vce = 7.5V
- Undervoltage lock-out at Vcc = 9.5V
- Over voltage lock out at Vcc = 20V

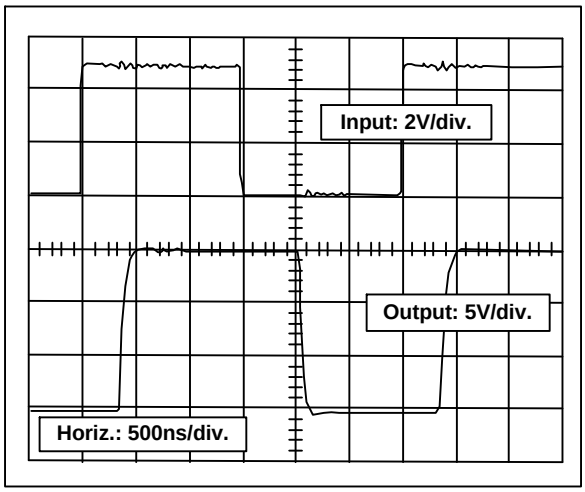


Figure 26b. Waveforms associated with the circuit of Figure 26a.

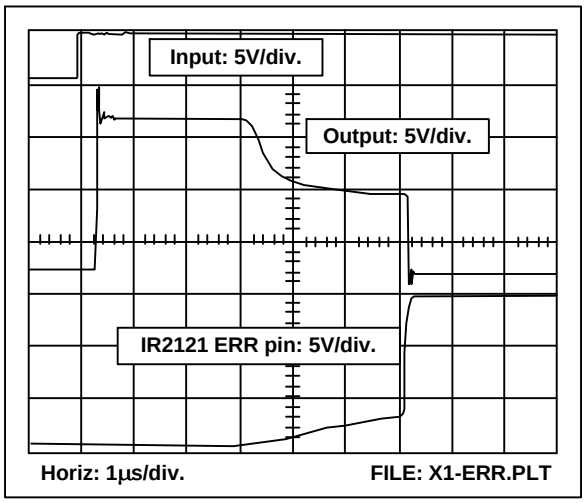


Figure 27. Shutdown due to high V_{CEsat}

The short circuit protection is implemented with a Vce sensing circuit in combination with the current sense input (CS) of IR2127/8. When the HO pin if U2 goes high R3 starts charging C5. Meanwhile the IGBT turns on, the collector voltage drops to the saturation level, D5 goes into conduction and C5 discharges. When the collector voltage is high, D5 is reverse biased and the voltage on C5 keeps raising. When C5 voltage exceeds 250mV the IR2127/8 shuts down the output. The fault to shut-down delay is approximately 2 microsecs.

For operation with a large duty cycle, several options are available. The circuits described in AN-950 use a saturating transformer to transfer the drive charge to the gate. The circuit shown in Figure 28a, on the other hand, achieves operation over a wide range of duty cycles by using the MGD as a latch. It has the following features:

- Frequency range from DC to 900kHz.
- Turn-on delay: 250ns.
- Turn-off delay 200ns
- Duty cycle range from 1% to 99% at 100kHz.
- Under voltage and over voltage lockout.
- Optional short circuit protection, as shown in Figure 26a

In the circuit of Figure 28a the transformer is small (8 turns), since it transmits only short pulses to the secondary side. The MGD on the secondary side of the transformer is latched by the feedback resistor R4. Figures 28b and 28c show the performance of this circuit at the two extremes of 900 kHz and 2.5 Hz

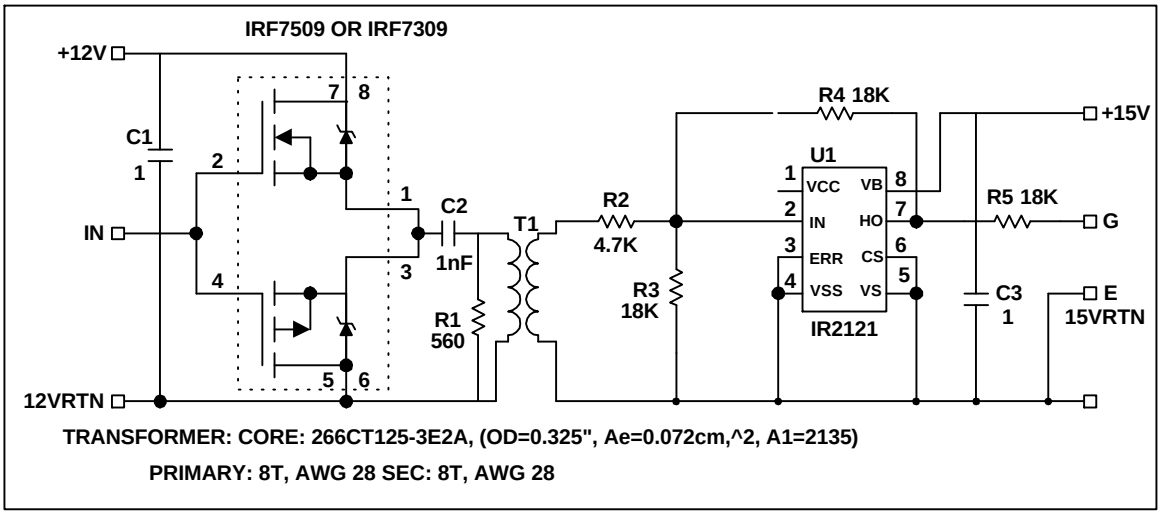


Figure 28a. Transformer-coupled MGD for operation from DC to 900 kHz

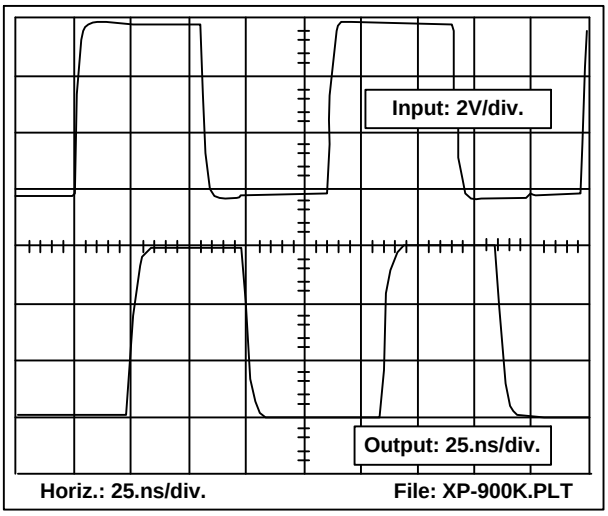


Figure 28b. Waveforms associated with the circuit of Figure 28a operated at 900 kHz

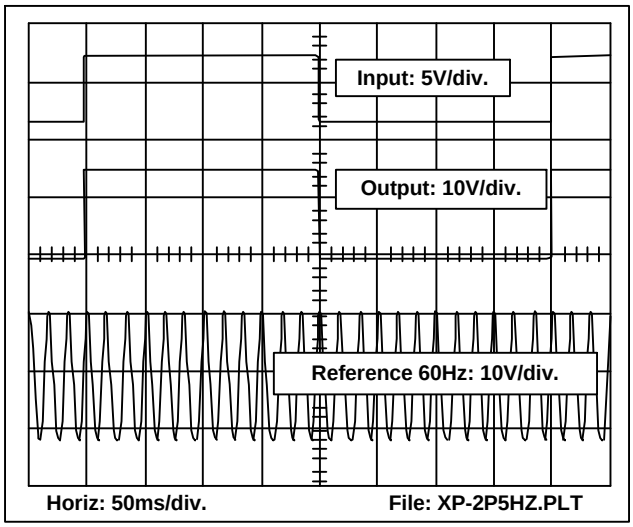


Figure 28c. Waveforms associated with the circuit of Figure 28a operated at 2.5 Hz

7.2 Chopping gate drives

Chopper circuits can maintain a gate drive signal for an indefinite period of time, have good noise immunity performance and, with some additional circuitry, the isolated supply can be avoided.

The basic operating principle is shown in Figure 29. To turn on the MOSFET, a burst of high frequency is transmitted to the secondary side. The MOSFET is turned off by interrupting the high frequency. The diode and the bipolar transistor form a crowbar that rapidly discharges the gate.

In addition to providing the gate drive signal, the high frequency transformer is frequently used to power auxiliary circuitry, like short-circuit protection, thus avoiding a dedicated supply.

8. DRIVE REQUIREMENTS AND SWITCHING CHARACTERISTICS OF LOGIC LEVEL HEXFET®S

Many applications require a power MOSFET to be driven directly from 5 V logic circuitry. The on-resistance of standard power MOSFETs is specified at 10 V gate drive, and are generally not suitable for direct interfacing to 5V logic unless an oversized MOSFET is employed.

Logic level HEXFET®s are specifically designed for operation from 5V logic and have guaranteed on-resistance at 5 or 4.5 V gate voltage. Some have guaranteed on-resistance at 2.7 V.

Some important considerations for driving logic level HEXFET®s are discussed in this section and typical switching performance of these is illustrated when driven by some common logic drive circuits.

8.1 Comparison to Standard HEXFET®s

Some devices are available as Logic-level HEXFET®s as well as standard HEXFET®s. The logic-level version uses a thinner gate oxide and different doping concentrations. This has the following effects on the input characteristics:

- Gate Threshold voltage is lower.
- Transconductance is higher.
- Input capacitance is higher.
- Gate-source breakdown voltage is lower.

While input characteristics are different, reverse transfer capacitance, on-resistance, drain-source breakdown voltage, avalanche energy rating, and output capacitance are all essentially the same. Table 3 summarizes the essential comparisons between standard and logic level HEXFET®s.

Characteristics and Ratings		Standard HEXFET® (IRF Series)	Comparable Logic Level HEXFET® (IRL Series)
Gate Threshold Voltage	$V_{GS(on)}$	2 - 4V	1 - 2V
On-Resistance	$R_{DS(on)}$	Logic level HEXFET® has same value of $R_{DS(on)}$ V_{GS} = 5V as standard HEXFET® at V_{GS} = 10V $R_{DS(on)}$ of logic level HEXFET® also speed at V_{GS} = 4V	
Transconductance	gfs	Typically 39% larger for logic level HEXFET®	
Input Capacitance	Crs	Typically 33% larger for logic level HEXFET®	
Output Capacitance	Crss	Essentially the same	
Reverse Transfer Capacitance	Crss	Essentially the same	
Gate Charge Gate-Source	Qgs	Essentially the same	
Gate-Drain	Qgd	Essentially the same	
Total	Qg	Essentially same as V_{GS} = 10V	Essentially same at V_{GS} = 5V
Drain Source Breakdown Voltage	BV_{DSS}	Same	
Continuous Drain Current	I_D	Same	
Single Pulse Avalanche Energy	E_{AS}	Same	
Max. Gate-Source Voltage	V_{GS}	± 20V	±10V

Table 3: Essential Comparisons of Standard and Logic Level HEXFET®s

The gate charge for full enhancement of the logic level HEXFET®is, however, about the same as for a standard HEXFET®because the higher input capacitance is counteracted by lower threshold voltage and higher transconductance. Since the logic level HEXFET®needs only one half the gate voltage, the drive energy is only about one half of that needed for the standard HEXFET®. Since the gate voltage is halved, the gate drive resistance needed to deliver the gate charge in a given time is also halved, relative to a standard HEXFET®. In other words, for the same switching speed as a standard HEXFET®power MOSFET, the drive circuit impedance for the logic level HEXFET®must be approximately halved.

The equivalence of switching times at one half the gate resistance for the logic level HEXFET®is illustrated by the typical switching times for the IRL540 and the IRF540 HEXFET®s shown in Table 4, using data sheet test conditions.

Gate Resistance	Gate Voltage	Drain Current	Typical Values (ns)			
R_G (Ω)	V_{GS} (V)	I_D (A)	$t_{D\ on}$	t_r	$t_{D\ on}$	t_r
9	10	28	15	72	40	50
4.5	5	28	15	72	44	56

Table 4: Typical Resistive Switching Times for IRL540 and IRF540

TTL families do not actually deliver 5V in their V_{OH} condition, even into an open circuit. The 5V level can, however, be reached by the addition of a pull-up resistor from the output pin to the 5V bus, as illustrated in Figure 30. Without the pull-up resistor, the $R_{DS(on)}$ value at V_{GS} = 5V may not be attained, and the value specified at V_{GS} = 4V should be used for worst case design.

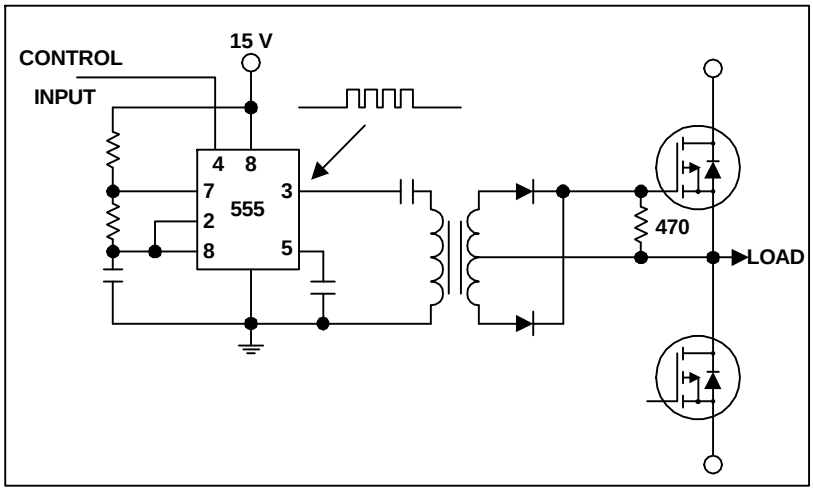


Figure 29.

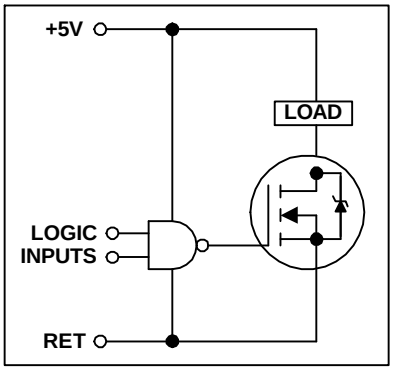


Figure 30. Pull-up resistor used to deliver 5V gate drive

8.2 Driving Logic Level HEXFET®s

The gate threshold voltage of MOSFETs decreases with temperature. At high temperature it can approach the $V_{OL(max)}$ specification of the logic driver. Care should be exercised to insure that $V_{TH(min)}$ at the highest operating temperature is greater than $V_{OL(max)}$ of the various logic families in order to guarantee complete turn off.

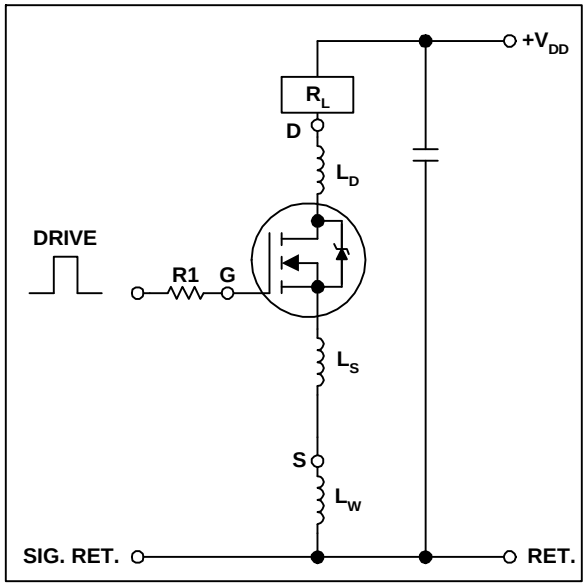


Figure 31a. High common mode inductance

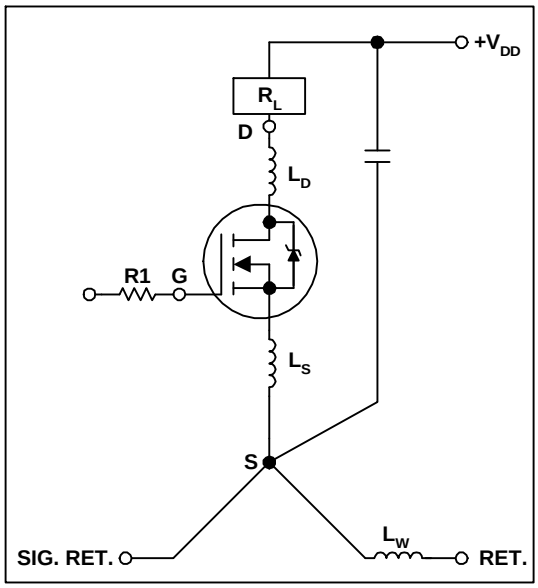


Figure 31b. Minimum common mode inductance

Common source inductance plays a significant role in switching performance. In the circuit of Figure 31a the switching performance is degraded due to the fact that V_{GS} is reduced by (L_S + L_W) di/dt, where di/dt is the rate of change of the drain current. By eliminating L_W from the drive circuit, V_{GS} can approach the applied drive voltage because only L_S (the internal source inductance) is common.

This can be done by separately connecting the power return and the drive signal return to the source pin of the switching HEXFET®, as shown in Figure 31b. Thus, the load current I_D does not flow through any of the external wiring of the drive circuit; consequently, only the internal source inductance L_S is common to both load and drive circuits.

In the case of logic level HEXFET®s, for which V_{GS} is 5V and not 10V, the loss of drive voltage due to common mode inductance has proportionately twice the effect as it would on a 10V drive signal, even though actual values of L_S and L_W are the same.

8.3 Resistive Switching Tests

In the following tests of switching performance, the physical layout of the test circuit was carefully executed so to minimize the common source inductance. The following precautions were also observed:

1. R_L was built by paralleling 0.5W resistors to achieve the desired load resistance (see Table 5).
2. To minimize inductance in the load circuit, a 10 μF low-ESR low-ESL capacitor was connected directly from $+V_{DD}$ to the source of the DUT.
3. To provide a low source impedance for the 5V gate pulse of the DUT, a 0.1 μF low-ESR low-ESL capacitor was connected directly between pin 14 and pin 7 of the driver IC.
4. To provide minimum common source impedance, the source of the DUT was the common return point of all ac and dc system grounds.
5. To reduce stray inductances and thus achieve maximum switching speeds, the physical size of the high current loop (R_L, DUT, 10 μF) was reduced to the smallest practical limits.

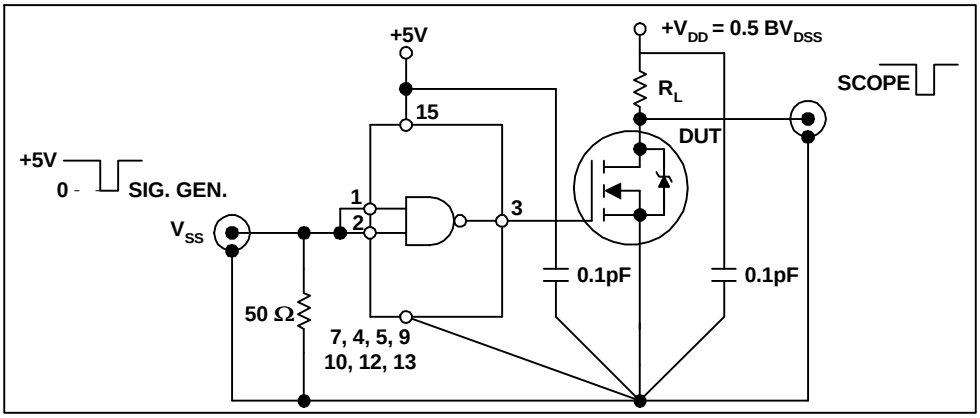


Figure 32. Switching test circuit. Logic level driver is one-quarter of a quad NAND gate.

Only the 5 volt families have been tested as logic level HEXFET®drives: bipolar and CMOS (and their derivatives), as indicated below.

TTL GATES

DM7400N:	Standard TTL
74F00PC:	High Speed TTL
DM74S00N:	Schottky TTL
DM74LS00N:	Low Power Schottky TTL
DM74AS00N:	Advanced Schottky TTL

CMOS GATES

74AC00PC: Advanced CMOS
74ACT00PC: TTL Compatible CMOS
MM74HC00N: Micro CMOS
MM74HCT00N: TTL Compatible Micro CMOS

BIPOLAR
DS0026: High Speed MOSFET Driver

The test conditions for the resistive switching performance is shown in Table 5. The resistive switching times obtained with the above TTL and CMOS gates are tabulated in Table 6. In this table t_{on} = Time in microseconds from 90% to 10% V_{DD} and t_{off} = Time in microseconds from 10% to 90% V_{DD}. Inductive switching gives faster voltage rise times than resistive switching due to the resonant charging of the output capacitance of the device. Voltage fall times are essentially the same.

LOGIC LEVEL HEXFET®	SWITCHING VOLTAGE (V)	SWITCHING CURRENT (A)	R_{DSON} (Ω)	R_L (Ω)
IRLZ14	30	8	0.24	3.25
IRLZ24	30	16	0.12	1.5
IRLZ34	30	24	0.06	1.2
IRLZ44	30	40	0.034	0.7
IRLZ514	50	5	0.60	9.5
IRLZ524	50	8	0.30	5.9
IRLZ524	50	12	0.18	4.0
IRLZ544	50	25	0.085	1.9

Table 5. Resistive Switching Conditions

Logic Family	Logic Level HEXFET®,															
Quad, Dual Input	IRLZ14		IRLZ24		IRLZ34		IRLZ44		IRL514		IRL524		IRL534		IRL544	
Nand Gate	t_{on}	t_{off}	t_{on}	t_{off}	t_{on}	t_{off}	t_{on}	t_{off}	t_{on}	t_{off}	t_{on}	t_{off}	t_{on}	t_{off}	t_{on}	t_{off}
DM7400N STANDARD TTL	0.173	0.018	0.663	0.026	0.700	0.076	1.491	0.146	0.151	0.022	0.238	0.041	0.263	0.060	0.616	0.124
7400FDOPC HIGH SPEED TTL	0.124	0.008	0.490	0.013	0.429	0.068	0.863	0.146	0.104	0.004	0.159	0.034	0.176	0.059	0.372	0.136
DM7400 SCHOTTKY TTL	0.133	0.092	0.549	0.020	0.503	0.032	1.068	0.142	0.116	0.006	0.183	0.041	0.212	0.057	0.441	0.132
DM74LS LOW POWER SCHOTTKY TTL	0.174	0.038	0.778	0.093	0.706	0.146	1.438	0.342	0.155	0.040	0.240	0.062	0.267	0.090	0.567	0.199
DM4SDON ADVANCED SCHOTTKY TTL	0.126	0.008	0.567	0.013	0.446	0.023	0.896	0.149	0.111	0.005	0.161	0.127	0.176	0.058	0.336	0.130
74ACOOPC ADVANCED CMOS	0.012	0.007	0.120	0.012	0.125	0.027	0.251	0.139	0.036	0.004	0.052	0.028	0.066	0.055	0.125	0.125
74ACTOOPC TTL COMPATIBLE CMOS	0.012	0.006	0.121	0.011	0.125	0.016	0.233	0.127	0.033	0.044	0.052	0.027	0.060	0.055	0.120	0.122
MM74CHCOON MICRO CMOS	0.066	0.039	0.179	0.091	0.227	0.147	0.508	0.328	0.058	0.044	0.092	0.068	0.111	0.096	0.232	0.213
MM74HCTCO4 TTL COMPATIBLE MICRO CMOS	0.066	0.030	0.179	0.060	0.227	0.123	0.504	0.269	0.068	0.035	0.092	0.051	0.111	0.086	0.232	0.186
DS0026 HIGH SPEED MOSFET DRIVER	0.052	0.005	0.016	0.005	0.014	0.007	0.032	0.016	0.021	0.004	0.036	0.004	0.036	0.005	0.029	0.009

Table 6. Results of the resistive load switching test

Typical Test Oscillograms

IRLZ24: 60V, 0.1 Ohm, N-Channel, TO-220 logic level HEXFET®was driven by each of the logic families listed in Table 4 and the comparative resistive switching times photographed.

9. SIMPLE AND INEXPENSIVE METHODS TO GENERATE ISOLATED GATE DRIVE SUPPLIES

.

In several applications, dc-to-dc converters are used to power the MOS Gate Driver. Although the gate drive requires little power, the noisy environment, the isolation voltage and creepage distance requirements and the high dv/dt between the primary and secondary size make the design of the DC-to-DC converter somewhat complicated. Its key parameters are listed below: OUTPUT VOLTAGE, CURRENT. The output voltage of the DC-to-DC converter is the sum of the positive and negative drive voltage to the gate. The load current required from the DC-to-DC converter is the sum of the current consumption of the drive circuit and the average drive current to the gate.

dv/dt CAPABILITY. When the DC-DC converter powers a high side switch, the secondary side of the converter is connected to the output of the power circuit. The rapid change of high voltage at the output of power circuit stresses the isolation of the transformer and injects noise to the primary side of the transformer. Switching noise at the primary side disturbs the operation of the converter and the control circuit for the power stage, causing false triggering and shoot-through. Therefore a transformer with high voltage isolation, appropriate creepage distances and low winding-to-winding capacitance is required in this application.

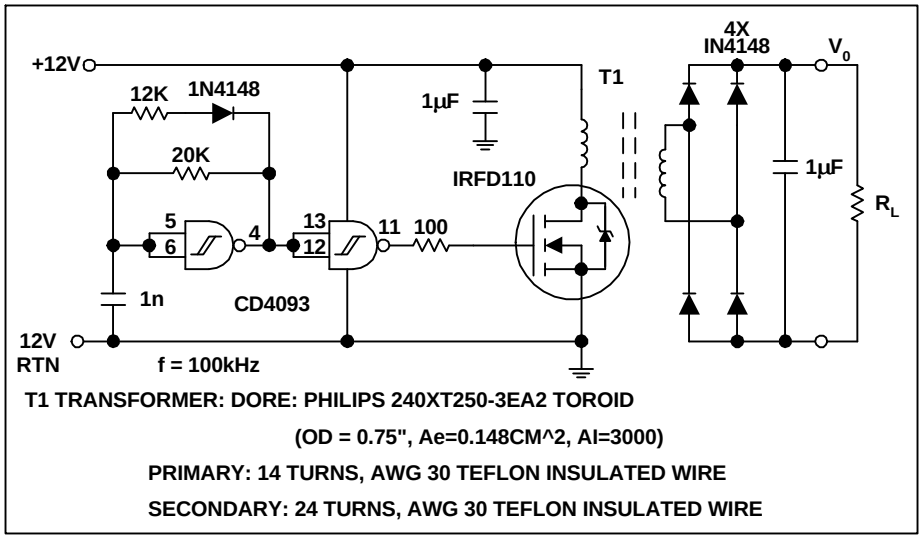


Figure 33a. 100 kHz Forward converter

SMALL SIZE. To reduce the interwinding capacitances the transformer must be made small. This implies operation at high frequency. Small size and compact layout help reducing the EMI and RFI generated by the converter. Figure 33a shows a forward converter made with two CD4093 gates to generate the clock and drive the MOSFET. Energy as transferred to the secondary when the MOSFET is on, in about 33% of the cycle. When the MOSFET is off, the secondary winding is used to demagnetize the transformer and transfer the magnetizing energy to the load, thus eliminating the need for a demagnetizing winding. The switching waveforms are shown in Figure 33b. The ringing in the drain voltage during the fly-back period is due to the loose coupling between the primary and the secondary windings. The load current vs. output voltage characteristic of the circuit is shown in Figure 34. When the output current falls below 5 mA, the circuit works as flyback converter because the demagnetizing current flows through the output. A minimum load of 5mA is required to limit the output voltage at 15V.

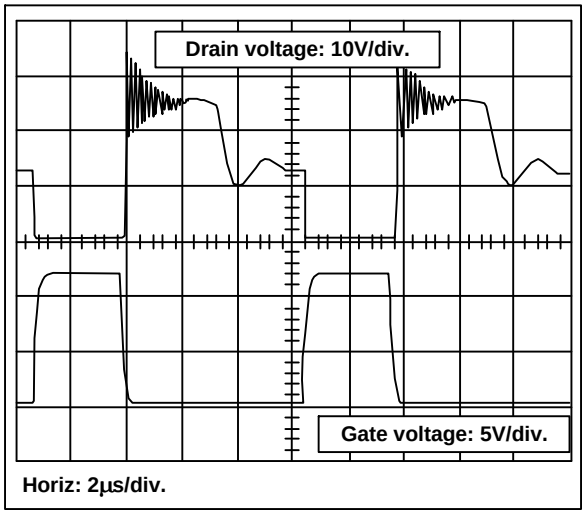


Figure 33b. Waveforms associated with the circuit in Figure 33a

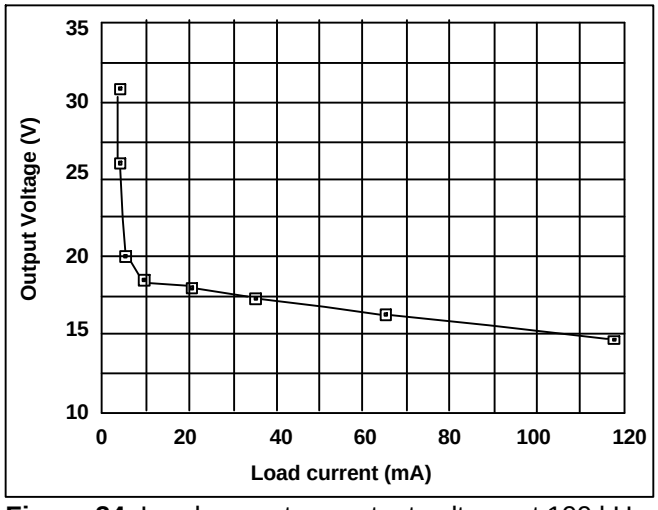


Figure 34. Load current vs. output voltage at 100 kHz, Rout = 27.7 Ohms

If the converter is loaded with a constant and predictable load, a zener can provide the necessary regulation. Otherwise a three-terminal regulator or a small zener-driven MOSFET may be necessary.

The circuit in Figure 35a is similar to the previous one, except that the higher switching frequency is higher (500 kHz) and the transformer is smaller. The remaining three gates in the package are connected in parallel to drive the MOSFET and reduce the switching losses. The switching waveforms are shown in Figure 35b. The output resistance (Rout) of this circuit is higher than the circuit shown in Figure 33a, mainly because the stray inductance of the smaller transformer is higher and the effects of the stray inductance are higher. Figure 37a shows a push-pull operated at 500 kHz. The single gate oscillator produces a 50% duty cycle output, while the remaining gates in the package are used to drive the push-pull output stage. The primary of the transformer sees half the voltage compared to the previous circuit, therefore the number of turns at the primary were reduced to half.

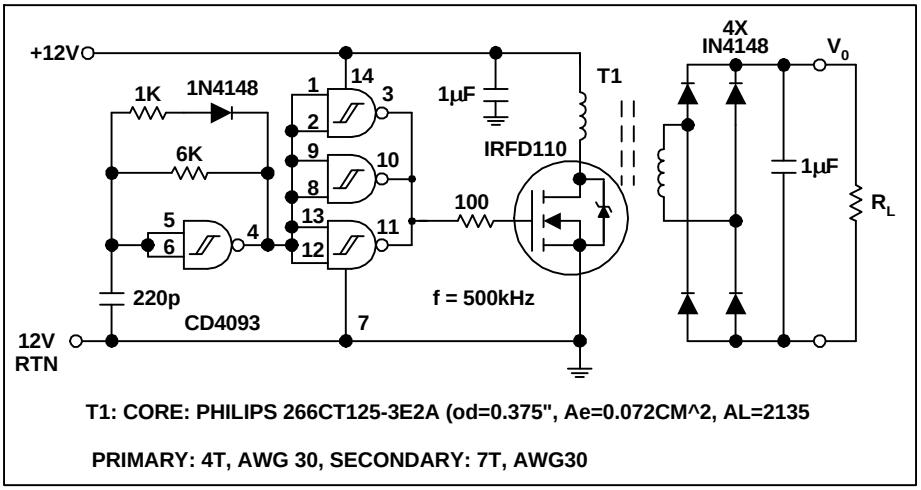


Figure 35a. 500 kHz Forward converter

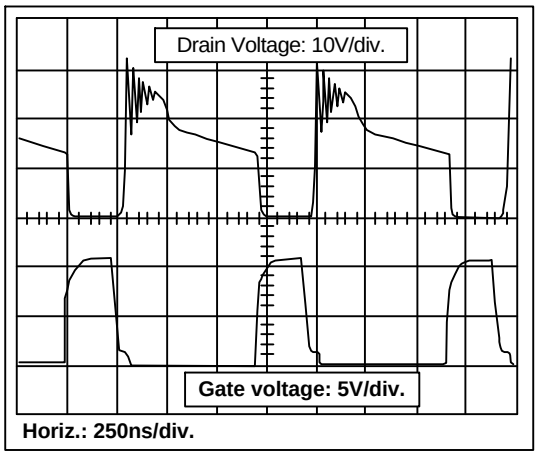


Figure 35b. Waveforms associated with the circuit in Figure 35a

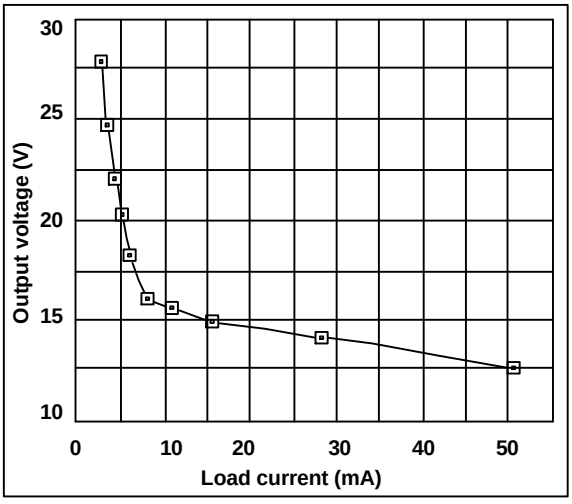


Figure 36. Load current vs. output voltage, Rout = 27.7 Ohms

10. PHOTOVOLTAIC GENERATORS AS GATE DRIVERS

A photovoltaic generator is a solid state power supply powered by light, normally an LED. The combination of the LED and the photovoltaic generator in one package is called a Photovoltaic Isolator or PVI and is available in a 8-pin DIP package. As a voltage source, the PVI can function as a "dc transformer" by providing an isolated low current to a load. While an optoisolator requires a bias supply to transmit a signal across a galvanic barrier, the PVI actually transmits the energy across the barrier. More information on the PVI can be found in Application Note GBAN-PVI-1 which appears in the Microelectronic Relay Designer' s Manual. This data book also contains the data sheet for the photovoltaic isolator, the PVI1050. A circuit is also provided in the AN to significantly speed up turn off of the switch. As a gate driver the PVI has significant limitations: its short circuit current is in the order of 30 microA with a very high internal impedance. Its simplicity, however, makes it appealing in solid-state relay replacements, where switching times are not important and switching transients are not present.

A typical application is the ac switch described below. The IGBT and the power MOSFET are not suited to switching AC waveforms directly. The IGBT can only conduct current in one direction while the power MOSFET has an anti-parallel diode that will conduct during every negative half-cycle. Bidirectional blocking capability can be achieved by connecting two power MOSFETs source to source, or two IGBTs with anti-parallel diodes emitter to emitter, as shown in Figure 39.

In the case of the MOSFET, there is the possibility that, for low current levels, the current flows through both MOSFET channels, instead that one MOSFETs and diode, thereby achieving lower overall voltage drop. The MOSFET channel is a bidirectional switch, that is, it can conduct current in the reverse direction.

If the voltage across the MOSFET channel is less than the VF of the intrinsic diode (which typically has a higher VF than discrete diodes), then the majority of the current will flow through the MOSFET channel instead of the intrinsic diode. The gate drive for both the MOSFETs and IGBTs must be referenced to the common sources or emitters of the devices. Since this node will be swinging with the AC waveform, an isolated drive is necessary. The PVI can be used, as shown in Figure 40.

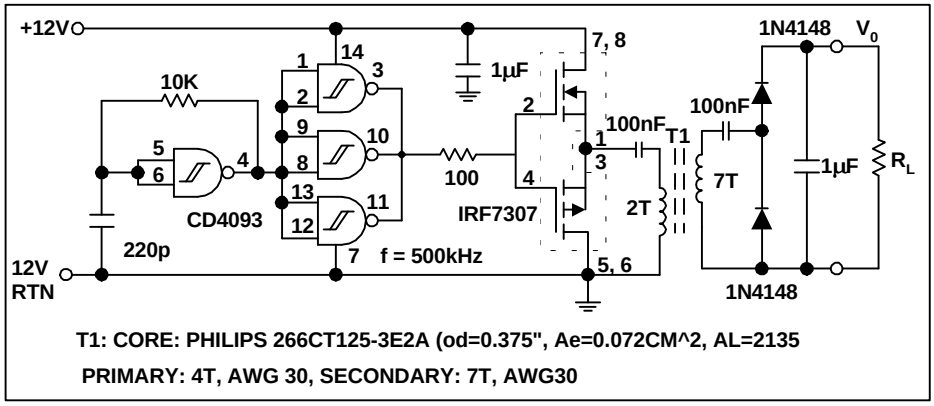


Figure 37a. 500 kHz Forward converter

11. RESONANT GATE DRIVE TECHNIQUES

As indicated in Section 14, gate drive losses in hard switching are equal to Qgs x Vgs x f. An IRF630 operated at 10 Mhz with a gate voltage of 12 V would have *gate drive losses* of 3.6 W, independent from the value of the gate drive resistor. Clearly, to achieve hard switching at this frequency, the resistance of the gate drive circuit is limited to whatever is associated with the internal impedance of the driver and with the gate structure of the device itself. Furthermore, the stray inductance of the gate drive circuit must be limited to tens of nH. The design and layout of such a circuit is not an easy task.

An alternative method to drive the gate in such an application is to design a resonant circuit that makes use of the gate capacitance and stray inductance as its reactive components, adding whatever inductance is necessary to achieve resonance at the desired frequency. This method can reduce the peak of the gate drive current and losses in half, while simplifying the design of the gate drive circuit itself. Since the gate charge is not dissipated at every switching transition, but stored in a reactive component, the gate drive losses are proportional to the resistance of the gate drive circuit, rather than being independent from it. More information on this gate drive method can be found in an article by El-Hamamsy: Design of High-Efficiency RF Class-D Power Amplifier and in references at the end of this article (IEEE Transactions on Power Electronics, May 1994, page 297).

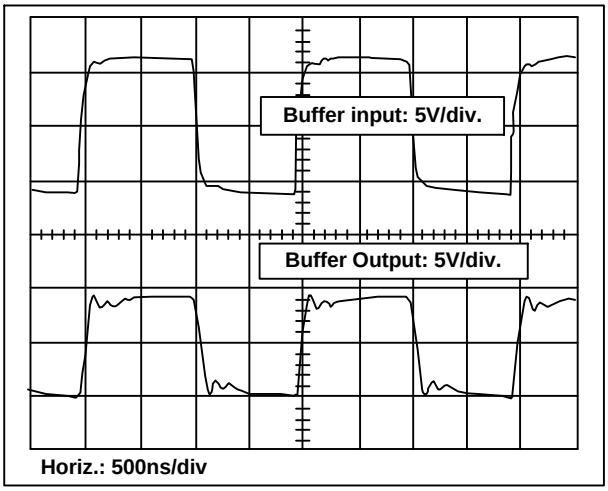


Figure 37b. Waveforms associated with the circuit in Figure 37a

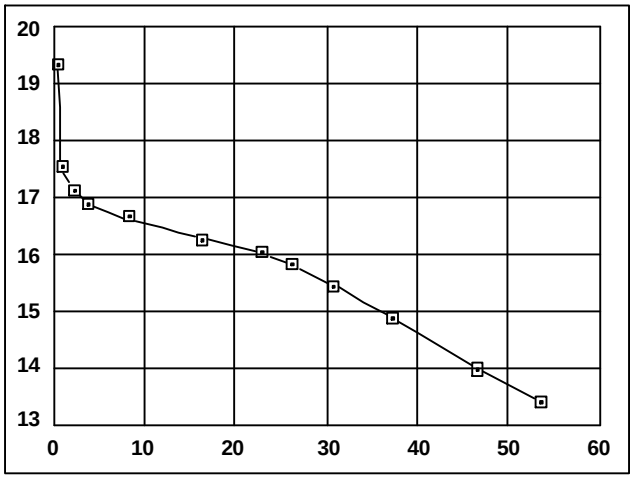


Figure 38. Load current vs. output voltage, Rout=27.7 Ohms

Related Topics

MOS-Gate Driver Ics
Transformer drive with wide duty cycle capability
Gate Charge
Three-phase MOS-Gate Driver
Photovoltaic Isolators (PVI)

International
IOR Rectifier

Application Note AN-941

Paralleling HEXFET® Power MOSFETs

Table of Contents

	Page
1. General Guidelines	1
2. Steady-State Sharing of Current	2
3. Dynamic Sharing of Current During Turn-On	4
4. Dynamic Sharing During Turn-Off	5

Whenever devices are operated in parallel, due consideration should be given to the sharing between devices to ensure that the individual units are operated within their limits. Items that must be considered to successfully parallel MOSFETs are: gate circuitry, layout considerations, current unbalance, and temperature unbalance between devices. Paralleling helps to reduce conduction losses and junction to case thermal resistance.

PARALLELING HEXFET POWER MOSFETs

(HEXFET Power MOSFET is the trademark for International Rectifier Power MOSFETs)

Summary:

- *General guidelines*
- *Steady state sharing*
- *Dynamic sharing at turn-on*
- *Dynamic sharing at turn-off*
- *Related topics*

Whenever devices are operated in parallel, due consideration should be given to the sharing between devices to ensure that the individual units are operated within their limits. Items that must be considered to successfully parallel MOSFETs are: gate circuitry, layout considerations, current unbalance, and temperature unbalance between devices. Paralleling helps to reduce conduction losses and junction to case thermal resistance. However, switching losses remain the same, or may even increase. If they are the dominant losses, only a thermal resistance improvement will be achieved by paralleling. Paralleling to take advantage of lower price of smaller devices should not be attempted without due consideration of the technical risks. Experimental results should be obtained at the extremes of the manufacturing tolerances.

The three most important parameters from this point of view are: voltage, current and junction temperature. Voltage unbalances will be briefly examined in a qualitative way in the next section with other general considerations. The effects of current and temperature unbalances will be analyzed in the following sections.

1. General guidelines on paralleling

Generally speaking, voltage equality is ensured by the fact that the devices are in parallel. However, under transient conditions, voltage differentials can appear across devices, due to di/dt effects in unequalized stray inductances.

The stray inductances of a typical power circuit, like the one shown in Figure 1, have different effects, depending on where they are situated. The effects of the emitter and collector inductances that are common to the paralleled pair have been analyzed in INT-936 and will be ignored here.

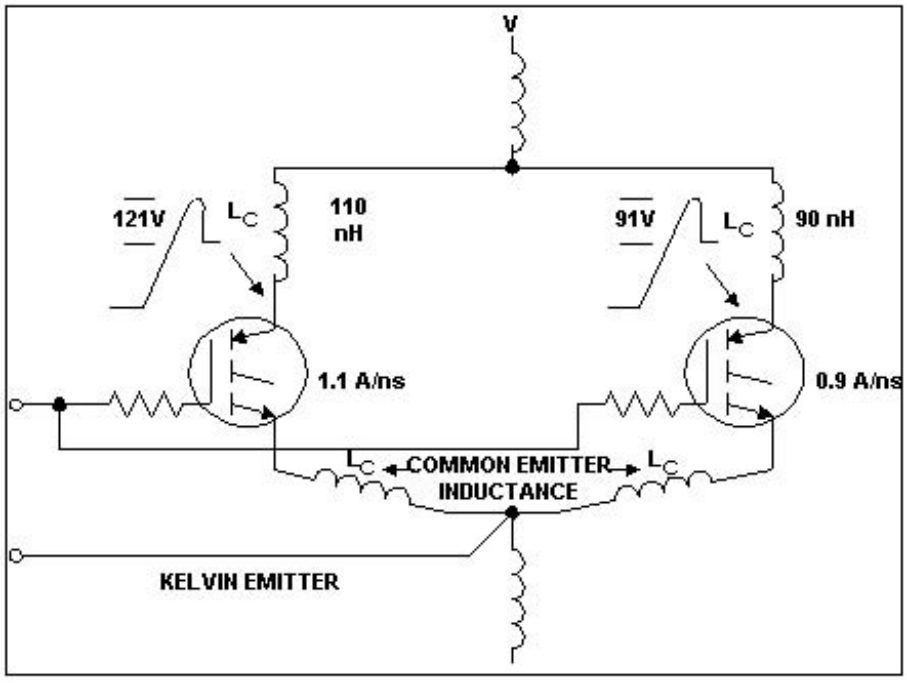


Figure 1. The effect of different di/dt and stray inductances on collector

An unbalance of 10% in the stray inductances that are in series with each collector, combined with a di/dt unbalance of 10% translates in an unbalance of 20% in the overshoot seen at turn-off (81 vs. 121V). To minimize these differentials both di/dt's and stray inductances have to be matched. However, if the overshoot does not violate the ratings of the MOSFET, the differential in the turn-off losses is negligible

The impact of the common emitter inductance on switching energy, on the other hand, is far from negligible, as explained in INT-937. Furthermore, the MOSFET with lower common source inductance turns off before the other, which is left to shoulder the entire load current during the turn-off transient. It follows that switchmode operation of paralleled IGBTs should not be undertaken unless the common emitter inductances are matched in value.

Figure 2 shows that when using paralleled devices, a low impedance path is generated that may be prone to parasitic self oscillations. This is analyzed in greater detail in Ref [1]. Individual gate resistors provide the necessary damping andgate decoupling to prevent oscillations.

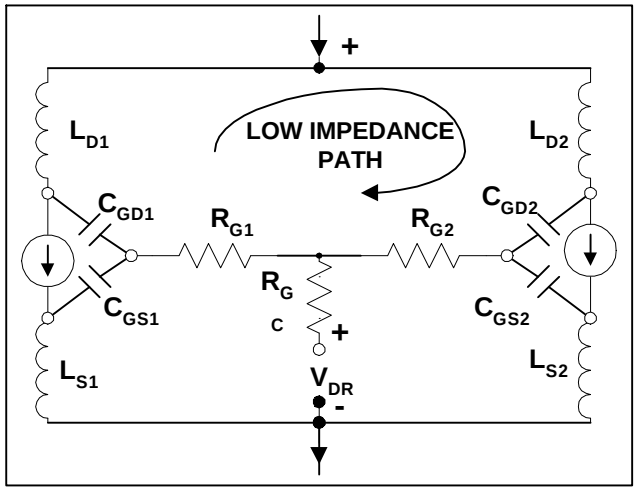


Figure 2. Low impedance Path for Parasitic Oscillation for Unbalanced Parallel Branches

In summary, the following general guidelines should be followed when paralleling MOSFETs:

* Use individual gate resistors to eliminate the risk of parasitic oscillation;

* Equalize common emitter inductance and reduce it to a value that does not greatly impact the total switching losses at the frequency of operation;

* Reduce stray inductance to values that give acceptable overshoots at the maximum operating current.

* Ensure the gate of the MOSFET is looking into a stiff (voltage) source with as little impedance as practical. This advice applies equally well to both paralleled and single device designs.

* Zener diodes in gate drive circuits may cause oscillations. Do not place them directly gate to emitter/source to control gate overvoltage, instead place them on the driver side of the gate isolation resistor(s), if required.

* Capacitors in gate drive circuits may also cause oscillations. Do not place them directly gate to emitter/source to control switching times, instead increase the gate isolation resistor. Capacitors slow down switching, thereby increasing the switching unbalance between devices.

Stray components are minimized by a tight layout and equalized by symmetrical position of components and routing of connections. INTDT 94-2 shows some preferred layout arrangements.

These guidelines ensure that the voltage and switching unbalances due to the layout are negligible with respect to those due to the MOSFETs themselves, analyzed in the next sections.

Two questions must be considered: (1) "steady-state" sharing of current, and (2) dynamic sharing of current under the transitional switching conditions.

2. Steady-State Sharing of Current

During the periods outside of the switching transitions, the current in a parallel group of HEXFET Power MOSFETs will distribute itself in the individual devices in inverse proportion to their ON resistance. The device with the lowest ON resistance will carry the highest current. This will, to an extent, be self-compensating, because the power loss in this device will be the highest. It will run hottest, and the increase in ON resistance due to heating will be more than that of the other devices, which will tend to equalize the current.

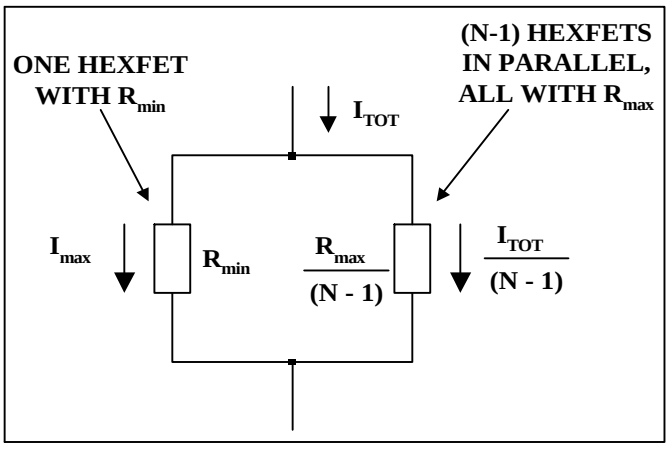


Figure 3. Simplified Equivalent Circuit for Estimating Worst Case Steady-State Current Unbalance

An analysis of the "worst case" device current in a group of "N" parallel connected devices can be based on the simplifying assumption that (N-1) devices have the highest limiting value of ON resistance, while just one lone device has the lowest limiting value of ON resistance. The analysis can then be concentrated on the current in this one device.

The equivalent electrical circuit shown in Figure 3 simplifies the analysis further by assuming the number of devices is sufficiently large that the current that flows through each of the high resistance devices is approximately $I_{TOT}/(N-1)$. On this assumption, the voltage drop across the lone low resistance device, and hence the current in it, can be calculated.

The ON resistance of each of the "high resistance" devices, at operating temperature, T, is given by:

$$\mathbf{R_{(max)T} = R_{(max)25}\left(1+\left[\left(T_A-25\right)+\frac{I^2{}_{TOT}}{(N-1)^2}R_{(max)T}\ R_{JA}\right]K\right)} \quad (1)$$

where $R_{(max)25}$ is the limiting maximum value of ON resistance at 25° C, R_{JA} is the total junction-to-ambient thermal resistance in deg. C/ W, and K is the per unit change of ON resistance per °C.

$$\therefore \quad \mathbf{R_{(max)T} = \frac{R_{(max)25}\left(1+\left[T_A-25\right]K\right)}{1-R_{(max)25}\frac{I^2{}_{TOT}}{(N-1)^2}R_{JA}K}} \quad (2)$$

The voltage drop, V, across the parallel group is: $\mathbf{V = \frac{I_{TOT}}{(N-1)} \bullet R_{(max)T}}$ (3)

The resistance of the one low resistance device at its operating temperature is:

$$\mathbf{R_{(min)T} = R_{(min)25}\left(1+\left[\left(T_A-25\right)+V\ I_{(max)}\ R_{JA}\right]K\right)} \quad (4)$$

where $R_{(min)\ 25}$ is the limiting minimum value of ON resistance at 25°C, and $I_{(max)}$ is the current in this device.

But,

$$R_{(min)T} = \frac{V}{I_{(max)}}$$

$$\therefore \quad I(max) = \frac{-b+\sqrt{(b^2+4aV)}}{2a} \quad (5)$$

where:
$b = R_{(min)\ 25}\ (1 + [T_A - 25]\ K)$
$a = R_{(min)\ 25}\ V\ R_{JA}\ K$

The following example shows the “worst case" degree of current sharing that can be expected, by applying the above relationships to the IRFP150 HEXFET Power MOSFET, and making the following assumptions:

$R_{(max)\ 25} = 0.045\Omega$
$R_{(min)\ 25} = 0.035\Omega$
$R_{JA} = 3$ deg. C/W

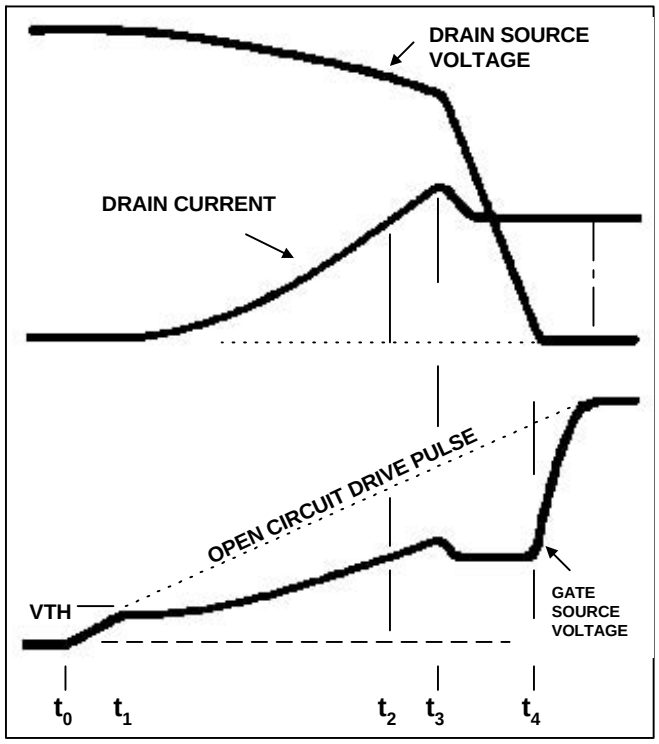
$$\mathbf{\frac{I_{TOT}}{(N-1)} = 20A} \quad (6)$$

K = 0.006 per degree C
$T_A = 35°C$

Using the relationships (2), (3), and (5) above, it can be calculated that the "worst case" maximum value of device current is 27A for the hypothetical situation where all devices but one have high limiting ON resistance, of 0.0452 and carry 20A each, whereas the remaining one has low limiting ON resistance of 0.03Ω.

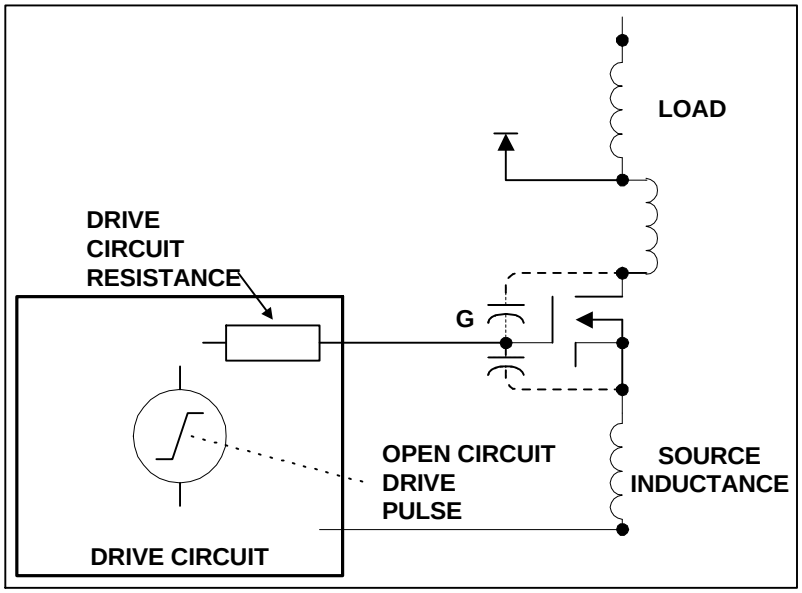
3. Dynamic Sharing of Current during turn-on

It is necessary to take positive steps to ensure that the current is distributed properly between a group of parallel connected devices during the switching transition. Since the HEXFET Power MOSFETs will not all have identical threshold and gain characteristics, some will tend to switch sooner than others, and attempt to take more than their share of the current. Adding to the problem is the fact that circuit inductance associated with each device may be different, and this will also contribute to unbalancing the current under switching conditions. Here we will provide a brief qualitative description of the different events that occur during a switching transition.



The problem will be introduced by considering the switching waveforms for a typical clamped inductive load. Figure 4 shows waveforms of drain current, drain-to-source voltage, and gate voltage during the turn-ON interval. For reasons of clarity we have shown the applied drive pulse increasing at a relatively slow rate.

At time, t_0, the drive pulse starts its rise. At t_1, it reaches the threshold voltage of the HEXFET Power MOSFET, and the drain current starts to increase. At this point, two things happen which make the gate-source voltage waveform deviate from its original “path." First, inductance in series with the source which is common to the gate circuit develops an induced voltage, as a result of the increasing source current. This voltage counteracts the applied gate drive voltage and slows down the rate-of-rise of voltage appearing directly across the gate and source terminals; this, in turn, slows down the rate-of-rise of the source current. This is a negative feedback effect; increasing current in the source produces a counteractive voltage at the gate, which tends to resist the change of current.



The second factor that influences the gate-source voltage is the so called "Miller" effect. During the period t_1 to t_2, some voltage is dropped across circuit inductance in series with the drain, and the drain-source voltage starts to fall. The decreasing drain-source voltage is reflected across the drain-gate capacitance, pulling a discharge current through it, and increasing the effective capacitance load on the drive circuit. This, in turn, increases the voltage drop across the impedance of the drive circuit and decreases the rate-of-rise of voltage appearing between the gate and source terminals. This also is a negative feedback effect; increasing current in the drain results in a fall of drain-to-source voltage, which, in turn, slows down the rise of gate-source voltage and tends to resist the increase of drain current. These effects are illustrated diagrammatically in Figure 5.

Figure 4. Waveforms at Turn-ON

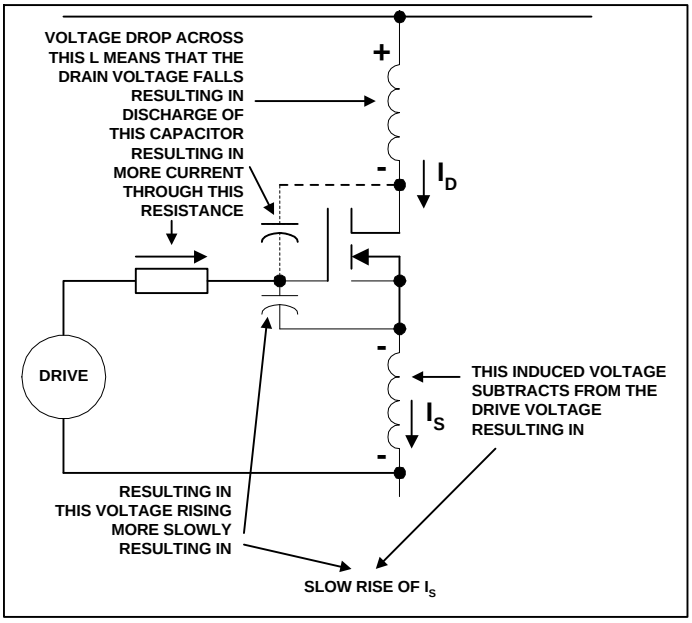


Figure 5. Diagrammatic Representation of Effects When Switching-ON

This state of affairs continues throughout the period t_1 to t_2, while the current in the HEXFET Power MOSFET rises to the level of the current, I_M, already flowing in the freewheeling rectifier, and it continues into the next period, t_2 to t_3, while the current increases further, due to the reverse recovery of the freewheeling rectifier.

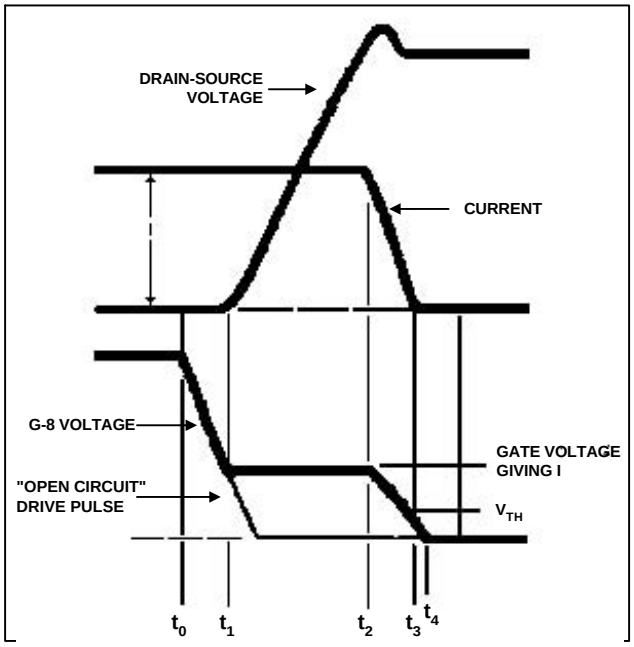
At time t_3, the freewheeling rectifier starts to support voltage, while the drain current and the drain voltage start to fall. The rate-of-fall of drain voltage is now governed by the Miller effect, and an equilibrium condition is reached, under which the drain voltage falls at just the rate necessary for the voltage between gate and source terminals to satisfy the level of drain current established by the load. This is why the gate-to-source voltage falls as the recovery current of the freewheeling rectifier falls, then stays constant at a level corresponding to the motor current, while the drain voltage is falling.

Finally, at time t_4, the HEXFET Power MOSFET is switched fully ON, and the gate-to-source voltage rises rapidly towards the applied "open circuit" value.

The above explanation, summarized in Figure 5, provides the clue to the difficulties that can be expected with parallel connected devices. The first potential difficulty is that if we apply a common drive signal to all gates in a parallel group, then the first device to turn ON—the one with the lowest threshold voltage—will tend to slow the rise of voltage on the gates of the others, and further delay the turn-ON of these devices. This will be due to the Miller effect. The inductive feedback effect, on the other hand, only influences the gate voltage of its own device (assuming that each source has its own separate inductance).
The second potential difficulty is that if the individual source inductances are unequal, then this will result in dynamic unbalance of current, even if the devices themselves are perfectly matched. Obviously, the solution to this is to ensure that inductances associated with the individual devices are as nearly equal as possible. This can be done by proper attention to the circuit layout.

As examined in detail in Ref. [1], there are several other circuit and device parameters that will contribute to dynamic unbalance. The conclusions presented in the above mentioned paper indicate, however, that the problem is not severe, as long as attention is paid to the following points, in order to ensure satisfactory sharing of current between parallel HEXFET Power MOSFETs at turn-ON:

- Threshold voltages should be within determined limits.
- Stray inductances throughout the circuit should be equalized by careful layout.
- Gates should be decoupled with individual resistors, but not more than strictly required, as it will be explained later.



4. Dynamic sharing during turn-Off

Similar considerations apply to the dynamic sharing of current during the turn-OFF interval. Figure 6 shows theoretical waveforms for the circuit of Figure 4 and 5 during the turn-OFF interval. At t_0, the gate drive starts to fall. At t_1, the gate voltage reaches a level that just sustains the drain current, I. The drain-to-source voltage now starts to rise. The Miller effect governs the rate-of-rise of drain voltage and holds the gate-to-source voltage at a level corresponding to the constant drain current. At t_3, the rise of drain voltage is complete, and the gate voltage starts to fall at a rate determined by the gate-source circuit impedance, while the drain current falls to zero.

Figure 6. Waveforms at TURN-OFF

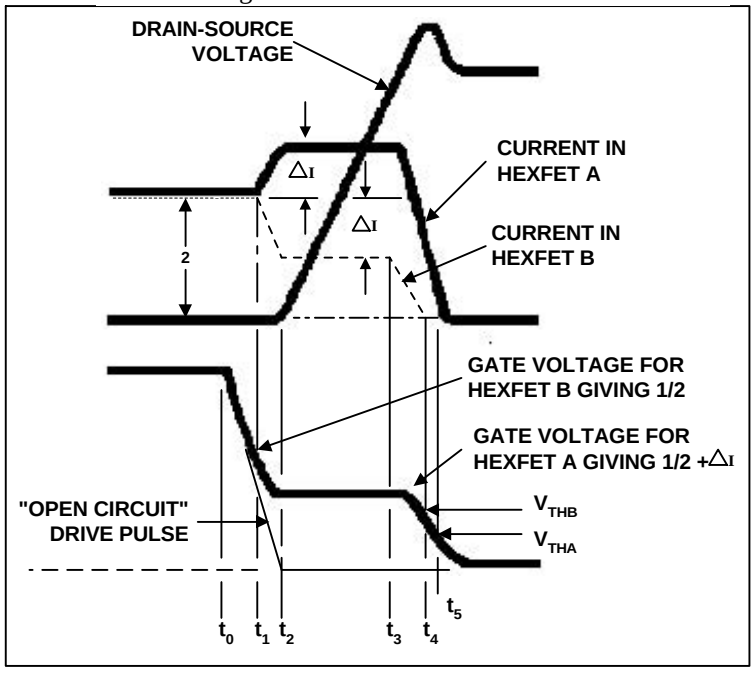


Figure 7 shows theoretical waveforms for two parallel connected HEXFET Power MOSFETs with their gates connected directly together. For purposes of discussion, the source inductance is assumed to be zero. At t_1, the gate voltage reaches the point at which HEXFET Power MOSFET B can no longer sustain its drain current.

The load current now redistributes; current in HEXFET Power MOSFET B decreases, while that in HEXFET Power MOSFET A increases. At t_2, HEXFET Power MOSFET B can no longer sustain its current; both HEXFET Power MOSFETs now operate in their 'linear" region, and the drain voltage starts to rise. The gate-to-source voltage is kept practically constant by the Miller effect, while the currents in the two HEXFET Power MOSFETs remain at their separate levels. Clearly, the unbalance of current in this example is significant.

While a turn-off unbalance is potentially a more serious problem, the analysis in Ref. [1] shows that this is not so in practice as long as the devices are turned off with a "hard" (very low impedance) gate drive. This by itself will almost guarantee limited dynamic unbalance at turn-off.

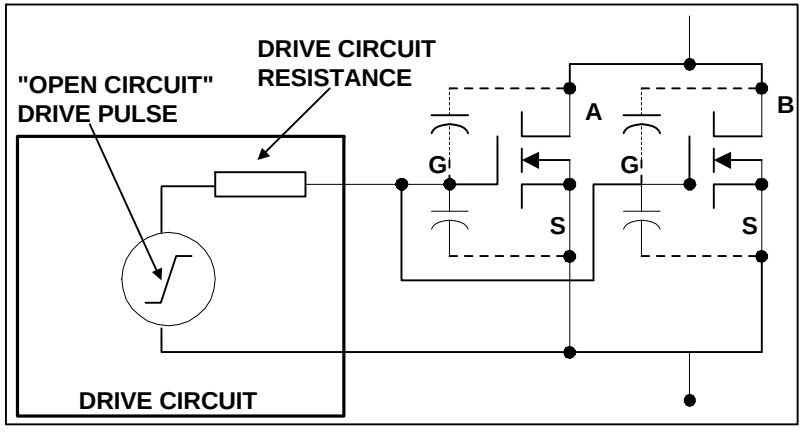


Figure 7. Waveforms at Turn-OFF - Two HEXFET Power MOSFETs with Common Gates and Common Sources

In summary, to achieve good sharing at turn-off the same precautions should be used as for turn-on, with the addition of a "hard" drive.

References:

1. J.B. Forsythe: "Paralleling of Power MOSFETs." IEEE-IAS Conference Record, October 1981.

Related topics:

Gate drive requirements
Thermal design
Paralleling of IGBTs

International
IOR Rectifier

Application Note AN-944

Use Gate Charge to Design the Gate Drive Circuit for Power MOSFETs and IGBTs

Table of Contents

Page

1. Input behavior of a MOS-gated transistor 1

2. Test Circuit 1

3. The Gate Charge Curve 3

4. Beware When Comparing Different Products 4

Designers unfamiliar with MOSFET or IGBT input characteristics begin drive circuit design by determining component values based on the gate-to-source, or input, capacitance listed on the data sheet. RC values based on the gate-to-source capacitance normally lead to a gate drive that is hopelessly inadequate.

Although the gate-to-source capacitance is an important value, the gate-to-drain capacitance is actually more significant—and more difficult to deal with—because it is a non-linear capacitance affected as a function of voltage; the gate-to-source capacitance is also affected as a voltage function, but to a much lesser extent.

Use Gate Charge to Design the Gate Drive Circuit for Power MOSFETs and IGBTs

Topics covered:

- *Background*
- *Test method*
- *How to interpret the gate charge curve*
- *How to estimate switching times*
- *How to compare different devices*

1. Input behavior of a MOS-gated transistor

Designers unfamiliar with MOSFET or IGBT input characteristics begin drive circuit design by determining component values based on the gate-to-source, or input, capacitance listed on the data sheet. RC value based on the gate-to-source capacitance normally lead to a gate drive that is hopelessly inadequate.

Although the gate-to-source capacitance is an important value, the gate-to-drain capacitance is actually more significant—and more difficult to deal with—because it is a non-linear capacitance affected as a function of voltage; the gate-to-source capacitance is also affected as a voltage function, but to a much lesser extent. This gate-to-drain capacitance function is similar to that found in vacuum tube amplifiers.

The gate-to-drain capacitance effect is akin to the “Miller” effect, a phenomenon by which a feedback path between the input and output of an electronic device is provided by the interelectrode capacitance. This affects the total input admittance of the device which results in the total dynamic input capacitance generally being greater than the sum of the static electrode capacitances. The phenomenon of the effects of the plate impedance and voltage gain on the input admittance was first studied in vacuum tube triode amplifier circuits by John M. Miller. Essentially, at high frequencies where the grid-to-plate (gate-to-drain) capacitance is not negligible, the circuit is not open but involves a capacitance that is a function of the voltage gain. Solving for the "Miller" effect is not exactly a straightforward process, even with vacuum tubes where much is known, but is even more difficult in MOSFETs. In actuality, the gate-to-drain capacitance though smaller in static value than the gate-to-source capacitance, goes through a voltage excursion that is often more than 20 times that of the gate-to-source capacity. Therefore, the gate-to-drain or “Miller” capacitance typically requires more actual charge than the input capacitance.

To account for both gate-to-source and gate-to-drain capacitance in a way readily usable by designers, International Rectifier supplies a “gate charge” specifications for its IGBTs and HEXFET POWER MOSFETs. that can be used to calculate drive circuit requirements. Gate charge is defined as the charge that must be supplied to the gate, either to swing the gate by a given amount, or to achieve full switching.

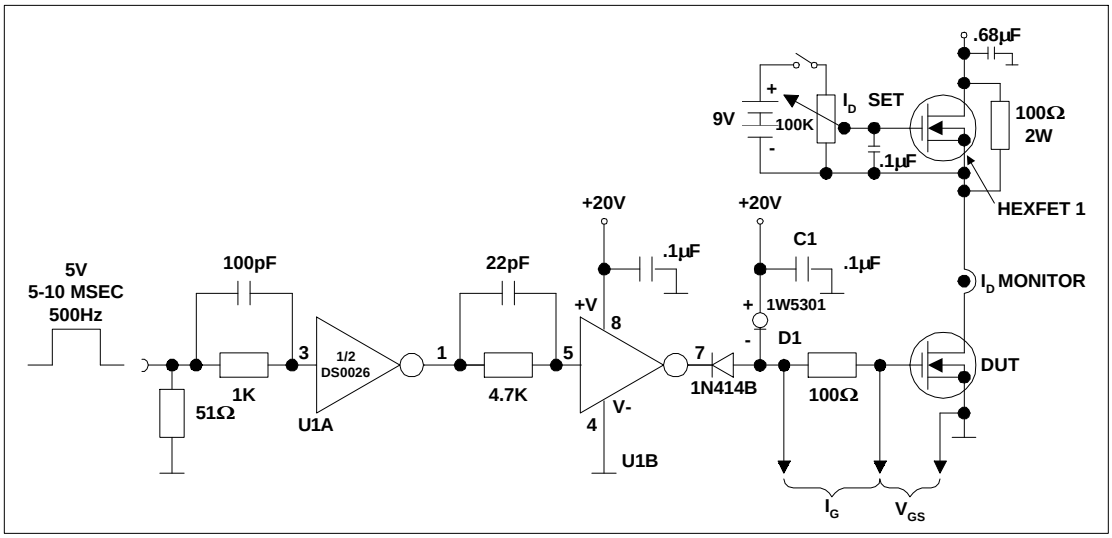


Figure 1. HEXFET POWER MOSFET Gate Charge Circuit.

2. Test Circuit

A typical test circuit that can be used to measure the gate charge is shown in Figure 1. In this circuit, an approximately constant current is supplied to the gate of the device-under-test from the 0.1 microfarad capacitor C1, through the regulator diode D1. A constant current in the drain circuit is set by setting the voltage on the gate of HEXFET POWER MOSFET 1, so the net measurement of the charge consumed by the gate is relative to a given current and voltage in the source-to-drain path.

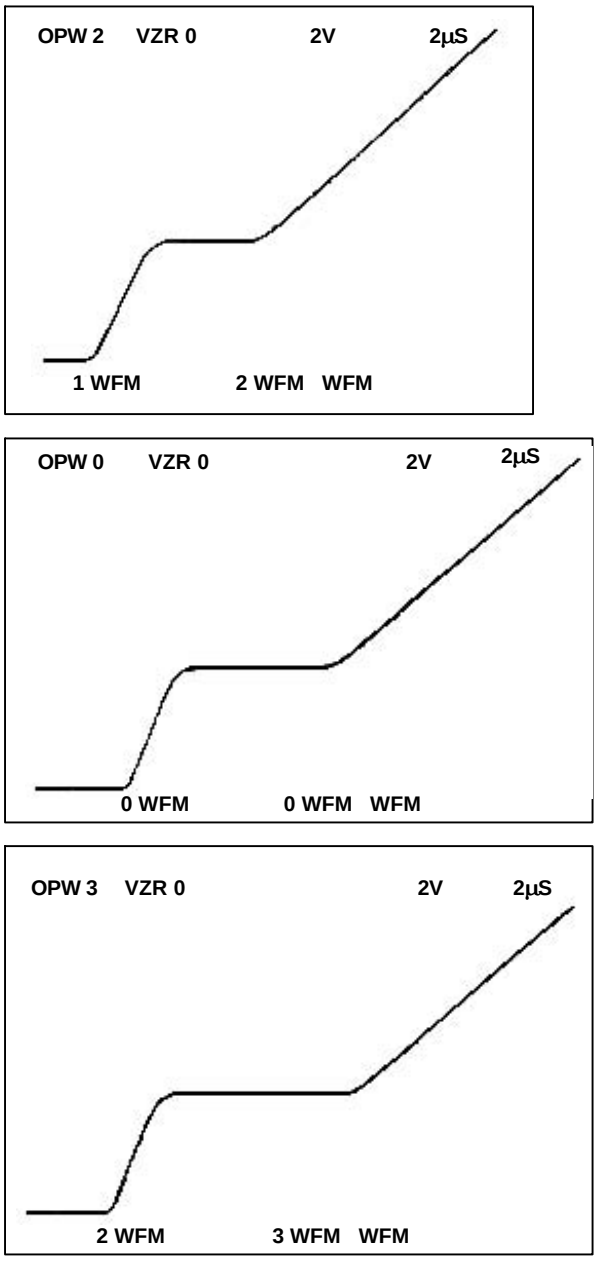


Figure 2. Gate Charge Waveform for Different Values of Drain Voltage (IRF130: l_G = 1.5 mA, I_D =1A, V_{DD} = 10, 40 and 80 volts).

The importance of the gate charge data to the designer is illustrated as follows. Taking the charge are required to switch a previous example, about 15 nanocoulombs of gate if 1.5 amps is supplied to the gate, the device will be drain voltage of 80 volts and a drain current of 12 amps. Since the 15 nC gate charge is the product of the gate input current and the switching time, switched in 10 nS. It follows that if 15 mA is supplied to the gate, then switching occurs in 1 us, and so on. These simple calculations immediately tell the designer the trade-offs between the amount of current available from the drive circuit and the achievable switching time. With gate charge known, the designer can develop a drive circuit appropriate to the switching time required.

Consider a typical practical example of a 100 kHz switcher, in which it is required to achieve a switching time of 100 nanoseconds.

An oscillogram of the gate-to-source voltage during testing, shown in Figure 2, relates the gate voltage to time. Since a constant current is supplied to the gate, the horizontal time scale is directly proportional to the charge supplied to the gate. With a suitable scaling factor, therefore, this oscillogram is a plot of gate voltage versus charge. The point on the oscillogram of the second voltage rise indicates where the device is fully switched on. During the first voltage rise, the gate-to-source capacitance is charging, and during the flat portion, the gate-to-drain capacitance is charging. This oscillogram therefore clearly differentiates between the charge required for the gate-source and gate-to-drain ("Miller") capacitances. At the second voltage rise, both capacitances are charged to the extent needed to switch the given voltage and current. A more detailed explanation of the interpretation of this data is given later. The graph in Figure 3 represents gate voltage versus gate charge in nanocoulombs for an IRF130. Although the second voltage rise indicates the point at which the switching operation is completed, normal design safety margins will dictate that the level of drive voltage applied to the gate is greater than that which is just required to switch the given drain current and voltage. The total charge consumed by the gate will therefore in practice be higher than the minimum required-but not necessarily significantly so. For example, the gate charge required to switch 12 amps at 80 volts is 15 nanocoulombs (point A), and the corresponding gate voltage is about 7 volts.

If the applied drive voltage has an amplitude of 10 volts (i.e. a 3 volt margin), then the total gate charge actually consumed would be about 20 nanocoulombs, (point B). As shown on the graph, whether switching 10 volts or 80 volts in the drain circuit, there is a much less than proportional difference in the charge required. This is because the "Miller" capacitance is a nonlinear function of voltage, and decreases with increasing voltage.

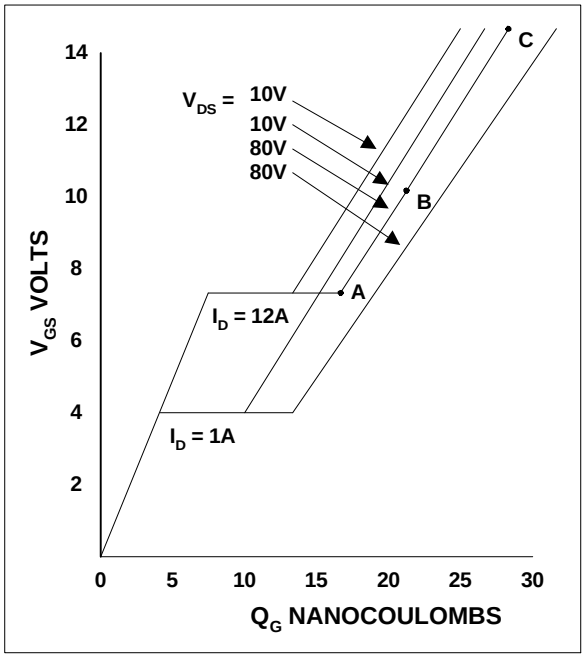


Figure 3. Gate Voltage Versus Gate Charge for the IRF130.

The required gate drive current is derived by simply dividing the gate charge, 15 X 10^{-9}, by the required switching time, 100 X 10^{-9}, giving 150 mA. From this calculation, the designer can further arrive at the drive circuit impedance. If the drive circuit applies 14 volts to the gate, for instance, then a drive impedance of about 50 ohms would be required. Note that throughout the "flat" part of the switching period (Figure 3), the gate voltage is constant at about 7 volts. The difference between the applied 14 volts and 7 volts is what is available to drive the required current through the drive circuit resistance.

The gate charge data also lets the designer quickly determine average gate drive power. The average gate drive power, P_{DRIVE}, is $Q_G V_G f$. Taking the above 100 kHz switcher as an example, and assuming a gate drive voltage V_G of 14 volts, the appropriate value of gate charge Q_G is 27 nanocoulombs (point C on Figure 3). The average drive power is therefore 27 X 10^{-9} X 14 X 10^5 = 0.038 Watts. Even though the 150 mA drive current which flows during the switching interval may appear to be relatively high. the average power is minuscule (0.004%) in relation to the power being switched in the drain current. This is because the drive current flows for such a short period that the average power is negligible. Thus actual drive power for MOSFETs is minute compared to bipolar requirements, which must sustain switching current during the entire ON condition. Average drive power, of course, increases at higher frequencies, but even at 5 MHz it would be only 1.9W.

3. The Gate Charge Curve

The oscillograms of the gate-to-source voltage in Figure 2 neatly delineate between the charge required for the gate-to-source capacitance, and the charge required for the gate-to-drain, or "Miller" capacitance. The accompanying simplified test circuit and waveform diagram (Figures 4 and 5 respectively) give the explanation. Before time t_0, the switch S is closed; the device under test (DUT) supports the full circuit voltage, V_{DD}, and the gate voltage and drain current are zero. S is opened at time t_0; the gate-to-source capacitance starts to charge, and the gate-to-source voltage increases. No current flows in the drain until the gate reaches the threshold voltage. During period T_1 to t_2, the gate-to-source capacitance continues to charge, the gate voltage continues to rise and the drain current rises proportionally. So long as the actual drain current is still building up towards the available drain current, I_D, the freewheeling rectifier stays in conduction, the voltage across it remains low, and the voltage across the DUT continues to be virtually the full circuit voltage, V_{DD}. The top end of the drain-to-gate capacitance C_{AD} therefore remains at a fixed potential, whilst the potential of the lower end moves with that of the gate. The charging current taken by C_{AD} during this period is small, and for practical purposes it can be neglected, since C_{AD} is numerically small by comparison with G_{CS}.

At time t_2, the drain current reaches I_D, and the freewheeling rectifier shuts off; the potential of the drain now is no longer tied to the supply voltage, V_{DD}. The drain current now stays constant at the value I_D enforced by the circuit, whilst the drain voltage starts to fall. Since the gate voltage is inextricably related to the drain current by the intrinsic transfer characteristic of the DUT (so long as operation remains in the "active" region), the gate voltage now stays constant because the "enforced" drain current is constant. For the time being therefore, no further charge is consumed by the gate-to-source capacitance, because the gate voltage remains constant. Thus the drive current now diverts, in its entirety, into the "Miller" capacitance C_{AD}, and the drive circuit charge now contributes exclusively to discharging the "Miller" capacitance.

The drain voltage excursion during the period t_2 to t_3 is relatively large, and hence the total drive charge is typically higher for the "Miller" capacitance C_{AD} than for the gate-to-source capacitance G_{CS}. At t_3 the drain voltage falls to a value equal to I_D x $R_{DS(ON)}$, and the DUT now comes out of the "active" region of operation. (In bipolar transistor terms, it has reached "saturation."

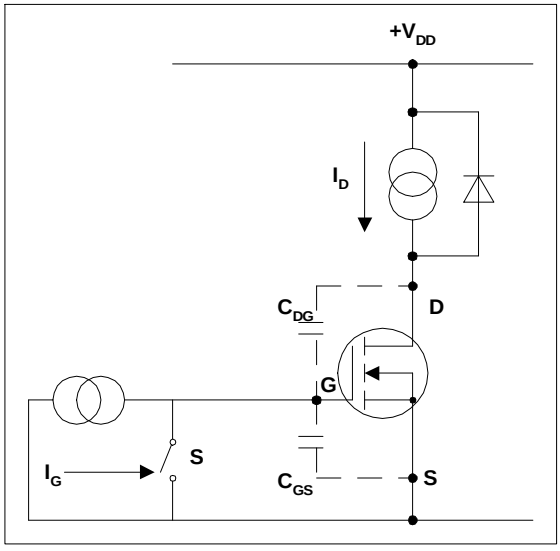


Figure 4. Basic Gate Charge Test Circuit

The gate voltage is now no longer constrained by the transfer characteristic of the device to relate to the drain current, and is free to increase. This it does, until time t_4, when the gate voltage becomes equal to the voltage "behind" the gate circuit current source. The time scale on the oscillogram of the gate-to-source voltage is directly proportional to the charge delivered by the drive circuit, because charge is equal to the product of current and time, and the current remains constant throughout the whole sequence. Thus the length of the period t_0 to t_1 represents the charge Q_{GS} consumed by the gate-to-source capacitance, whilst the length of the period t_2 to t_3 represents the charge Q_{GD} consumed by the gate-to-drain or "Miller" capacitance. The total charge at time t_3 is the charge required to switch the given voltage V_{DD} and current I_D. The additional charge consumed after time t_3 does not represent "switching" charge; it is simply the excess charge which will be delivered by the drive circuit because the amplitude of the applied gate drive voltage normally will be higher (as a matter of good design practice) than the bare minimum required to accomplish switching.

4. Beware When Comparing Different Products

Manufacturers sometimes make technical claims for their products that appear to be plausible, but which in actuality do not stand up to scrutiny. A case in point concerns the input capacitance of a power MOSFET. Statements such as "the input capacitance of device Y is less than that of device X, ergo Y is a faster switch than X", are frequently bandied about, but are just as frequently erroneous. Apart from the obvious speciousness of many such statements — "apples" are frequently not compared with "apples", and obviously larger chips have more self capacitance than smaller ones—the more basic fundamentals are generally overlooked. As this application note shows, of "bottom line" importance is the total gate charge required for switching. The lower the charge, the lower is the gate drive current needed to achieve a given switching time.

A general comparison between hypothetical MOSFETs brands "X" and "Y" is illustrated in the Figure. Device X has a higher input capacitance; hence the initial slope of its gate charge characteristic is less than that of device Y. Q_{GS} of device X is, however, about the same as that of device Y, because it has a higher transconductance and therefore requires less voltage on its gate for the given amount of drain current (V_{GX} is less than V_{GY}) The "Miller" charge consumed by device X is considerably less than that consumed by device Y. The overall result is that the total charge required to switch device X, Q_X, is considerably *less* than that required to switch device Y, Q_Y. Had the comparison between devices X and Y been made on the more superficial basis of input capacitances, it would have been concluded— erroneously— that Y is "better" than X. Another consideration is the energy required for switching. Again, device X scores handsomely over device Y in this example. The energy is the product of the gate charge and the gate voltage, and is represented by the area of the rectangle whose corner lies at the "switching point". (Point 1 for device X, and point 2 for device Y.) It is obvious that X requires significantly less gate energy than Y. To summarize: beware of superficial comparisons. Check the full facts before deciding which MOSFET really has the edge in switching performance.

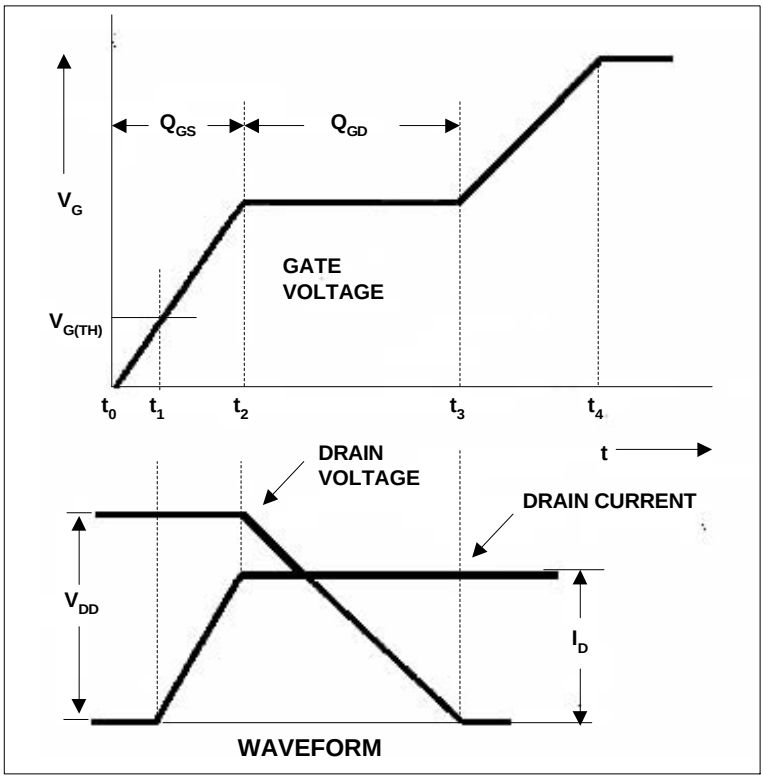


Figure 5. Basic Gate Charge Waveforms

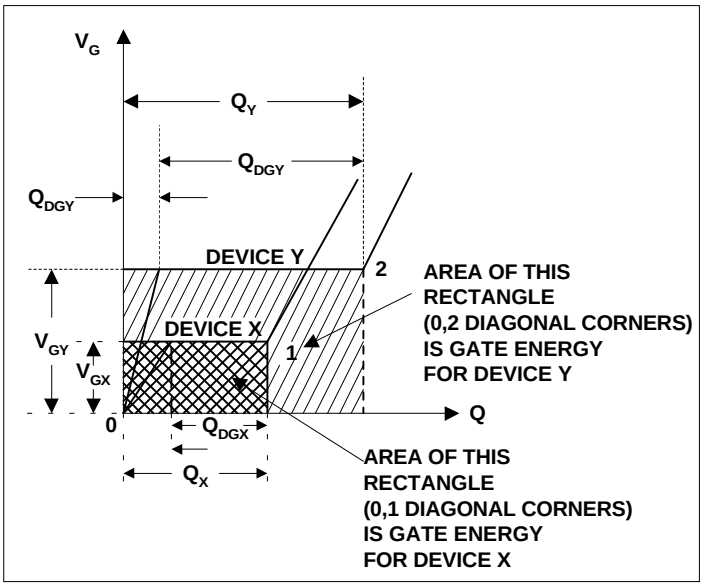


Figure 6. Comparison of Gate Charge Characteristics of Different Device Types.

Related Topics:

Gate drive considerations for IGBT modules
Gate drive characteristics of IGBTs
Gate drive requirements of MOS-gated transistors
High-voltage gate drive ICs
Three-phase gate drive IC
Gate drive IC for ballasts
Transformer-isolated gate driver

International
IOR Rectifier

Application Note AN-948

Linear Power Amplifer Using Complementary HEXFET® Power MOSFETs

Table of Contents

	Page
1. Circuit Description	1
2. Amplifier Performance	3
3. Power Supply Regulation Requirements	5
4. Set-Up Procedure and Troubleshooting	5
5. Performance Summary	6

The class AB amplifier described in this application note uses a complementary pair of HEXFET® power MOSFET devices as the output stage. This feature offers performance improvements over the equivalent bipolar output stage and allows a reduction in the complexity of the driver circuit, the output devices being driven by a single class A driver.

Linear Power Amplifier Using Complementary HEXFETs®

(HEXFET is the trademark for International Rectifier Power MOSFETs)

Topics covered:

A description of the circuit
Performance
Power supply requirements
Set-up and troubleshooting
Performance summary
Related topics

The class AB amplifier described in this application note uses a complementary pair of HEXFET Power MOSFET devices as the output stage. This feature offers performance improvements over the equivalent bipolar output stage and allows a reduction in the complexity of the driver circuit, the output devices being driven by a single class A driver. At the time the work was done, logic level HEXFET Power MOSFETs were not available and standard devices were used. The use of low-threshold devices would simplify the bias circuit and reduce the quiescent power dissipation The design described will deliver 60W rms into a 4 ohm load when working from +30V supplies. The bandwidth is in, excess of l00kHz, but may be tailored to the user requirements by making component value changes.

1. CIRCUIT DESCRIPTION

The amplifier circuit diagram is shown in Figure 1, and the components listing in Table 1. Split power supply rails ($\pm V_{DD}$) are used, giving improved rejection of power supply ripple and allowing the load, R_{load}, to be direct coupled. The output devices Q_5, Q_6 , operate in source follower configuration. This offers a twofold advantage; a) the possibility of oscillation in the power output stage is reduced as the voltage gain is less than unity and b) signal feedback through the heatsink on which the devices are assembled is eliminated as the drain terminal, which is electrically connected to the tab on the TO220 package, is at dc voltage.

Symmetrical output is achieved by providing "bootstrapped" drive to the gate of the n-channel device, Q_5, from the output. The use of the bootstrap circuit, C_4, R_8, R_9 also allows the driver transistor, Q_4, to operate at near constant current which improves the linearity of the driver stage. The diode D_1 acts as a clamp for the bootstrap circuit, restricting the positive voltage at the gate of Q_5 to $+V_{DD}$. This allows symmetry to be maintained under overload conditions.

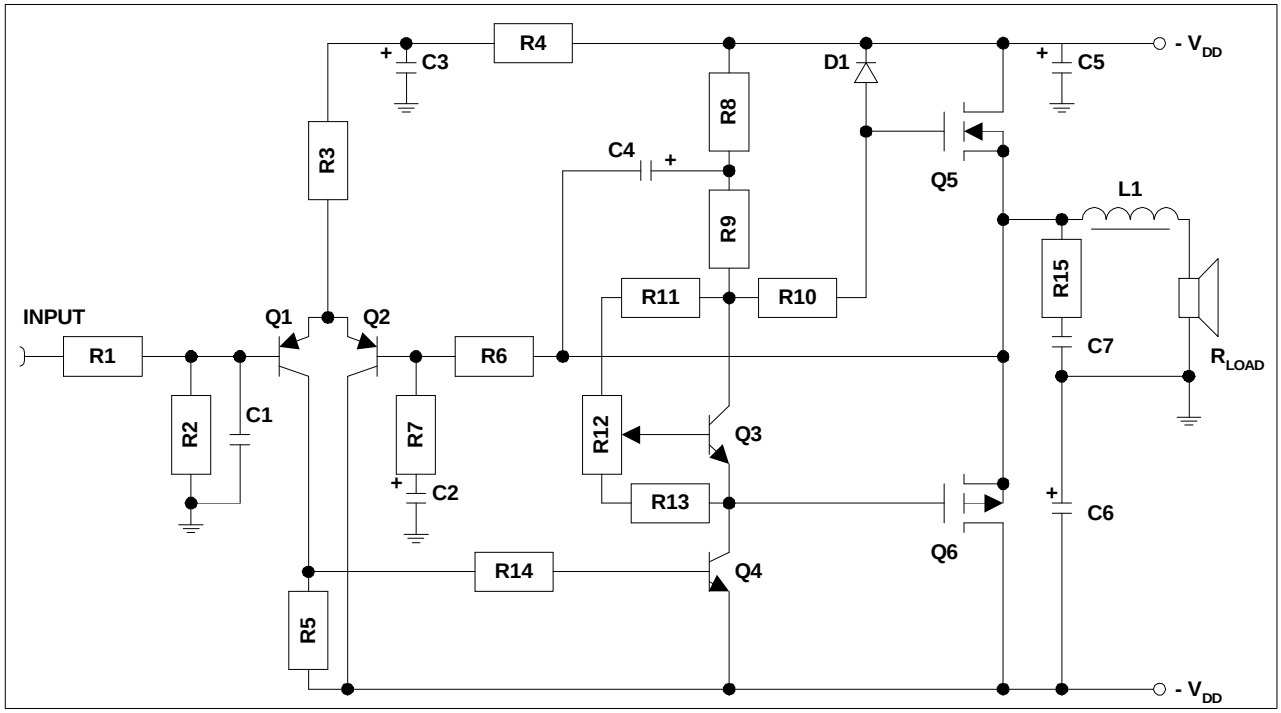


Figure 1. Class AB Amplifier Circuit Diagram

Transistor Q_3 and resistors R_{11}, R_{12}, R_{13} provide gate-source offset voltage for the output devices. R_{12} is variable, allowing adjustment of the output quiescent current for variation in HEXFET Power MOSFET threshold voltage. A degree of temperature compensation is built into the circuit as both the bipolar transistor, Q_3, emitter base voltage and the combined threshold voltages of the HEXFET Power MOSFETs, Q_5, Q_6 have a temperature coefficient of -0.3%/°C.

Components List							
R_1	4.7KΩ	R_9	2.7KΩ	C_1	220 pF	Q_1 , Q_2 ,	2N4356, 2N5086
R_2	47KΩ	R_{10}	680Ω	C_2	100 µF 10V		or equivalent
R_3	15KΩ	R_{11}	10KΩ	C_3	47 µF 40V	Q_3 , Q_4 ,	2N4410, 2N5088
R_4	1.2KΩ	R_{12}	1KΩ Pot.	C_4	47 µF 40V		or equivalent
R_5	560Ω	R_{13}	820Ω	C_5	2200 µF 40V	Q_5	IRF532
R_6	47KΩ	R_{14}	4.7KΩ	C_6	2200 µF 40V	Q_6	IRF9532
R_7	4.70Ω	R_{15}	10 Ω 1 W	C_7	68 nF	D_1	IN4002
R_8	2.7KΩ	R_{load}	8/4 Ω	L_1	3 µH aircored		

Table 1. Components List

The class A driver transistor, Q_4, operates at a bias current determined by resistors R_8, R_9, nominally 5mA. Q_4 is driven by a PNP differential input pair, Q_1, Q_2. The bias current in the input stage is set to 2mA by resistor R_3. Negative feedback from the output of the amplifier is fed to the base of Q_2 by resistor R_6. Components R_7, C_2 set the closed loop gain of the amplifier (R_6 /R_7) and provide low frequency gain boosting. The additional components R_{15}, C_7 connected between the output node and ground suppress the high frequency response of the output stage, allowing the h.f. performance of the amplifier to be determined by the input circuit. Components R_1, R_2, C_1 at the input of the amplifier define the input impedance (47Kohm) and suppress noise.

The amplifier input stage requires additional power supply ripple suppression which is provided by components R_4, C_3.

Additional circuit components have been added to ensure high frequency stability of the complete amplifier. Placement of the components and component values will depend to some extent on the printed circuit board layout. The following rules should be followed when designing the printed circuit board:

(a) A 'common ground' principle should be

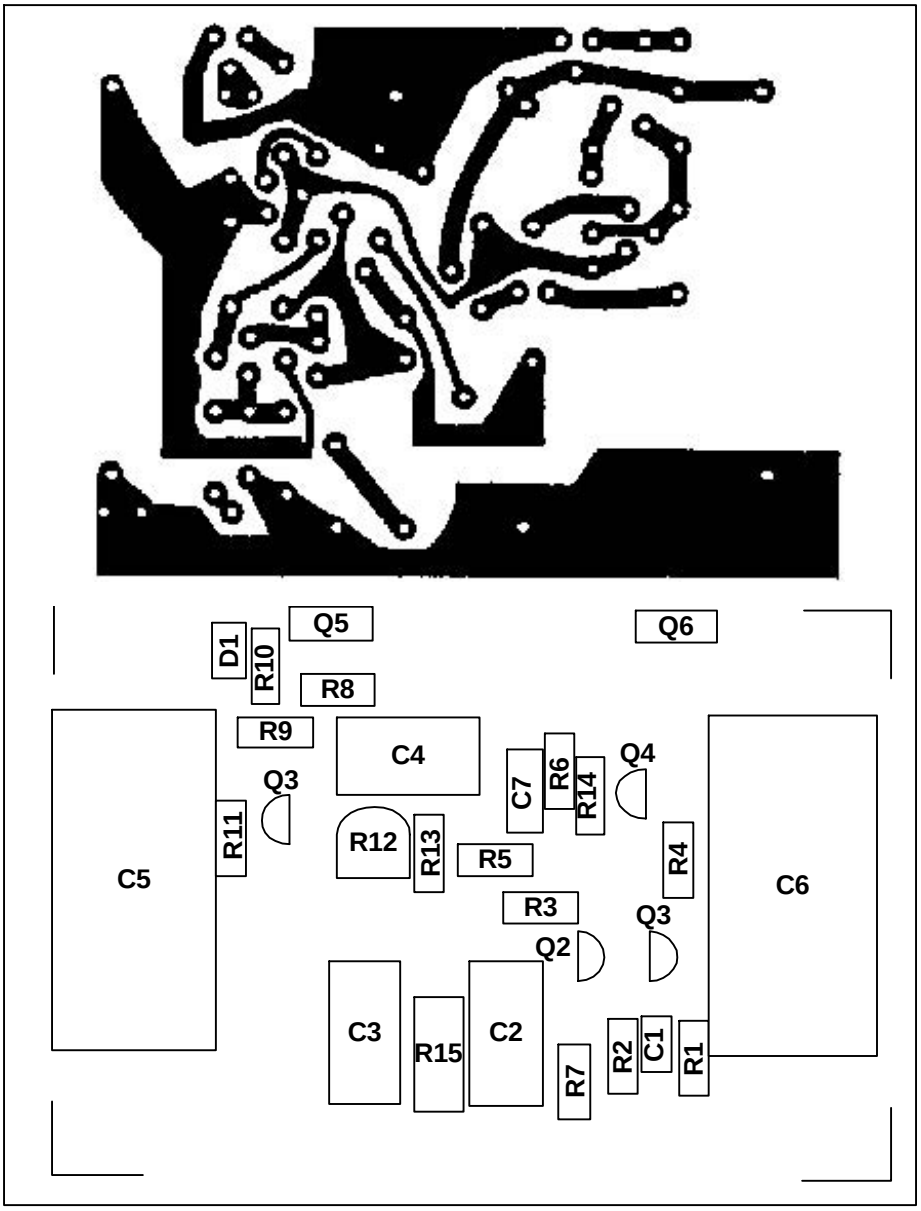


Figure 2. Amplifier Printed Circuit Board Layout

adopted, i.e., power supply decoupling capacitors, load and input stage biasing components should all be taken to ground in close proximity, eliminating the effects of common node ground current. Similarly, a "common output node" should be used, the load, feedback resistor and h.f. suppression components being taken from a common point on the pcb.

(b) The length of connecting lead to the gate terminals of HEXFET Power MOSFETs Q_5, Q_6 should be an absolute minimum to avoid oscillation of the power output stage. A series gate resistor, R_{10}, may be used to suppress oscillation, but too high a resistor value will limit the slew rate. Oscillation of the amplifier caused by capacitive coupling to the base of the driver transistor, Q_4, is suppressed by the addition of a series resistor, R_{14}.

(c) Phase shift in the amplifier when driving a reactive load can lead to high frequency instability. With a capacitive load, the addition of a small, air-cored choke (3µH with an 8 ohm, 2µF load) will restore stability. The final value of the choke is defined by experiment.

Figure 2 shows a printed circuit layout which can be used for the circuit shown in Figure 1. The preceding design rules have been followed.

2. AMPLIFIER PERFORMANCE

(a) *Output Power:* To achieve 60W rms into a 4 ohm load, the current in the load is 3.9A rms or 5.5A pk. This information is derived from equations (1) and (2):

$$P_O = I^2_{rms} R_{load} = \frac{V^2_{rms}}{R_{load}} \qquad (1)$$

$$I_{rms} = \frac{I_{pk}}{\sqrt{2}} \qquad (2)$$

$$V_{rms} = \frac{V_{pk}}{\sqrt{2}} \qquad (3)$$

Also from equation (1), the voltage developed across the load at 60W output is 15.5V rms or 22V pk. To sustain a source current of 5.5A, the n-channel HEXFET Power MOSFET, IRF530, requires a gate-source voltage of 5V. One can conclude that the gate bias voltage to achieve peak power in the positive sense is $V_{pk} + V_p = 27V$. A similar calculation for the negative peak, using the P channel HEXFET Power MOSFET IRF9530, shows that a negative gate bias supply of -28V is required.

Consequently, a ±30V supply will be adequate for a 60W output, provided that the supply voltage does not fall below ±28V when loaded, i.e., the power supply impedance should be better than 1 ohm. The relationship between the power delivered to the load and the power absorbed from the supply is shown in Figure 3, assuming a sinusoidal waveform and a ±30V supply. The curve representing the power delivered to the load can be easily plotted with the help of equation (1) for different values of load current.

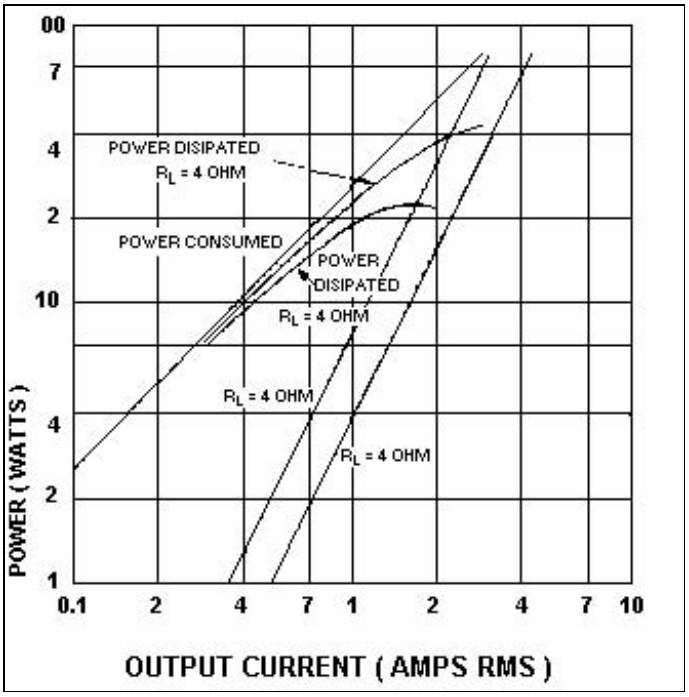


Figure 3. Power Curves for the Amplifier with 4 and 8 Ohm Loads and ± 30V Power Suuplies

The power absorbed from the supply has been plotted with the help of the following relationship:

$$P_S = \left(V_{avg}\right)\left(I_{avg}\right)$$
$$= \left(2V_{DC}\right)\left(\frac{\sqrt{2}}{\pi} I_{rms}\right) \qquad (4)$$

The difference between the two is the power dissipated in the HEXFET Power MOSFETs and as it can be seen from Figure 3, it has a peak of approximately 46W. Assuming a maximum ambient temperature of 55°C, the total thermal resistance between the junction of the two HEXFET Power MOSFETs and the ambient will have to be less than 2°C/W. Considering that the IRF530 and IRF9530 have a thermal resistance between junction and case of 1.7°C/W each, the maximum case temperature will have to be less than 110°C and the thermal resistance of the heatsink will have to be less than 1.15°C/W to ambient.

(b) *Frequency Response*: Open shown in Figure 4. The open loop gain, measured with gate and source connections to the HEXFET Power MOSFETs broken, is 30db, -3db points occurring at 15 Hz and 60kHz. Closed loop curves are shown for amplifier gains of 100 (R_7 = 470 ohm) and 20 (R_7 = 2.2K ohm). In both instances the curves remain flat to within +1db between 15 Hz and 100kHz with an 8 ohm load. The slew rate of the amplifier, measured with a 2V pk-pk square wave input is 13V/µs positive going and 16V/µs negative going. The discrepancy could be balanced out by addition of a series gate resistor for Q_6.

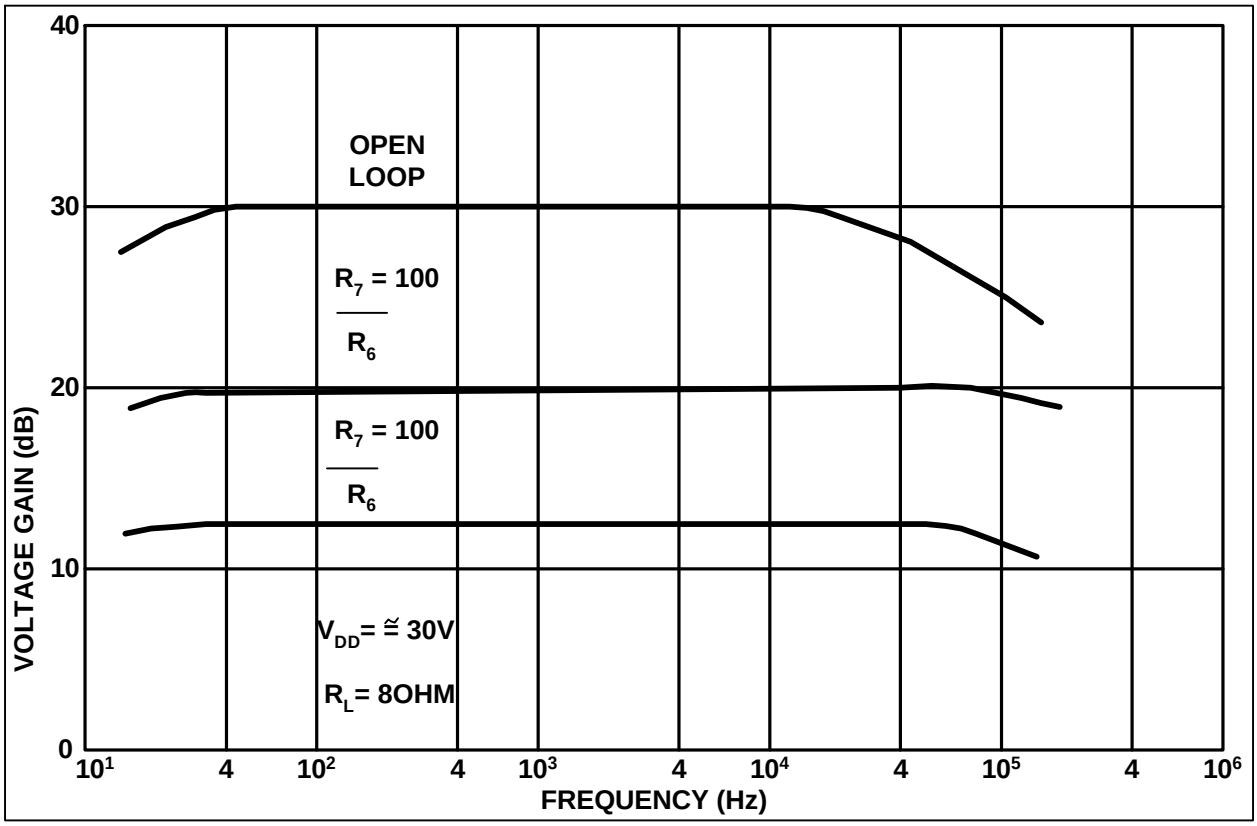


Figure 4. Amplifier Open and Closed Loop Frequency Response Curves

(c) *Total Harmonic Distortion*: The fidelity of the amplifier is shown in the distortion curves, Figure 5, and is limited by the loop gain. Reduction of the closed loop gain from 100 to 20 produces a significant improvement in distortion figure. The output stage quiescent current was adjusted to 100mA and can influence the distortion measurement significantly if allowed to fall below 50mA.

(d) Quiescent Operating: The dependence of the quiescent current in the output stage and of the offset voltage on power supply voltage are illustrated in Table 2. The quiescent current is set by first adjusting the potentiometer, R_{12}, for minimum offset voltage-turned fully anticlockwise if the pcb layout in Figure 2 is used - and applying the power supply voltage, the positive

supply passing through an ammeter with 1A f.s.d. R_{12} is then adjusted until the meter reading is 100mA with ±30V supplies. The meter should be removed from circuit before applying an input signal to the amplifier.

3. POWER SUPPLY REGULATION REQUIREMENTS

A simple line derived power suitable for the class AB amplifier is illustrated in Figure 6. The ±30V supplies are taken from the centertapped secondary of the line transformer. The 2200µF decoupling capacitors, (Figure 1) which should be mounted as close as possible to the amplifier output stage, reduce the supply frequency ripple to 5.5V pk-pk at full load.

4. SET-UP PROCEDURE AND TROUBLESHOOTING

It is unlikely that any experienced experimenter will have difficulty in achieving satisfactory results when building an amplifier to this design. The printed circuit board shown in Figure 2 is intended to assist in this respect. The major problems anticipated are those associated with the faulty assembly of components and damage to the HEXFET Power MOSFETs through handling or circuit oscillation.

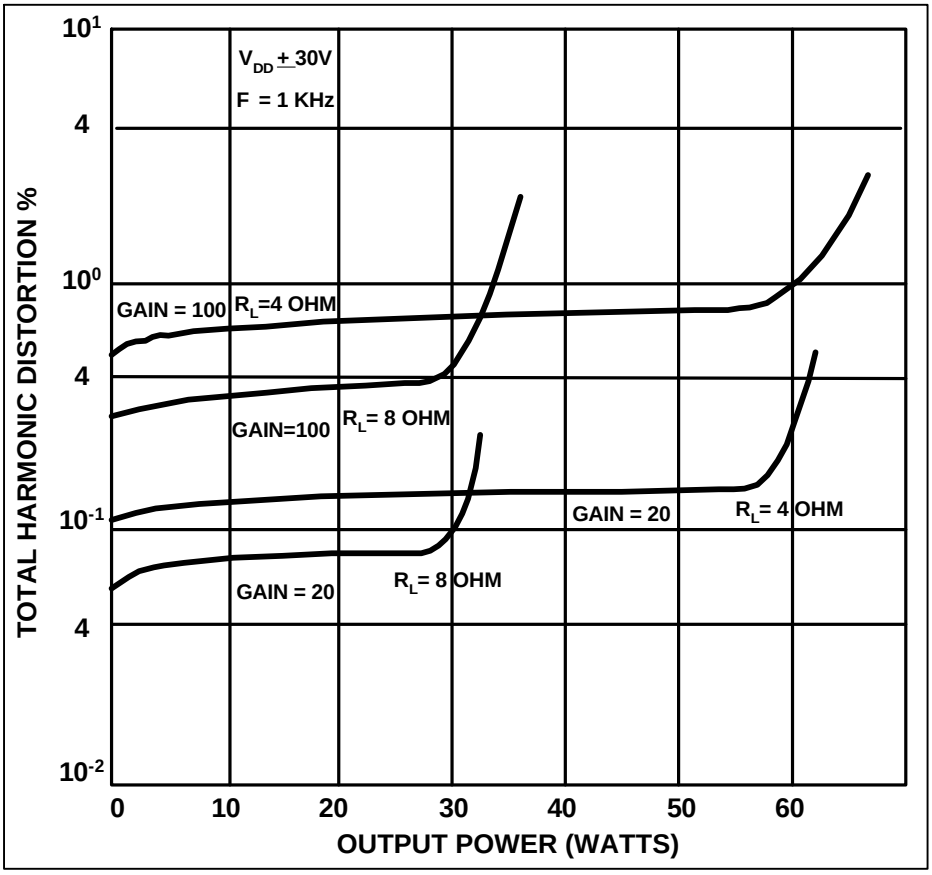


Figure 5. Amplifier Total Harmonic Distortion Curves

The following troubleshooting checklist is offered as a guide to the experimenter:

1. When assembling the printed circuit board, mount the passive components first, ensuring the correct polarity of electrolytic capacitors. Then solder in the transistors Q_1 - Q_4 checking for correct pin identification. Finally, mount the HEXFET Power MOSFETs, avoiding static discharge by shorting the pins together to ground and using a grounded soldering iron. Check the assembled board for correct component placement. A component side overlay as shown in Figure 2 is useful for this purpose. Check the copper side of the board for solder bridges between tracks, and remove them. Check for dry solder joints visually and electrically using a resistance meter and rework, if necessary.

2. Power can now be applied to the amplifier and the output stage quiescent current set to between 50 and 100mA. The potentiometer, R_{12} is first adjusted for minimum offset (fully anticlockwise on the pcb layout in Figure 2). An ammeter is connected in series with the positive supply and selected to read 1A f.s.d R_{12} is adjusted until the ammeter reads between 50mA and 100mA.

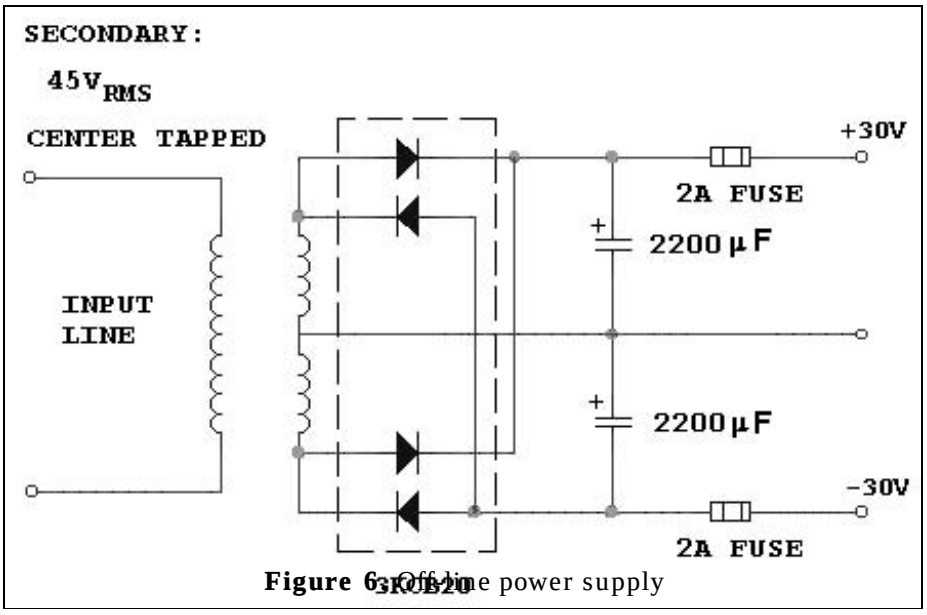


Figure 6. Off-line power supply

Quiescent current setting can be performed without the load connected.
If, however, a loudspeaker load is connected in circuit, it can be protected by a fuse from dc overload. With the quiescent current set, the output offset voltage can be confirmed to be less than 100mV. Excessive and erratic variation in quiescent current as R_{12} is adjusted are indicative of circuit oscillation or faulty wiring. The solutions described in "Circuit Description" (series gate resistors, minimized gate wiring and common earthing) should be adopted. Also, supply decoupling capacitors should be mounted in close proximity to the amplifier output stage and load ground point. Quiescent current setting should be performed with the HEXFET Power MOSFETs mounted on their heatsink to avoid overdissipation.

3. With the quiescent current set, the ammeter should be removed from the positive supply and a signal can be applied to the amplifier input. Signal requirements for full rated output are:

gain = 100:
R_L = 4 Ω, V_{in} = 150mV rms
R_L = 8 Ω, V_{in} = 160mV rms

gain = 20:
R_L = 4 Ω, V_{in} = 770mV rms
R_L = 8 Ω, V_{in} = 800mV rms

Variation in Output Offset Voltage and Output Quiescent Current with Supply Voltage

Supply Voltage + V_{DD} V	Output Offset V_{OS} mV	Output Quiescent Current Iq mA
35	-40	135
30	-20	100
25	+4	75
20	+30	54

Table 2. Variation in Output offset voltage and Output quiescent current with supply voltage

"Clipping" of the output waveform when operating at rated power indicates poor supply regulation and can be remedied by reducing the input signal amplitude and derating the amplifier. Alternatively, a lower impedance supply should be used. The frequency response of the amplifier can be checked over the frequency range 15 Hz - 100kHz with the aid of an audio test set or signal generator and oscilloscope. Distortion of the output waveform at high frequency is indicative of a reactive load and adjustment of the output choke will be required to restore the waveform. The high frequency response may be tailored with a compensation capacitor in parallel with R_6. The low frequency response is controlled by components R_7, C_2.

4. Hum pickup will be more likely to occur in a high gain circuit. Pickup at the high impedance input is minimized by use of a shielded cable, grounded at the signal source. Supply frequency ripple injected through the supply to the input stage of the amplifier can be detected across capacitor, C_3. This is attenuated by the common mode rejection ratio of Q_1, Q_2 before being amplified. However, if this is the source of hum, adjustment to the values of C_3, R_5 can be made to suppress the signal amplitude.

5. In the event of the output stage being destroyed, either through short circuit load or h.f. oscillation, both HEXFET Power MOSFETs should be replaced. It is unlikely, however, that other circuit components will have been affected. The setup procedure should, of course, be repeated with the new devices in circuit.

5. PERFORMANCE SUMMARY

Using a complementary pair of HEXFET Power MOSFETs, IRF530 and IRF9530 and with a ±30V supply, the following performance can be achieved:

Maximum rms output power:

60W into 4Ω
32W into 8Ω

Bandwidth:

15 Hz to 100kHz ±1db

Total harmonic distortion (1kHz):

0.15% at 60W into 4Ω
0.08% at 32W into 8Ω

Voltage gain:

Adjustable, X100 to X20

Input impedance: 47KΩ

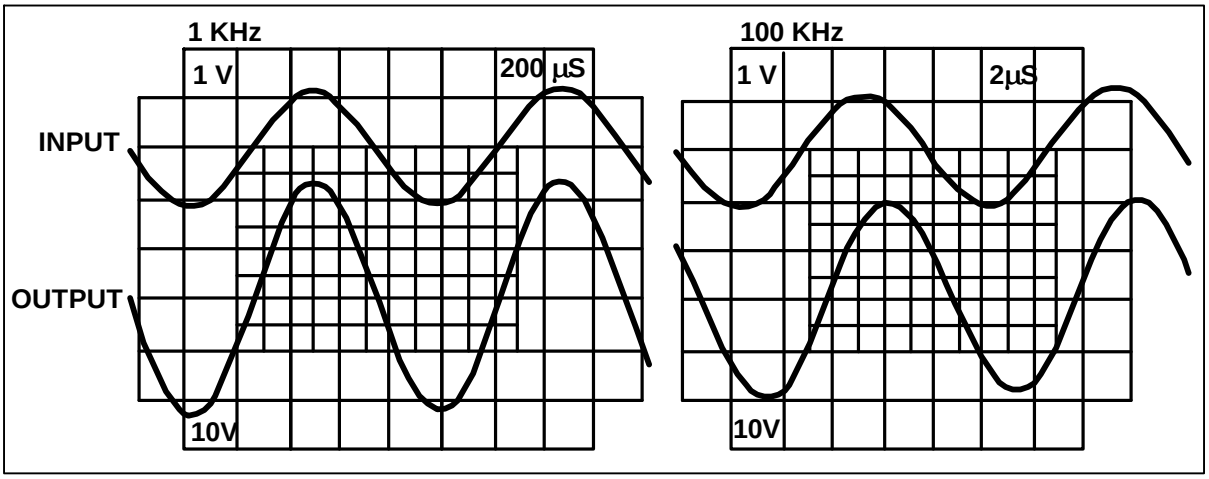


Figure 7a: Sinewave Frequency Response

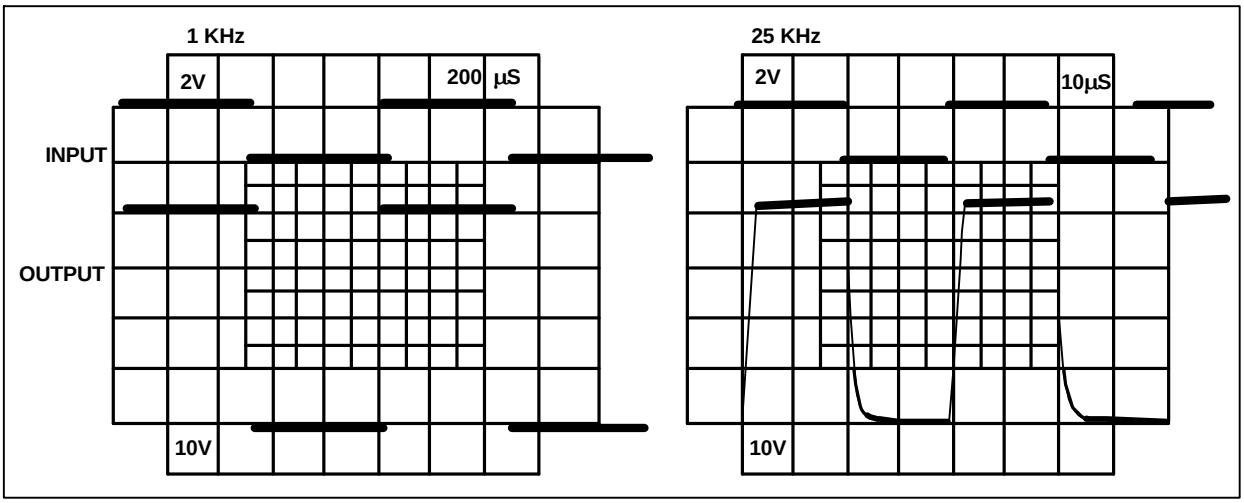


Figure 7b: Square Wave Response

Figure 7a & 7b: Amplifier Waveforms illustrating Frequency Response

Figure 7 illustrates the amplifier response to 1kHz and 100kHz sinewave input signals and also the square wave response at 1kHz and 25kHz.

Related topics:

P-Channel HEXFET Power MOSFETs
Gate drive requirements

International
IOR Rectifier

WORLD HEADQUARTERS: 233 Kansas St., El Segundo California 90245, Tel: (310) 322 3331
EUROPEAN HEADQUARTERS: Hurst Green, Oxted, Surrey RH8 9BB, UK Tel: ++ 44 1883 732020
IR CANADA: 7321 Victoria Park Ave., Suite 201, Markham, Ontario L3R 2Z8, Tel: (905) 475 1897
IR GERMANY: Saalburgstrasse 157, 61350 Bad Homburg Tel: ++ 49 6172 96590
IR ITALY: Via Liguria 49, 10071 Borgaro, Torino Tel: ++39 11 451 0111

IR FAR EAST: K&H Bldg., 2F, 3-30-4 Nishi-Ikeburo 3-Chrome, Toshima-Ki, Tokyo Japan 171 Tel: 81 3 3983 0086
IR SOUTHEAST ASIA: 315 Outram Road, #10-02 Tan Boon Liat Building, Singapore 0316 Tel: 65 221 8371
Data and specification subject to change without notice.

AND8199

Thermal Stability of MOSFETs

Prepared by: Alan Ball
ON Semiconductor
Application Engineering Manager

ON Semiconductor®

http://onsemi.com

A variety of applications use hot−swap controllers, often to increase the reliability of a system. However, a failure in the hot−swap circuit would defeat that purpose. When you use MOSFETs in their active region to control current, such as you would for a controller that operates in a constant−current mode of operation, they have an inherent failure mechanism. In this mode, the MOSFET can get hot spots and fail, long before the device exceeds its Safe Operating Area (SOA) ratings.

Engineers have long understood that MOSFETs are positive temperature coefficient devices. Therefore, as the temperature of the device increases, the resistance increases. In other words, higher temperatures result in lower currents. This fact is important if you want to operate MOSFETs in parallel. With a good thermal path between devices, the positive temperature coefficient reduces the current in the hottest device and forces more of it to flow in the cooler device, thereby avoiding thermal runaway.

Engineers often think of a MOSFET as a single power transistor, but it is a collection of thousands of tiny power FET cells connected in parallel. In terms of sharing current, the same application of the positive temperature coefficient applies. In this case, the thermal path between the cells is better than that of separate packaged devices, because the cells are all on the same die.

As the current density of a small group of cells increases, those cells heat up, increasing the resistivity of those cells and forcing current to flow in neighboring cells, which minimizes the thermal gradient and avoids hot spots. This process is an essential physical tenet that allows the parallel array of cells to function reliably.

If the MOSFET exhibits a negative thermal coefficient, today's parallel cell structure would cause serious reliability issues. In fact, in some modes of operation, the thermal coefficient goes negative. You can easily understand this phenomenon by looking at the transconductance curves for a FET device (refer to Reference 1).

A typical set of transconductance curves clearly demonstrates this effect as shown by Figure 1. Below are curves from three typical devices used in hot swap applications.

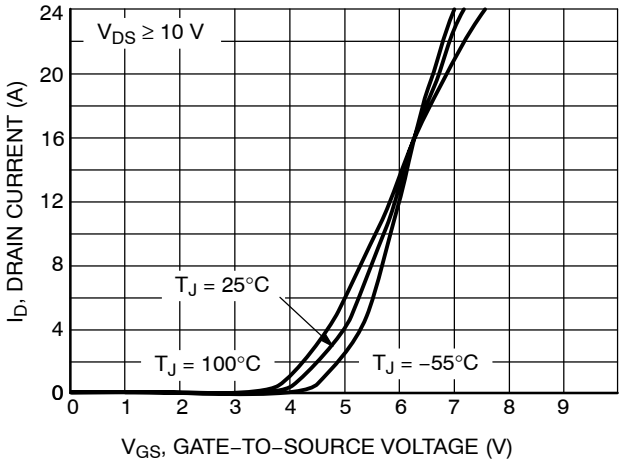


Figure 1. Transfer Characteristics for NTD12N10

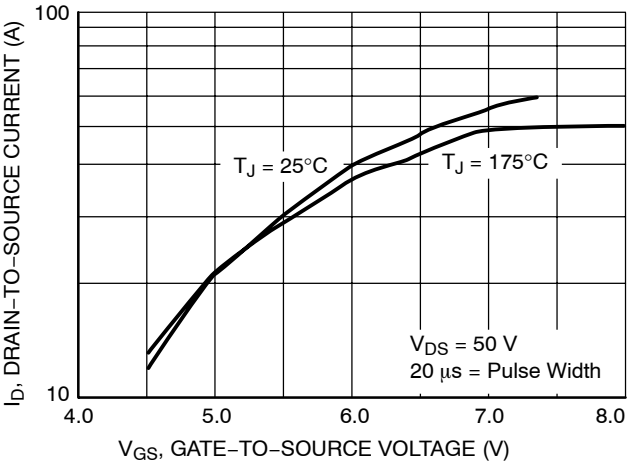


Figure 2. International Rectifier IRF530

© Semiconductor Components Industries, LLC, 2005

January, 2005 − Rev. 0

Publication Order Number:
AND8199/D

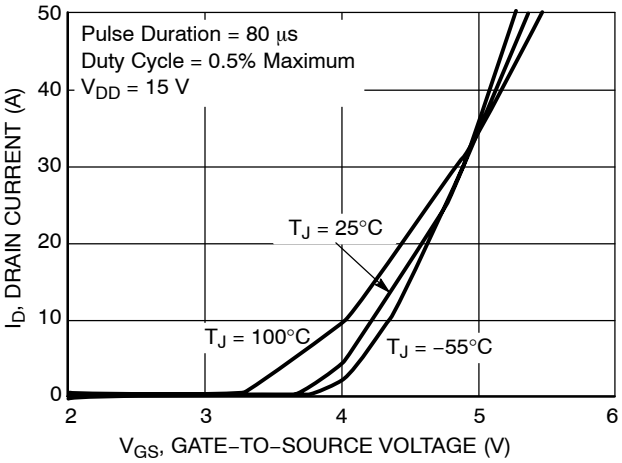


Figure 3. Fairchild HUF75631SK8

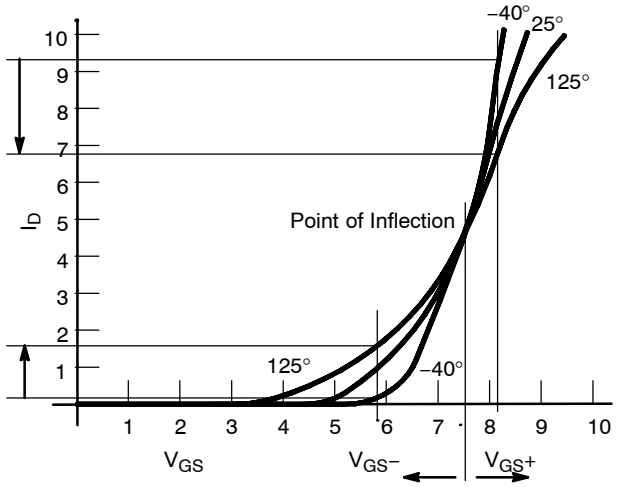


Figure 4. Typical Transconductance Curve

All three devices shown have one thing in common: a point of inflection at which the temperature coefficient is zero. At greater gate−to−source voltages, the coefficient is positive, and, at lower gate−to−source voltages it is negative.

Figure 4 illustrates the change from negative to positive. At a gate−to−source voltage greater than that of the inflection point (V_{GS}+), a positive temperature coefficient exists. At this gate voltage, the drain conducts more than 9.0 A of current. However, at 125°C the drain current reduces to less than 7.0 A. The arrow at the left of Figure 4, which shows the current decreasing due to an increase in temperature, indicates this drop.

At a gate−to−source voltage below the inflection point (V_{GS}−) a negative temperature coefficient exists. At −40°C, the drain current is close to zero. However, at 125°C, the drain current is more than 1.0 A. A second arrow at the left of Figure 4 indicates this relationship, and the current rises due to an increase in temperature.

The implication is that when you are controlling the FET with a gate−to−source voltage below the inflection point, thermal runaway can occur. When one cell or a small group of cells becomes hotter than the surrounding cells, they tend to conduct more current. This situation, in turn, creates more heat, which allows more current to flow. These cells can pull a large amount of current and, if not limited in time, can cause the device to fail.

This situation is similar to the well−known phenomenon of secondary breakdown that occurs in bipolar transistors except that a bipolar junction transistor is a single device, and you can take steps to avoid its destruction. A power MOSFET contains thousands of parallel devices that are internal to the die, and you cannot individually protect them. If hot spots occur, the SOA characteristics of the heavily conducting cells differ greatly from those of the marginally conducting cells.

The thermal−runaway situation occurs when you use large devices at low current−limit settings. Even though it would appear to be desirable to use a very large MOSFET for an application such as a hot swap and limit it to a low current, it may be an inappropriate approach. Use of a very−low−on−resistance device offers low losses for steady−state operation but may cause the device to fail during a short circuit or an overload.

One way to overcome this problem is to directly sense the die temperature of the MOSFET by integrating the MOSFET with the controller using a monolithic approach. ON Semiconductor takes this approach with its new line of hot−swap ICs. In this case, the temperature can be sensed directly on the FET die. The location of the sense element on the die is critical for ensured protection of the device. If a hot spot occurs too far from the sense location, the device may be unable to protect itself.

Discrete hot−swap controllers employ a number of protection schemes. Thermal instability is an issue only if the controller can go into a constant−current mode of operation. Some protection circuits simply shut off the MOSFET switch if a number of conditions indicate a dangerous area of operation. Controllers that use a constant−current method of protection can use timers or other schemes along with the current−limit circuit to reduce the risk of failure.

Because system efficiency is an important parameter, it is tempting to use the largest MOSFET possible to reduce losses. It is important to keep in mind, however, that this approach may require you to make a trade−off with the system reliability if you are not mindful of the possible thermal instability. You can reliably use a large power device at a low current limit level if you handle it properly.

Reference

1. *Thermal Instability of Low Voltage Power−MOSFET's.* IEEE Transactions on Power Electronics, Vol. 15, No. 3, May 2000, Alfio Consoli et al.

ON Semiconductor and are registered trademarks of Semiconductor Components Industries, LLC (SCILLC). SCILLC reserves the right to make changes without further notice to any products herein. SCILLC makes no warranty, representation or guarantee regarding the suitability of its products for any particular purpose, nor does SCILLC assume any liability arising out of the application or use of any product or circuit, and specifically disclaims any and all liability, including without limitation special, consequential or incidental damages. "Typical" parameters which may be provided in SCILLC data sheets and/or specifications can and do vary in different applications and actual performance may vary over time. All operating parameters, including "Typicals" must be validated for each customer application by customer's technical experts. SCILLC does not convey any license under its patent rights nor the rights of others. SCILLC products are not designed, intended, or authorized for use as components in systems intended for surgical implant into the body, or other applications intended to support or sustain life, or for any other application in which the failure of the SCILLC product could create a situation where personal injury or death may occur. Should Buyer purchase or use SCILLC products for any such unintended or unauthorized application, Buyer shall indemnify and hold SCILLC and its officers, employees, subsidiaries, affiliates, and distributors harmless against all claims, costs, damages, and expenses, and reasonable attorney fees arising out of, directly or indirectly, any claim of personal injury or death associated with such unintended or unauthorized use, even if such claim alleges that SCILLC was negligent regarding the design or manufacture of the part. SCILLC is an Equal Opportunity/Affirmative Action Employer. This literature is subject to all applicable copyright laws and is not for resale in any manner.

PUBLICATION ORDERING INFORMATION

LITERATURE FULFILLMENT:
Literature Distribution Center for ON Semiconductor
P.O. Box 61312, Phoenix, Arizona 85082−1312 USA
Phone: 480−829−7710 or 800−344−3860 Toll Free USA/Canada
Fax: 480−829−7709 or 800−344−3867 Toll Free USA/Canada
Email: orderlit@onsemi.com

N. American Technical Support: 800−282−9855 Toll Free USA/Canada

Japan: ON Semiconductor, Japan Customer Focus Center
2−9−1 Kamimeguro, Meguro−ku, Tokyo, Japan 153−0051
Phone: 81−3−5773−3850

ON Semiconductor Website: http://onsemi.com

Order Literature: http://www.onsemi.com/litorder

For additional information, please contact your local Sales Representative.

AND8199/D

ADVANCED POWER TECHNOLOGY®

APT0002
By: Richard Frey, P.E.
Denis Grafham
Tom Mackewicz

APPLICATION NOTE

New 500V Linear MOSFETs for a 120 kW Active Load

Presented at PCIM 2000
Nuremberg, Germany
June 7, 2000

New 500V Linear MOSFETs for a 120 kW Active Load

Authors: Richard Frey, Denis Grafham - APT, Tom Mackewicz - TDI-Dynaload.

Summary

Power MOSFETs are generally designed as switches. Total power dissipated is the sum of "on"-state losses (Id^2 x Rds(on)) plus losses generated during the very short switching intervals. Desirable characteristics include the lowest possible ON resistance, high breakdown voltage, very high gain (g_{fs} or G_m), minimum switching losses, and a low gate threshold voltage $V_{GS(th)}$.

In those applications requiring operation in the linear region, these characteristics are not ideal. Firstly, G_m is too high and, secondly, $V_{GS(th)}$ has a negative temperature coefficient ($\Delta V_{th}/\Delta T_J$), which makes it impossible to maintain constant drain current without negative feedback. Finally, and most dangerous of all, large switchmode MOSFETs exhibit a phenomenon known as "hot spotting" or current tunnelling.

In an ideal device, current density across the die is uniform and the resulting junction temperature profile is also uniform. In practice, however, minor imperfections such as small voids in the die attachment solder or non-homogeneous junction doping etc, cause non-uniform current distribution and variations in the junction temperature profile. Hot spotting is then triggered by the negative temperature coefficient of gate threshold voltage acting in a localized manner. As one part of the die heats up more than another, the local junction $V_{GS(th)}$ falls and G_m then causes that area to turn on harder, drawing even more current. Since G_m increases with current, the runaway process accelerates rapidly until failure occurs.

The development of the 500V APL501JN linear power MOSFET and its 1kV APL1001JN sibling was stimulated by the expressed desire of linear MOSFET customers to have greater usable power dissipation. Understanding the nature of hotspotting led to innovative re-engineering to lower the V_{GS} versus I_C temperature coefficient crossover point. In these new devices, the source structure has been modified to produce an asymmetric channel, that is one where some parts of the interdigitated single channel are longer than others. The effect of this is to reduce current density in the channel region which in turn lowers the tempco crossover point. The onset of forward biased thermal runaway is retarded, enabling up to 50% more usable power dissipation in the linear mode.

One such application that has benefited from this new technology is the TDI Dynaload range of water-cooled dynamic loads. Here, individual 12kW modules are paralleled together to produce integrated and very compact power systems capable of dissipating up to 120,000 Watts, for such end uses as power supply testing.

Introduction

It is not appreciated in the all-pervasive switchmode world that power MOSFETs cannot and do not meet their published Forward Biased Safe Operating Area (FBSOA) ratings, because the average designer working in this discipline does not care. Such is not the case, however, among those engineers constrained to use these devices in the “linear mode”, where significant and concurrent values of drain voltage and current produce very high internal power dissipation.

FBSOA Testing

There is no easy way to test for FBSOA, because it is by nature a destructive test. Furthermore, there are significant variations in capability from unit to unit within a production lot and even more between die lots. Statistical analysis of the results will yield both typical and limit values for the part. For commercial production, a screening test is used which tests to a minimum level.

The SOA test fixture uses a water-cooled heat sink to keep the case close to 25C. A test value of drain voltage is selected, typically 50 to 80 percent of the rated drain voltage. The drain current is sensed by a sense-resistor in series with the source. An op amp, with its non inverting input tied to a reference voltage, establishes the device drain current as a function of the sense-resistor voltage.. The test procedure is to gradually increase the drain current in steps until the DUT fails. Each of the steps is held for a minute to ensure that the junction temperature reaches equilibrium. The power dissipation at which failure occurs is recorded for each device, and this defines the DC Safe Operating Area.

If the current is pulsed and the pulse width kept less than the thermal time constant of the device, the failure point can be several times the rated power dissipation of the device. This demonstrates that the failure mechanism is hot spotting. Under pulsed conditions, the junction temperature is more uniform so that the "hot spot" does not have time to develop.

Forward biased safe operating area for a semiconductor device (FBSOA or just SOA for linear devices) is a triangular area bounded on the left side by a positive sloped line defined by the Rds(on) of the device, and on the right side by a negative slope line determined by the power dissipation of the device. A typical SOA curve is shown in Figure 1.

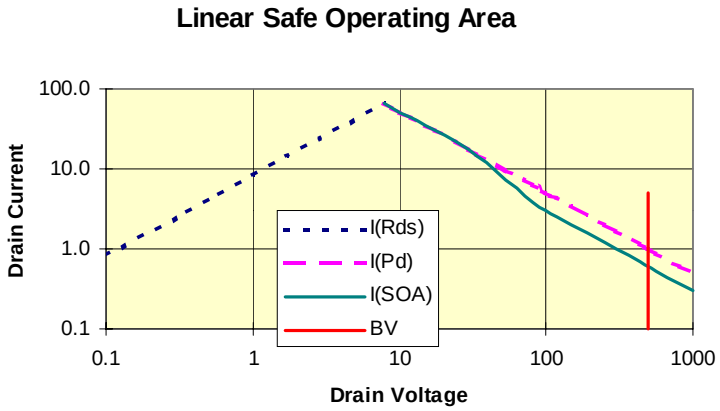


Figure 1 Typical FBSOA Curve

This figure appears on virtually every power device data sheet and, athough the $R_{\Theta JC}$ and $R_{DS(on)}$ parameters on which it is based are carefully measured and controlled by the manufacturer, the figure itself is derived by calculation and is seldom if ever rigorously validated.

In the case of switchmode devices, FBSOA is of little importance. The device is operated between two points on the figure, one on the Rds line at some current and one off the chart at some voltage and zero current. Power dissipation in a switched application is the sum of the on loss, $I_D^2 \times R_{DS(on)}$,and the dynamic switching losses which occur while making the transition between the two operating points. The characteristics which concern a switchmode designer are the $R_{DS(on)}$ characteristic and the parameters which determine the dynamic losses. Linear operation is different. The device is operated in a non-saturated mode, away from the $R_{DS(on)}$ line and somewhere within the SOA. In this mode, it behaves as a variable resistance, which is is exactly its role as a programmable active load.

Circuit for an active load

Linear operation of a MOSFET cannot be reliably accomplished without negative feedback. Whether as a "programmable resistor" in an active load or as the control element in a linear series regulator, the device must always be used inside a control loop. Within such a loop, thermal variations of V_{th} and G_m can be accommodated.

As a design example, consider the active device requirements for a programmable load to be used for testing a regulated power supply. The power supply has a maximum output of 3.5A at 300V. The active load will therefore need a device with a BV_{DSS} of at least 300V and must be capable of dissipating the entire output of the supply plus a safety factor to permit testing of the over current protection feature.

Devices suitable for the voltage requirement are fairly common, but a single device with more than 500W of usable power dissipation is not. The question is then to decide how many devices must be paralleled.

The APL501J device has a rated FBSOA of 300W and a rated power dissipation P_D of 520W, but as already explained this latter is not really usable in linear operation. The critical parameter defining allowable power capacity is $R_{\Theta JC}$. For an APL501J, $R_{\Theta JC}$ is 0.24K/W and $R_{\theta CS}$ 0.6K/W, so if the desired operating T_J is 130°C and the sink is water cooled at 30°C, the allowable (switchmode) power dissipation is (130-30)/0.3 = 330W. However, since this exceeds the 300W FBSOA rating of the device, the 300W max FBSOA limit must prevail. Applying a 20% safety factor reduces this to 240W.

If the power supply being tested is capable of delivering 3.5A at 300V, the MOSFET load-bank must absorb 1050W. 1050W/240W = 4.4, so five devices in parallel are required, assuming that five devices can be configured to share the load equally.

Paralleling devices

No matter how carefully devices are matched for one given characteristic, G_m for example, differences in other parameters and variations in mechanical assembly will generally conspire to unbalance paralleled devices. Some form of feedback control is mandatory to assure current sharing, and the easiest way to do this is to install source resistors to monitor the drain current of each MOSFET. Resistor tolerances then determine the relative matching between the MOSFETs. The resulting resistor voltages are applied to the inverting inputs of operational amplifiers driving the gates, the non-inverting inputs being tied to a current-dependent reference voltage.

The value of the source resistor depends on the dynamic range of adjustment required, the noise content in the output, the minimum load resistance and the stability requirements of the system. A resistor outputting 1 to 2 volts maximum is ideal. The temperature coefficient of the resistors will determine the temperature stability of the system. Since the feedback is unique to each device, any number of MOSFETs may be paralleled without risk. The same current-dependent reference voltage can be applied to all the opamps. A typical arrangement is shown in Figure 2.

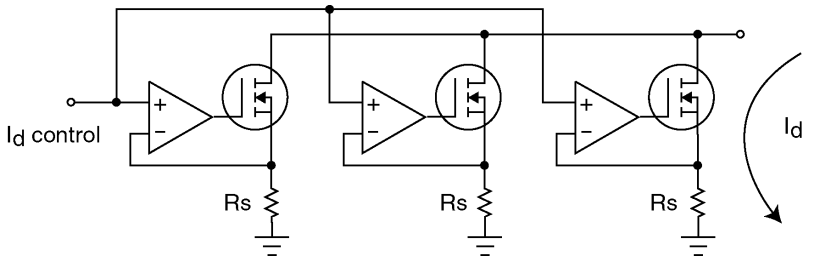


Figure 2- Paralleling several MOSFETS

Device changes to improve SOA

The development of the 500V APL501JN along with its 1kV APT1001JN stable-mate was stimulated by the expressed desire of potential users to have substantially increased allowable power dissipation in the linear mode. As already postulated, the fundamental mechanism behind MOSFET FBSOA failures is localized thermal runaway caused by “hot spotting”. It results from the interaction between non-uniform heating over the surface of the die and the negative temperature coefficient of gate threshold voltage as illustrated in Figure 3.

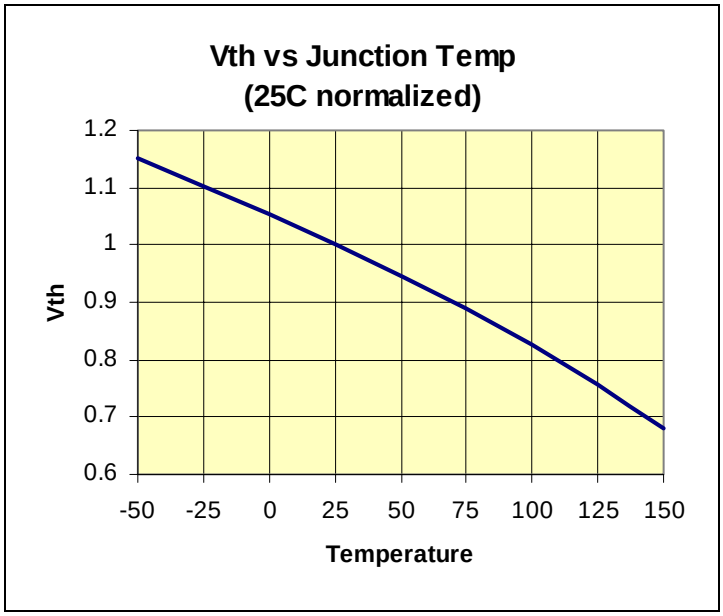


Figure 3

The phenomenon can be viewed as a classic example of an unstable feedback loop. From a device standpoint, only three variables can be manipulated to improve the stability of the system: $R_{\Theta JC}$, G_m, and the tempco of V_{th}.

- In the case of the APL501 and APL1001, $R_{\Theta JC}$ is a function of die size and the package, so is more or less fixed. In any event, since a bigger die would yield a lower $R_{\Theta JC}$ and a higher G_m, the net loop gain would remain unchanged.
- Within limits, the very high G_m of a switchmode device can be reduced considerably before difficulties start to appear in linear applications. Common MOSFET design rules indicate that G_m is generally inversely proportional to $R_{DS(on)}$ and proportional to channel width and die area.
- The primary means for improving SOA of the new design was to reduce its crossover or turnover current. Crossover current of the original switchmode part was 30A, this being where the transfer characteristic curves cross in Figure 4, or the current at which the slope is zero in Figure 5. By modifying the gate-source geometry using a fabrication process dubbed "asymmetric channel", the crossover current was reduced by half, with commensurate improvements in linear-mode FBSOA and usable power dissipation.

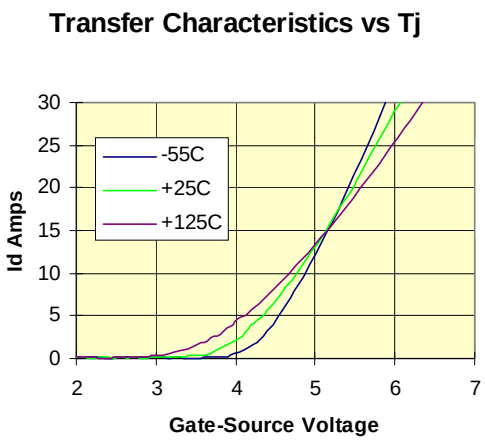


Figure 4

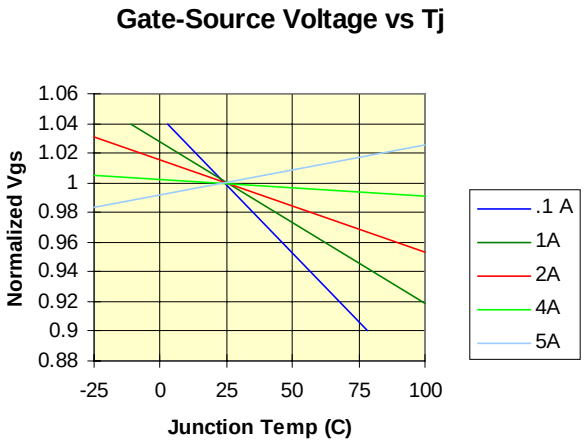


Figure 5

As can be seen from Table 1, crossover performance differs markedly among nominally similar devices - all having chip areas of about 128,000 sq mils. Note that the later high performance switchmode designs- characterized by higher speed, lower gate charge, tighter gate/source pitch, lower on-state resistance etc, all display progressively worse FBSOA performance as measured by crossover current. As a result, the new optimized linear design was based on the older MOS IV process, rather than on the more recent MOS V or MOS VI developments.

Device	Crossover Current	Device Process
APL501J	15A	New Linear Process
APT5010JN	30A	MOS IV®
APT5010JVR	62A	MOS V®
APT5010LLC	100A	MOS VI®

Table 1 Performance versus Process

A schematic cross section of the APL501J chip is depicted in Fig 6. Unlike the vast majority of competitive power MOSFETs that feature multi-cellular chip structures, APT devices are manufactured with a single-cell interdigitated gate and source structure. A proprietary process change results in the formation of a shorter channel length along one side of the structure. The net effect is the creation of two distributed parallel MOSFETs with different transfer characteristics. The portion of the die with the shorter channel length has a slightly lower gate threshold voltage (V_{th}) and a lower zero temperature coefficient point than the portion with the longer channel. At low drain currents, the low V_{th} portion of the die dominates the current flow. Since this portion of the die is operating closer to the zero temperature coefficient point, hot spotting and subsequent thermal runaway is less likely to occur. At higher currents, the current sharing between the two portions of channel becomes more uniform and operation approaches that of a device with a normal symmetrical channel structure. In this situation, the usable FBSOA is 50% higher than that of the original APT PowerMOS IV® switchmode device.

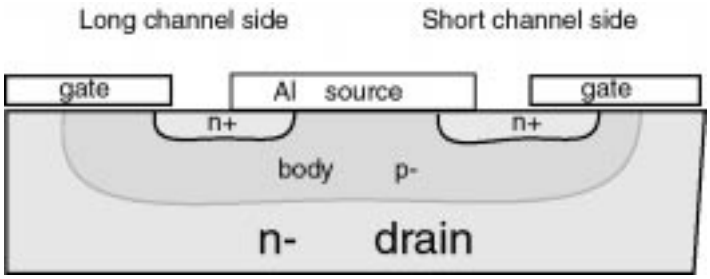


Figure 6 – Asymmetric Source

Conclusions

Incorporated into TDI-Dynaload's WCL series of 12kW water-cooled electronic load modules, which may be stacked together to dissipate up to 120kW at 400V or 10,000A per system, APT's new APL501J linear MOSFETs operate safely and consistently at power levels in excess of 300W at 400V. In comparison, SOT227 packaged

switchmode devices from other manufacturers are limited to about 100W at 400V. This vast improvement in FBSOA has allowed TDI to develop the highest power density electronic loads available today. Although the APL501J does exhibit the power roll off with applied drain voltage characteristic of all switchmode devices, this can be put to good advantage in the new device, in that the SOA curve shows a dramatic increase in power capability at lower voltages. At voltages below 100V, Dynaload has demonstrated that the APL501J is capable of dissipating up to 500W without failure.

Affiliations

Richard Frey, Senior Applications Engineer, Advanced Power Technology Inc., Bend, USA-97702 Oregon
Denis R. Grafham, European Applications Engineer, Advanced Power Technology, F-24300 Hautefaye
Tom Mackewicz, Senior Design Engineer, TDI-Dynaload Inc., Randolph, USA-07869 New Jersey

405 S.W. Columbia Street
Bend, Oregon 97702 USA
Phone: (541) 382-8028
Fax: (541) 330-0694

Chemin de Magret
33700 Merignac, France
Phone: 33-557 92 15 15
Fax: 33-556 47 97 61

http://www.advancedpower.com

Printed - June 2000

SOLID STATE OPTRONICS

Application Note 050
Failure Mode of SCR-Output Relays vs. MOSFET-Output Relays at High Temperature

Introduction

High temperature is a common environmental stress in many solid state relay applications. Application engineers must have a solid understanding of how a relay can be expected to fail if the temperature exceeds the rating of the package. This paper aims to compare the theoretical and experimental failure modes of SCR-output relays vs. MOSFET-output relays at high temperature.

Let us first define failure for an SSR as a condition in which we lose control of the output. For example, if we want to turn the output off and are unable to do so, that will be considered a failure. Similarly, if we try to turn the output on and we are stuck in an open condition, that is a failure.

Solid state relays are typically rated for an operating temperature in the neighborhood of 85º C. This nominal figure depends on the maximum junction temperature of the output structure. We consider high temperature to be any temperature above 85º C.

This report is broken up into two parts: 1) theoretical discussion of SCR and MOSFET at high temperature; 2) experimental results supporting theory.

Theoretical Discussion

SCRs at High Temperature

An SCR is simply a p-n-p-n structure with a gate terminal (Fig.1). We can break the structure down as back-to-back transistors, one p-n-p, the other n-p-n. With that simplification, we can see that temperature analysis of the bipolar transistor extends logically to the SCR structure.

It is a well known empirical fact that leakage current approximately doubles with every 10° C increase in temperature.[1] In a bipolar transistor, this increase in leakage is accentuated by the "transistor action" of the device. This can be explained by using an n-p-n transistor as an example. As we increase the temperature, more and more electrons are able to jump the barrier from the emitter to the base. This further biases the base region with respect to the emitter and collector, causing an increase in collector current. In fact, a transistor can be turned on simply by applying high temperature - sufficient leakage current can be generated to trigger the transistor action.

This discussion extends to the SCR, which is nothing more than two bipolar transistors driving each other. Any effect felt by the bipolar transistor is only magnified when discussing the SCR. The effect is not additive, it is multiplicative.

There is another temperature-related phenomenon we must point out: as we increase temperature, diode voltage decreases at an approximate rate of 2 mV/°C.[2] Therefore, a transistor in the on state will have a tendency to not only stay on at high temperature, but to conduct even more fully; i.e., the barrier between p- and n-type regions is reduced even more.

The result of these two phenomena is that the bipolar transistor has a negative temperature coefficient; the higher the temperature, the higher the collector current at a given base drive.

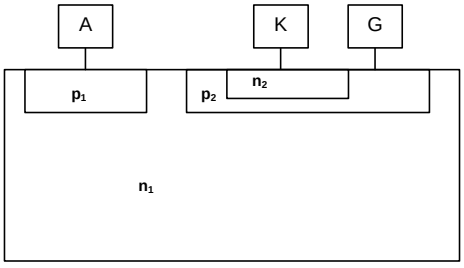


Fig. 1: SCR structure

Let's now extend the discussion to the SCR, specifically in Solid State Optronics device S742. The S742 uses two SCRs in the output; thus, it can only be utilized in AC applications. This is because the only way to prevent these particular SCRs from conducting once they are turned on is to reverse the voltage across their terminals. This is dictated by the output of the application. In a DC application, once the SCR is turned on, there is no way of turning it off. Under DC, the SCR never experiences the reverse voltage condition across its terminals necessary to prevent conduction.

An SCR in the off state will tend to turn on and stay on (latch) at high temperatures. Of course, in the S742 one SCR will always be non-conducting because of the reverse voltage on the output. But the other will tend to turn on even without an input signal because of the above considerations.

MOSFETs at High Temperature

Under no gate bias, a MOSFET can be thought of as a pair of back-to-back p-n diodes (Fig. 2), from source-to-bulk and bulk-to-drain. Again, we point to the empirical fact that leakage current approximately doubles with every 10° C increase in temperature. However, as we increase temperature, we can think of the leakage current from each diode cancelling each other out, resulting in no net change in current.

The dominant temperature-related mechanism for the MOSFET is the reduction in carrier mobility. This reduction exists because of the increase of scattering a carrier experiences due to the increased excitation

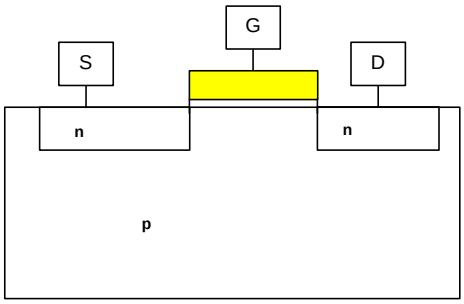


Fig. 2: MOSFET structure

of lattice sites in silicon at high temperature. The mobility of carriers in the channel of a MOSFET behaves according to the following empirical expression[3]:

$\mu(T) = \mu(300\ K)/(T/300)^a$- where T is measured in kelvins and *a* is 1.0-1.5

This relation shows that mobility may decrease as much as 40% for a 100° C temperature increase. Decreased mobility leads directly to an increase in on-resistance. Thus, we can say that the MOSFET's on-resistance has a positive temperature coefficient.

On the other hand, V_T, the threshold voltage of a MOSFET, typically decreases by approximately 1.5 mV/°C.[4] The reason for this effect is that at high temperature, we can expect an increase in thermally generated carriers (leakage). With more n-type carriers available in the conducting channel, we need less of a gate voltage to achieve the same amount of conduction.

This effect would tend to counter the decrease in mobility, however, the effect is relatively weak. V_T decreases by approximately 0.15 V for a 100° C increase in temperature - this is relatively insignificant. We would expect the reduction in mobility to be the dominant mechanism at high temperature.

The reduction in V_T is further offset by the reduction in the voltage of the gate driver circuit. In SSO's MOSFET-output devices, the driver consists of 14 series diodes. These diodes generate sufficient voltage to drive the gate of the output MOSFET, allowing conduction. As mentioned above in the SCR discussion, diode voltage drops at higher temperature. This in turn reduces the drive voltage to the gate of the MOSFET. Empirical results have shown that the drive voltage is reduced by approximately the same amount as the reduction in V_T, thereby virtually cancelling the effect.

Now, let's look at the potential failure modes. A MOSFET in the off state will not turn on when exposed only to high temperature. This is because high temperature alone will not be able to create the inversion layer beneath the gate necessary for conduction. A gate voltage is required to do this. Furthermore, a MOSFET in the on state will tend to conduct less and ultimately shut off at high temperatures due to the reduction in carrier mobility.

Experimental Results

In order to provide support for the theoretical conclusions, we ran a set of experiments with an SSO SCR-output device (SSO P/N S742) and an SSO MOSFET-output device (SSO P/N AD8C111) at high temperature. Please see attached documented procedure for details. In short, the test procedure consisted of 4 parts:

1) Take 20 S742s and put in oven cycling ON/OFF (10 HZ input square wave, 100 mA load); raise temperature until device fails and note failure mode.

2) Take 20 S742s and put in oven OFF (blocking 250 V_{RMS}); raise temperature until device fails and note failure mode.

3) Take 20 AD8C111s and put in oven cycling ON/OFF (10 HZ input square wave, 100 mA load); raise temperature until device fails and note failure mode.

4) Take 20 AD8C111s and put in oven OFF (blocking 275 V_{RMS}); raise temperature until device fails and note failure mode.

The following is a summary of the results of these experiments (please see attached documentation for further detail):

1) 18 of the 20 relays failed short; 1 failed open. The final relay didn't fail up to 145º C, at which point the oven's fuse mechanism prevented a further increase in temperature.

2) 18 of the 20 relays failed short. The other 2 relays didn't fail up to 175º C, at which point the oven's fuse mechanism prevented a further increase in temperature.

3) All 20 relays failed open - 11 of the 20 were "partially" conducting; the other 9 were completely OFF. (This result clearly shows the reduction in carrier mobility).

4) 13 of the 20 relays failed open. The other 7 didn't fail up to 150º C, at which point the oven's fuse mechanism prevented a further increase in temperature.

Conclusion

These results support theoretical expectations; namely, SCR-output devices can be expected to fail short at high temperature, while MOSFET-output devices can be expected to fail open.

REFERENCES

[1] Yang, "Fundamentals of Semiconductor Devices", P. 105
[2] Yang, P. 105
[3] Hodges/Jackson, "Analysis and Design of Digital Integrated Circuits", P. 48
[4] Supertex Databook 1996, P. 3-1

Elliott Sound Products **Project 51**

Balanced Line Driver & Receiver

Rod Elliott (ESP)

PCBs are available for the updated version of this project (P87). Click the image for details.

Introduction

Sometimes, you just can't get rid of that %$#*& hum, no matter what you do. Especially with long interconnects (such as to a powered sub-woofer), earth loops can be a real pain. For this reason, just about all professional equipment uses balanced lines, which, if properly executed, will eliminate the hum completely.

With this simple project, you can have balanced lines too, simply adapting the unbalanced inputs and outputs of your hi-fi gear to become balanced, and then back to unbalanced at the other end. You can even be extra cunning, and power the remote converter from the cables carrying the signal. Professionally, this is called "Phantom Feed", and is used to power microphones and other low current equipment. The version I have shown is actually a differential feed. Whilst not as good as a true 48V phantom powering circuit, it does work, and makes an interesting experiment (if nothing else).

Description

Before we start, a brief description of the standard (unbalanced) and balanced line is in order. An unbalanced line is the type you have on the hi-fi, typically using an RCA connector, and feeding the signal through a coaxial cable. The inner cable carries the signal, and the outer shield is a screen, to prevent RF interference and general airborne noise from being picked up on the signal lead.

This is fine, except for one small detail - the shield must also carry the signal! This is the return path, and is required in all electrical connections - otherwise there is no current flow and the system will probably just hum softly (or loudly) with none of the wanted signal.

The problem with electricity (like water and most people) is that it always takes the path of least resistance, so when two pieces of equipment are connected, most likely there will be signal plus hum, because of the dreaded earth loop. This is formed when both items are connected to the mains earth, and also have their earth (zero Volt) points joined via the shields of the signal leads.

In some cases it is possible to disconnect the earth at one end of the cable - some people have also disconnected the mains (safety) earth. Both achieve the same result, but disconnecting the mains earth is extremely dangerous. Unfortunately, the result is not always as one would hope. RF interference can become much worse, and other noises become apparent that were absent before.

In contrast, a balanced connection uses two wires for the signal (much like the telephone circuit), with the signal equal in amplitude in each wire, but opposite in phase. Only the out of phase signal is detected by the remote balanced receiver, and any in phase (common mode) signal is rejected. RF interference and other noise will be picked up equally by both wires in the cable and so will be in phase. It will therefore be rejected by the receiver. In this way, it is possible to have long interconnects, with the shield connected at one end only. This cuts the earth loop, and the balanced connection ensures that only the wanted signal is passed through to the amplifier(s).

It is very important that the two signal leads are twisted together, and the tighter the twist, the better. The shield prevents RF and other interfering signals from causing too much trouble, and the final signal should be free from hum and noise. The shield serves the same function in an unbalanced circuit, but is less effective due to the fact that it usually serves as the signal return path, and any signal that does get through becomes part of the signal.

The idea of this project is to give you some options, and to assist in creating a solution - it should not be seen as a complete solution in itself. There are many variables - far too many to be able to say with complete confidence that this WILL prevent all hum and other interference. It might, but it is likely that some experimentation will be needed to get the results you want.

Note that for both transmitter and receiver, it is essential that 1% (or better) tolerance resistors are used. If the trimming option is implemented, then you could use 5% resistors, and you will be able to adjust the circuit to get maximum common mode rejection - however I recommend that you use the 1% metal film resistors. For the small extra cost you get much higher stability, and lower noise.

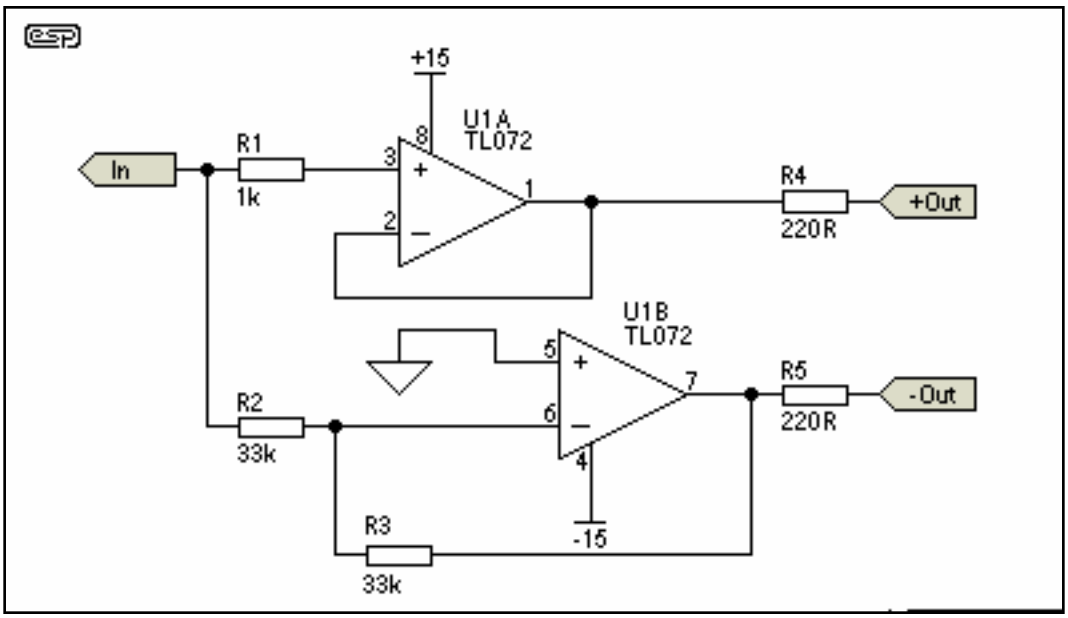


Figure 1 - Balanced Line Transmitter

The transmitter uses one opamp to buffer the signal, and the other to buffer and invert it. This creates a balanced signal, where as the signal swings positive on one lead, it swings exactly the same amount negative on the other. The 220 Ohm resistors at the output ensure stability with any lead, and are also used to attenuate the signal slightly. The signal swing from the transmitter (across both wires) is double the voltage of the input signal.

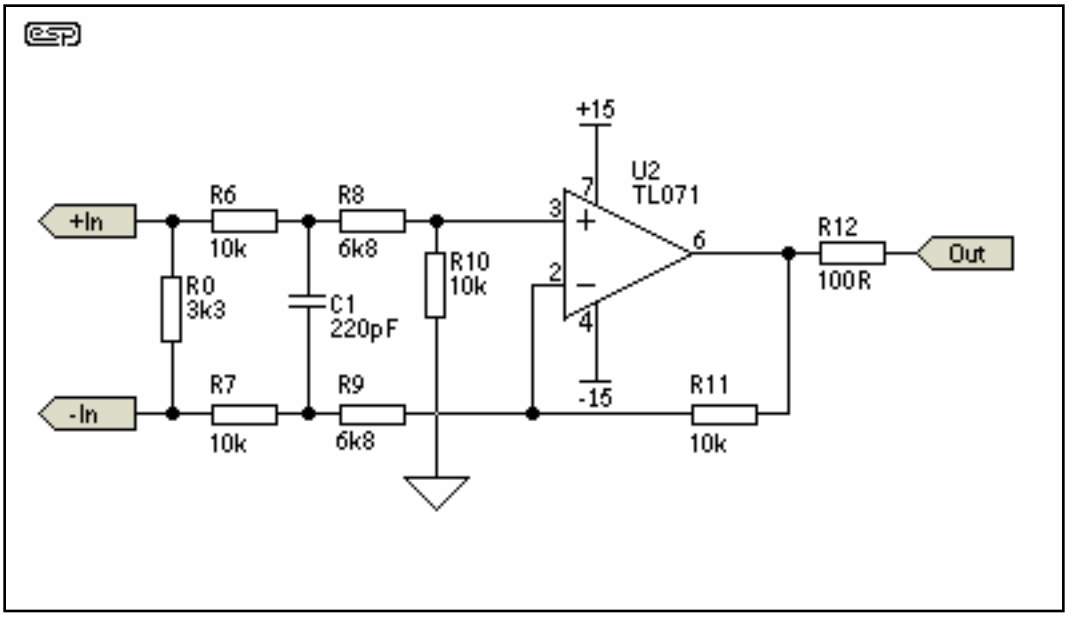


Figure 2 - Balanced Line Receiver

The receiver has an optional 3.3k resistor across the inputs (RO) to help balance the input against minor variations in cable impedance between the individual lines. The 220pF capacitor is for HF rolloff, and will attenuate any RF that might get picked up by the lead. Any common mode signal - where both leads provide a signal of the same polarity to the receiver circuit; typically noise - is rejected, leaving only the wanted signal.

The rest of the circuit is a conventional balanced input stage. This particular configuration is somewhat notorious for having unequal input impedances referred to earth. The 3.3k resistor helps this (a little, anyway), and the 220pF capacitor also assists at higher frequencies. A more complex circuit could have been used, but that would require 3 opamps, and for the intended task would offer few real advantages.

With the capacitor value chosen, there is about 0.1dB attenuation at 20kHz - if you don't like this idea, reduce the value to 100pF, however since 0.1dB is quite inaudible, there seems little point.

With the values shown, there is a very slight overall gain of just over 0.3dB. This is unlikely to be a problem. The circuit is designed to send the maximum level possible across the balanced cable, and most of the attenuation is performed at the receiver. This will reduce any noise picked up by a further 6dB for the transmitter / receiver pair.

It is also possible to ensure that the common mode rejection is as good as it can possibly get, by making R10 variable. I suggest that you use an 8.2k fixed resistor, with a 5k multi-turn trimpot in series. To balance the circuit, you may use an oscillator and millivoltmeter, or just a small battery and a multimeter.

Join the two inputs together, and connect the battery or audio oscillator between the two joined inputs and earth. Adjust the trimpot until there is 0V at the output - the common mode signal is now gone completely. Typically, this circuit will give a common mode rejection of about 40dB if not trimmed as described, but trimming will let you improve on this considerably.

Although this transmitter and receiver pair will probably allow the use of unshielded interconnects, I don't recommend this. Use a good quality shielded twin microphone cable. The earthing of the shield should normally be done at the receiver end, but in some cases you might find that the noise rejection is better if the transmitter end is earthed. Experimentation will be needed.

Phantom Power (For the Experimenter)

It is possible to run this unit with the signal leads also carrying the power for the receiver. We could use conventional phantom feed (using a 48V supply), but it is easier to use a differential feed, with the +ve and -ve supply voltages on the signal leads. The basic scheme is shown in Figure 3. This may be found to reduce common mode rejection, and it is essential that the power is completely noise free, or it will become part of the signal! If this method is to be tried, use the trimming option, so the supply feed resistors can be catered for. Alignment with a battery will no longer be possible, and a signal generator will have to be used - with coupling capacitors to each signal line.

The resistor RO must be removed in this configuration. I would strongly recommend that an output coupling capacitor is used from the Out terminal of the receiver, since it is likely that there will be some DC offset due to capacitor leakage currents.

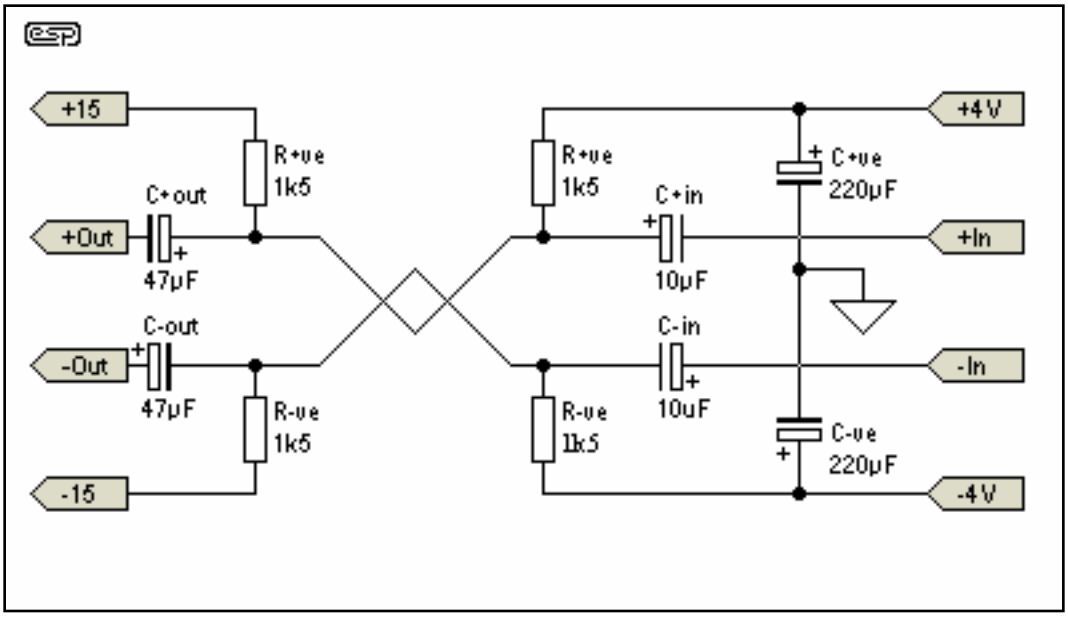


Figure 3 - Differential "Phantom" Powering

The voltage to the receiver opamp is reduced by this technique, and the maximum signal level will be reduced too. Only by experimenting will you be able to determine the exact power losses and maximum signal level attainable. The tests I did indicate that you should not expect more than about 1V RMS, but you might get more depending on the opamp used for the receiver. The power feed resistors also load the transmitter, and reduce its output capability somewhat. You might want to experiment with a low-power opamp (such as an LF351) as the receiver, as this will allow a higher supply voltage and more signal before distortion.

I would expect that the most likely use for this arrangement would be for a remote sub-woofer, where it may be very inconvenient to have to create an additional power supply. I can't say that I am completely happy with this arrangement, but it does work. A 48V phantom supply would be better, but it is not likely that too many constructors will want to go to this trouble.

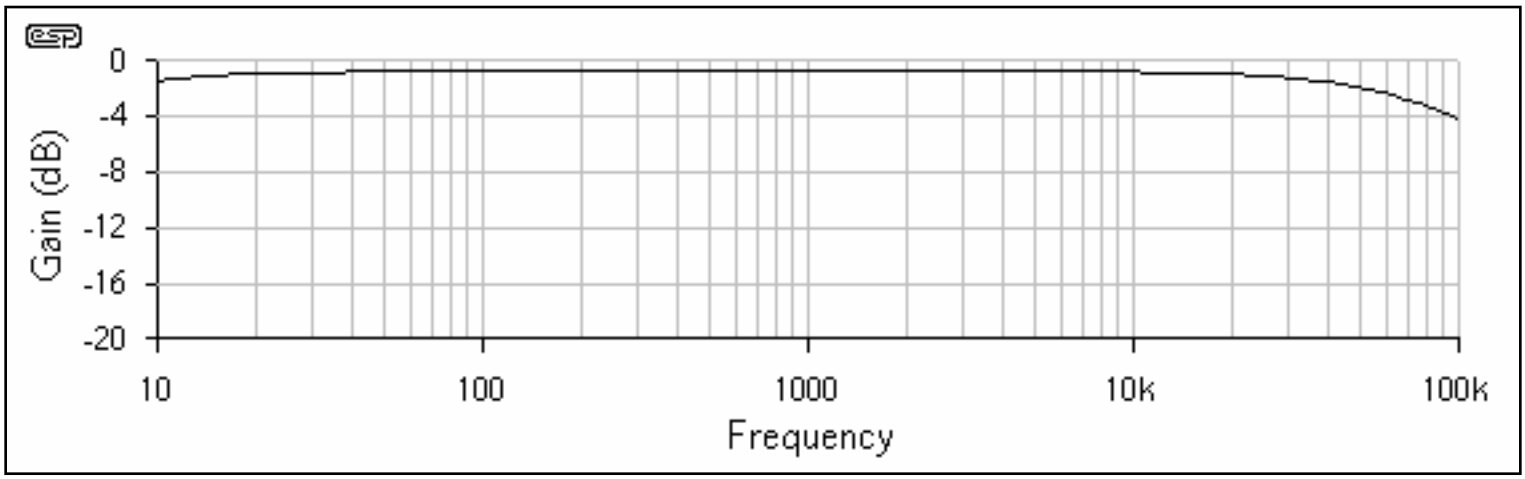


Figure 4 - Overall Frequency Response of Differential Feed and Both Circuits

The shield will now have to be connected at each end, but one end can be earthed using a 10 Ohm resistor, which should be bypassed with a 100nF capacitor. Again, experimentation is needed to determine which end should have the "hard" earth. Make sure that the connectors are polarised so that power cannot be connected the wrong way around. Diodes may be added if desired to provide proper protection. These should be in parallel with the receiver filter caps (C+ve and C-ve), because a series connection will reduce the voltage further (there is not a lot to start with, so a further reduction would be a disaster).

Use of a multi-cored cable and suitable connectors will allow you to run the power supply on separate wires in the cable, and the additional cost of the cable and connectors is likely to be offset by the simpler circuit and better performance. This may not always be possible, hence the differential phantom feed.

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright (c) 1999. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright (c) 05 Feb 2000

The Audio Pages

Elliott Sound Products **Project 87**

Balanced Transmitter and Receiver II

Rod Elliott (ESP) / Uwe Beis *

Updated 01 April 2002

PCBs are available for this project. Click the image for details.

Introduction

This is essentially an update to the original article on the subject, and includes some ideas to stimulate further thought on the subject. This is especially true of the last section (Hey! That's Cheating) - everyone wants balanced outputs free, well you can have them free (well, near enough anyway).

The balanced transmitter and receiver described in Project 51 work very well, but both are less than optimum under difficult conditions. Uwe has written an article (published on The Audio Pages) describing an active balanced transmitter that has performance almost equivalent to a transformer. There are ICs available that (almost) manage the same thing, and the principle uses feedback to equalise the levels from each transmitter opamp.

While Uwe has gone to a great deal of trouble to get his circuit to match a transformer as well as possible, this is not an easy circuit to get working well, and it requires 0.1% tolerance resistors and wide bandwidth opamps.

Enter ESP and the "simplification methodology" that I tend to use wherever I can. The result is a transmitter (in particular) that is extremely good, and matches the performance of a transformer to a quite acceptable degree. It's not perfect, but it is very stable, and requires no adjustments or close tolerance parts (1% resistors will provide a maximum error of 1/100, or 40dB common mode rejection and balance).

Use of closer tolerance resistors - and good or premium opamps - will give a circuit with excellent performance, and it will come very close to that of a transformer balanced circuit with none of the associated cost. Response is flat to at least 50kHz, with a low frequency limit of DC (as shown). Capacitors can be used to limit the low frequency limit if desired.

Mk II Receiver

The receiver is shown in Figure 1, and as shown does not have any RF protection. This configuration is somewhat better behaved than that shown in the original article, and presents exactly the same impedance to each of the balanced lines in the cable. This is also the case with the original version shown, but *only* if the source is balanced.

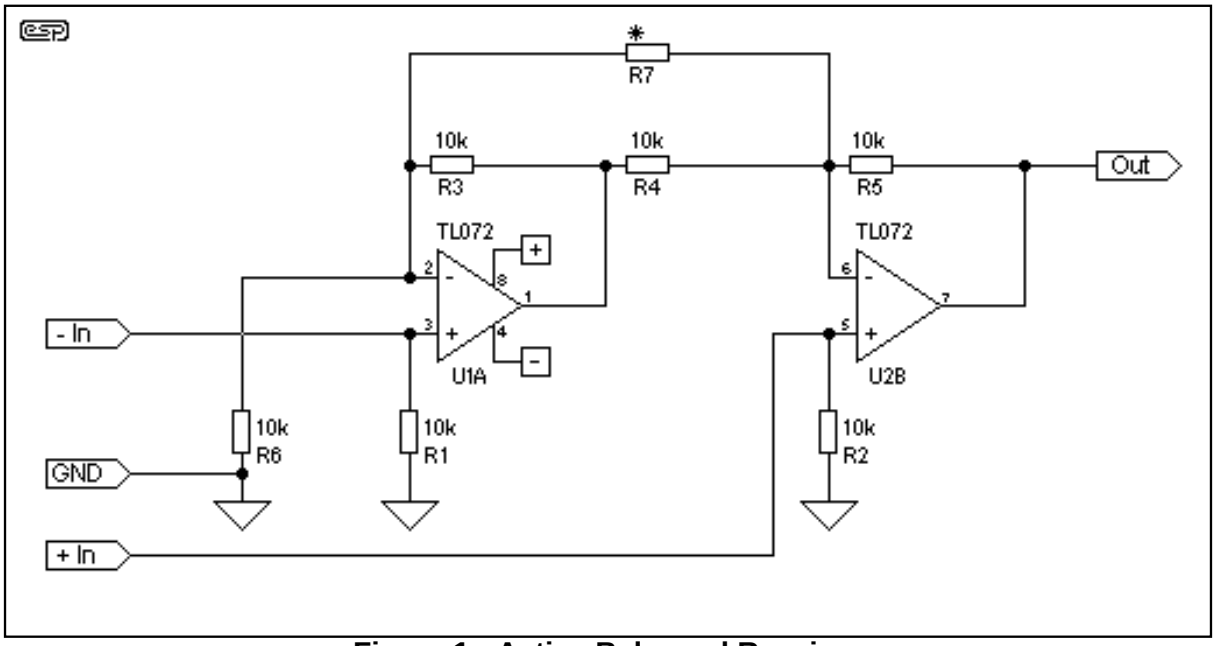
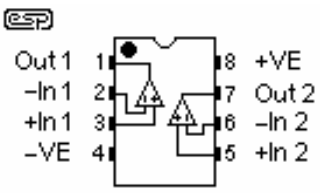


Figure 1 - Active Balanced Receiver

The resistor marked * (R7) may be left out, and the circuit will have a gain of 2. Installing this resistor will increase gain, but will have no effect on the input impedances or balance performance. The minimum gain for this circuit is 1.5 (if R6 and R7 are omitted), and this increases to 2 with R6 installed. The gain setting resistor R7 still works if R6 is omitted, with a value of 10k providing a gain of 3.5, and 1k giving a gain of 21 (26dB).



The standard pinouts for a dual opamp are shown (top view of device). It is suggested that a bypass capacitor (typically a 100nF ceramic or polyester) be connected between pins 4 and 8, as close to each opamp package as possible. Even with "slow" opamps, it is always a good idea to use a bypass cap to prevent possible instability at high frequencies.

The exact same scheme as shown in the original project could be used for the inputs on this version. One possible connection is shown in Figure 2. This is virtually identical to the configuration shown in Project 51, and will provide a very high noise rejection.

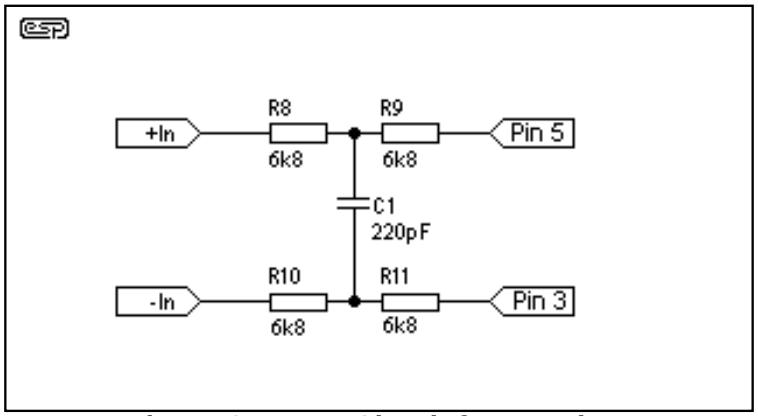


Figure 2 - Input Circuit for Receiver

The 10k resistors (R1 and R2) to ground are still needed, and this arrangement will reduce the gain by a little over 7db. Use lower value resistors for less attenuation, but remember that as the attenuation is reduced, so is noise immunity. 1k resistors would be the lowest value I'd recommend, and will cause minimum attenuation (about 1.6dB). C1 may be increased if desired, but if too high may cause rolloff of the signal source.

Mk II Transmitter

This transmitter is quite a bit more complex than the Project 51 version, but this is the price one pays for higher performance. The input is unbalanced, and has an input impedance of 11k. This must be driven from a low impedance source (such as an opamp's output) or performance will be degraded.

With the values shown, the circuit has a gain of 6dB when measured from the input, and between +Out and -Out. R13 and R14 are not absolutely essential in this version, but are recommended. They enforce a balance on the circuit, and prevent the possibility of "latch-up" where the outputs get stuck to a supply rail. This is extremely unlikely with the values shown, but the precaution is worth the very minor effort.

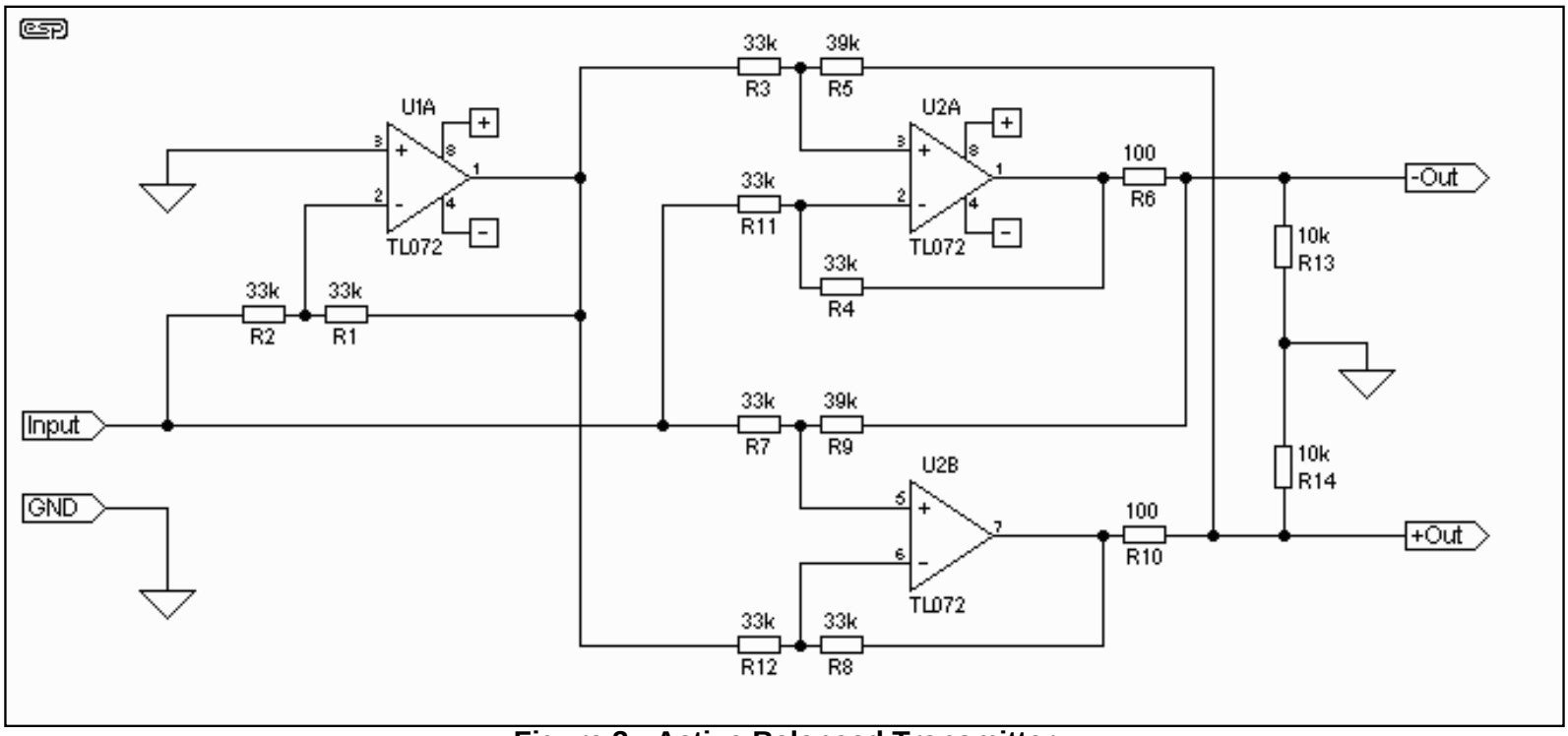


Figure 3 - Active Balanced Transmitter

If either output is shorted to ground by connection to an unbalanced input, the output voltage is only 0.4dB less than when operating in fully balanced mode. When one output is shorted, the feedback path to the other opamp is removed, so it provides (almost) the full swing that would normally be available between both opamps. This is the way a transformer (without centre tap) works, so the behaviour of this design is much closer to that of a transformer than the "standard" balanced output circuit.

1 Apr 2002 - I tested the circuit shown using 1458 dual opamps and 5% resistors. If the circuit is reliable and shows no bad habits with very basic opamps (basically dual 741s) and ordinary carbon resistors then I know that it will work when you use better components. Indeed, my test version is both stable and surprisingly accurate, despite the lowly parts used to test the circuit's operation. Because the crossed feedback paths are reduced from the optimum (by virtue of making R5 and R9 1.2 times the "correct" value), the overall stability and frequency response is much less dependent on the component values and opamp quality. I was able to verify that even using 741 type opamps, frequency response is less than 1dB down at 75kHz.

Somewhat surprisingly, output impedance is 100 ohms, and not the 200 ohms one would expect - there are 100 ohm resistors in series with each output after all. The crossed feedback paths make the difference here, and both simulation and measurement confirm that this is the case.

Construction Hints

Both the transmitter and receiver circuits require at least 1% tolerance resistors, or common mode rejection will be unacceptable. Even with 1% tolerance, the worst case rejection is only 40dB, and if you can use your multimeter to match the resistors to closer tolerance this will improve the performance.

Although the transmitter and receiver are shown with (mainly) 33k and 10k resistors respectively, these may be changed if desired. Any value between 10k and 100k could be used, but remember that higher value resistors create more thermal noise. R5 and R9 in the transmitter are approximately 1.2 times the other resistors - the next E12 value up. For example, if you elected to use 22k resistors throughout, then R5 and R9 would be 27k. Also remember that for the transmitter's input, the impedance is 1/3 of the resistor value used - 10k resistors would therefore give an input impedance of about 3k.

Both circuits require a balanced +/-12V or +/-15V supply (Project 05 or similar power supply), and it must be free of noise. Make sure that 100nF ceramic caps are placed between the supplies as close as possible to the supply pins of the opamps. This is especially important if you use high speed opamps.

Hey! That's Cheating

Finally, for those who want a balanced output that is really simple, try the circuit shown in Figure 4. Now look at it again - it's not balanced at all ... or is it?

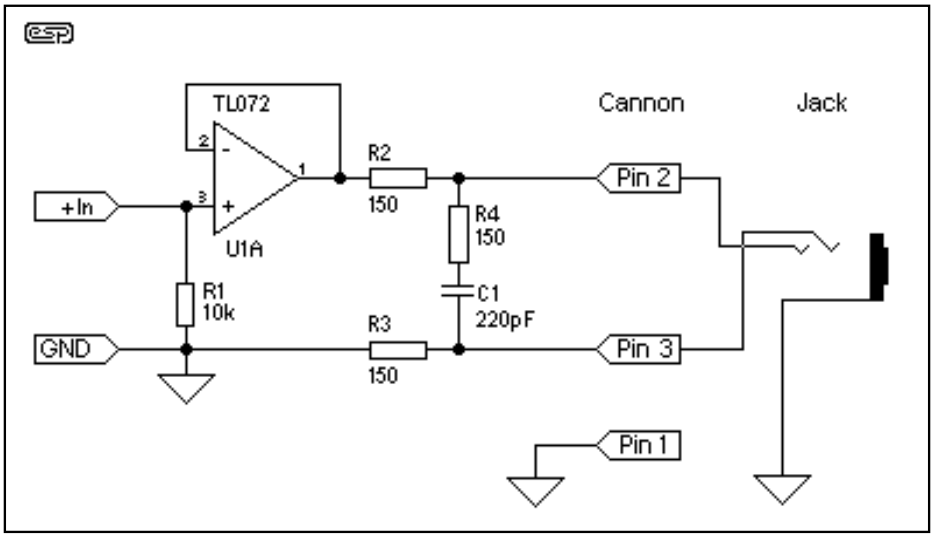


Figure 4 - Simplest Possible Balanced Output

Now, before you get all horrified, let's have a proper look at what is happening. The main trick with a balanced circuit is that the receiver should "see" the same impedance on each input. It doesn't actually care that much if there is signal on either or both wires (indeed, that is merely an *expectation* on our part), but even if the wanted signal is only on one wire, any induced noise will still be common mode, and will still be in phase across both wires. The noise gets cancelled either way, and the signal gets amplified, which is just what we want.

Yes, it's cheating - but it works. Apparently, this trick is used on some of the popular stage mixers, as well as some very well regarded phantom feed microphones (although as far as I know they don't use the Zobel network - this is optional BTW). There is less signal than one would expect (most balanced transmitter circuits have an effective gain of 6dB), but this is generally not an issue. In the case of a

microphone, the signal is the same as it would normally be anyway, and with a line output, 6dB of additional gain is usually not a problem to achieve. The amplifier as shown in Figure 4 only needs 2 x 10k resistors in the feedback path to achieve this (10k from output to -ve input (i.e. -in, *not* -ve supply pin), and 10k from -ve input to ground).

In most cases, this will work *just as well* as a true balanced output circuit. From the perspective of the balanced input circuit (the receiver), this arrangement provides exactly the same signal as if the circuit were fully balanced. Although 150 ohm resistors are shown for the balancing network, these can be changed if desired. Normally, I would expect that the values shown will be fine for almost all applications (effective output impedance is 300 ohms).

Somewhat predictably, the signal is only on one lead, so the cables had better be wired correctly if it is feeding an unbalanced input (but this is something that should be regarded as essential for all stage and studio work anyway).

Caveats - the output impedance of the opamp should be flat to a suitable high frequency, and this will not often be the case. Although I have shown a TL072 (or half thereof), I suggest that a better opamp should be used. A very high speed opamp (Such as the Texas Instruments THS6012 - See The Need For Speed) will help enormously here, but at considerable cost and inconvenience - unfortunately. In general, the output opamp should be as good as you can afford. A cheap opamp will give a cheap result. The optional zobel network will help ensure that the line appears properly balanced at all frequencies including RF, but cannot guarantee perfect results with any opamp. Resistor tolerances are just as important here as with any of the more complex versions - 1% is the minimum acceptable tolerance.

Inspiration

This new project article was inspired by Uwe Beis, and his article on the (almost) perfect balanced transmitter is published on these pages. Although the material here is somewhat "off topic" from his approach, the inspiration to experiment and try the various techniques came from his submission. I recommend that you read the full article, as it explains the operation of the balanced transmitter shown in Figure 3 very well, and will give you an idea of the dedication of some people (Uwe in this case) to the advancement of their understanding of analogue electronics.

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright © Rod Elliott 28 Mar 2002./ Updated 1 Apr 2002 - tested new circuit as described in update.

Elliott Sound Products **Project 66**

Low Noise Balanced Microphone Preamp

By Phil Allison
(Edited by Rod Elliott - ESP)

PCBs are available for this project. Click image for details.

Introduction

This simple design has very low noise, close to the theoretical minimum, high hum rejection and variable gain with a single rotary pot. It is similar to that used in many professional grade mixing desks and can form the basis of a no compromise recording mixer for live work.

The design consists of differential compound pairs of transistors with a common mode (floating) gain control connecting the emitters of the pair. The compound pairs of 2N4403 and BC549s are far more linear than any single transistor. The circuit is differential in and out and therefore requires a balanced to unbalanced buffer to give suitable output for the next signal stages of a channel in a mixing desk. This is provided by a high performance op-amp differential gain stage, which can be a TL071 or similar IC of your choice. The stage has a gain of six or 15 dB and that sets the maximum input level at about 1.5 volts rms before clipping. This equals an SPL of over 150dB with a typical microphone!

Full gain is 1000 times or 60dB. Distortion is low to unmeasurable because it is below the noise level at high gains. The CMRR (Common Mode Rejection Ratio) is well over 60 dB and better than any available mic cable as far as hum rejection is concerned. The bandwidth extends beyond 100kHz, and no RF suppression is shown as it has proved unnecessary in practice. The input impedance or load on the mic is set by the two 3.3k ohm resistors. This will suit almost any mic with a nominal impedance of 150 to 600 ohms.

Description

The input stage is configured for least noise and this has meant a non IC approach. There are some special ICs that can be used for mic pre-amps, they contain a circuit like this one except fabricated on one chip. Examples include the SSM2017 (now obsolete) or the replacement INA217.

Components should all be readily available except for the 10 k ohm pot for the gain control. This needs to be a reverse log taper - or else use a multi-position switch with 6 dB gain steps covering the 60 dB range of the circuit. Make sure it is make before break.

Editor's Note - Alternatively, a standard log pot can be used, but wired "backwards". This will work fine if it is labelled "Attenuation" instead of "Gain". As the pot is advanced clockwise, the gain is reduced (attenuation is increased). Maximum gain will therefore be applied when the pot is fully anti-clockwise.

The +/-15 Volt power supply is important too, it must be regulated and low noise. If the usual voltage regulator ICs are used I recommend fitting a post filter consisting of a 10 ohm resistor and a 470 uF capacitor to remove any noise generated in the ICs (as shown in Figure 1). Some 7815 ICs could be sold as noise generators, the adjustable voltage ones (LM317, LM337) are very much quieter. A single regulator board may be used to power multiple preamps, with each preamp having its own post filter circuits. Because of the extensive filtering applied, the P05 (Rev-A) power supply is recommended for this preamp.

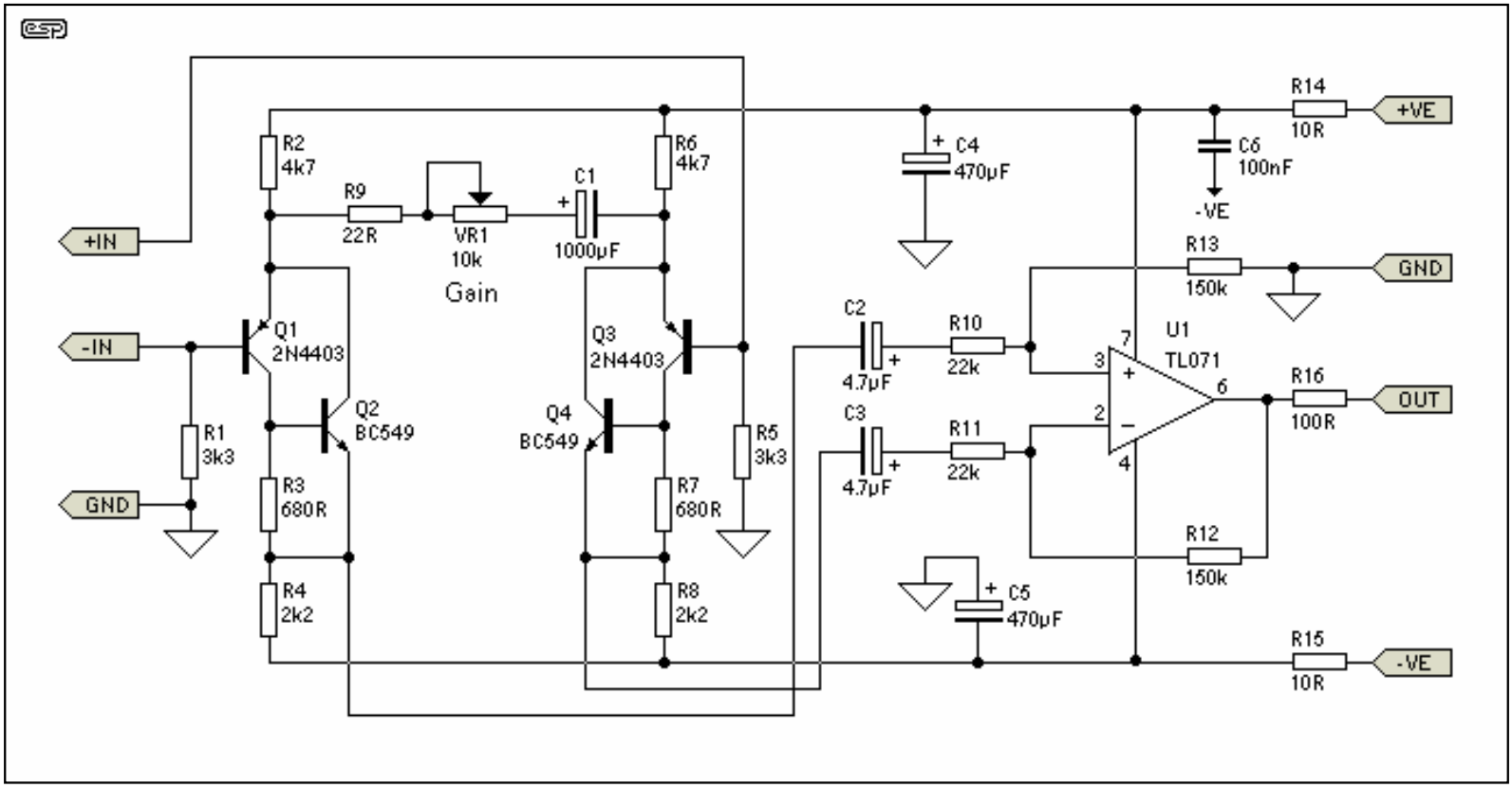


Figure 1 - Complete Microphone Preamp

Good quality components should be used with metal film resistors in the collectors and emitters of the input pairs for least noise. Where a resistor has significant DC voltage imposed on it in high gain circuits always use low noise types. Metal film resistors are about the best only bettered by wire wound which is a bit impractical. Avoid cermet, metal glaze, and very old carbon composition types. Also avoid bead tantalum capacitors, as they go leaky and crackle. They are just about the most fragile electronic components made. The 100nF capacitor (C6) should be mounted as close as possible to the opamp supply pins - a ceramic cap is recommended for best bypass performance at high frequencies.

The 1000uF capacitor can be a normal electrolytic of 10 or 16 volts rating. There is usually no problem

with zero DC bias on modern electros. All other electros should be 25V rating as a minimum.

Upon checking the published specs for the SSM2017 in regards to noise, my workshop version of the preamp measures at least as good with a 200 ohm source resistance (typical of most dynamic microphones).

> EIN = 0.27 uV rms in 20 kHz bandwidth with 200 ohm source.
> = 1.9 nV per root Hz (equal to spec for SSM 2017)
> Noise Figure = 0.9 dB rel 200 ohms.

Editor's Comment

I would suggest that 1% metal film resistors should be used throughout this circuit - the additional cost is negligible, and this will also ensure that the balanced buffer stage (U1) is properly balanced. Even a small error in the input and feedback components will degrade the common mode rejection.

Like Phil, I also recommend against the use of tantalum capacitors, and regular readers will notice that I have not suggested them for any project (although there was *one* suggestion that you could use them if you wanted to). The only capacitor fault I have ever had to track down with an *intermittent* short circuit was a tantalum bead type - it was neither fun, nor easy to find :-(

As with all circuits presented on these pages, feel free to experiment. The 2N4403 transistors may prove difficult for some readers to obtain, and BC559s can be substituted with some possible increase in noise. I would expect that any increase will be acceptable for most applications. Performance should otherwise be much the same as described.

The preamp is ideal for portable use, and can be operated from a pair of 9V batteries.

Note: PCBs are available for this preamp. There are a couple of very minor changes to the circuit, and the board is a dual preamplifier - two completely independent microphone preamps on one PCB. Included with the construction data (available when you purchase the PCB) is a switched gain control, which provides much more linear control than you will get from a pot.

As is my usual practice, the prototype was constructed from standard 5% carbon resistors. If I get good performance from this, then yours will be better if you use 1% metal film resistors - lower noise, and higher common mode rejection.

I also substituted some of the component values (only because I didn't have the correct ones immediately to hand), and the preamp worked flawlessly from the start. This means that you also have some flexibility, provided you understand the possible consequence of any substituted parts.

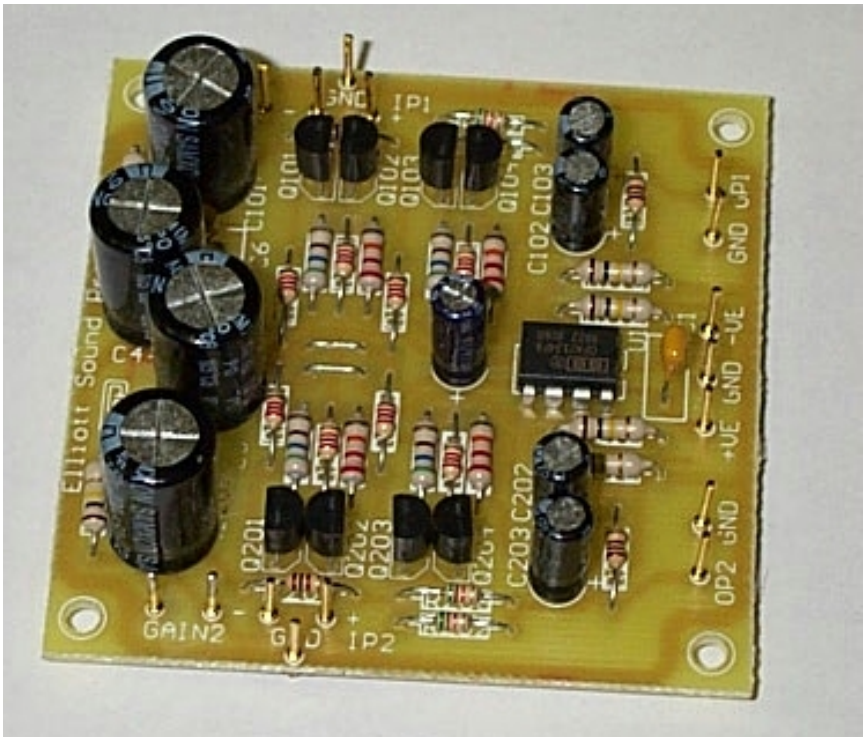


Figure 2 - Photo of the completed prototype PCB

In all, this preamp is highly recommended for professional or semi-professional use, wildlife recording or just experimenting. As you can see from the photo, the board is very compact, and I have described a phantom feed supply and distribution board elsewhere in the project section, along with a phantom powered microphone amplifier and a series of microphone projects.

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Phil Allison and Rod Elliott, and is Copyright (c) 2000. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Phil Allison) and editor (Rod Elliott) grant the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott and Phil Allison.

Page Created and Copyright (c) Rod Elliott 23 Aug 2000./ Updated 12 May 2001 - added noise measurements./ 30 Jun - PCB available

Not Recommended for New Designs

This product was manufactured for Maxim by an outside wafer foundry using a process that is no longer available. It is not recommended for new designs. The data sheet remains available for existing users.

A Maxim replacement or an industry second-source may be available. Please see the QuickView data sheet for this part or contact technical support for assistance.

For further information, contact Maxim's Applications Tech Support.

19-0042; Rev 1; 4/93

MAXIM

Wideband Transconductance Amplifiers

General Description

The MAX435 and MAX436 are high-speed, wideband transconductance amplifiers (WTAs) with true differential, high-impedance inputs. Their unique architecture provides accurate gain without negative feedback, eliminating closed-loop phase shift — a primary cause of circuit oscillation in conventional high-speed amplifiers. The output of the WTA is a current that is proportional to the applied differential input voltage, providing inherent short-circuit protection for the outputs. Circuit gain is set by the ratio of two impedances and an internally set current gain factor (K).

The current output and the absence of negative feedback allow the differential output MAX435 to achieve a bandwidth of 275MHz, an 800V/µs slew rate, and a 1% settling time of 18ns to a 0.5V step input. The single-ended output MAX436 achieves a bandwidth of 200MHz, an 850V/µs slew rate, and a 1% settling time of 18ns to a 0.5V step input. Both amplifiers offer exceptional wideband common-mode rejection, with a CMRR of 53dB at 10MHz. 300µV input offset voltage offers a level of DC precision rarely found in high-speed op amps.

Unlike current-feedback amplifiers, the MAX435/MAX436 have fully symmetrical, high-impedance inputs that tolerate wide differential input voltages without destructive failure or amplifier saturation, virtually eliminating overload recovery time. The unique performance features of these WTAs suit them for a wide variety of applications, such as high-speed instrumentation amplifiers and wideband, high-gain bandpass amplifiers. And, with its differential output, the MAX435 can be used in high-speed differential line driver and receiver applications.

Applications

High-Speed Instrumentation Amplifiers

High-Speed Filters

Wideband, High-Gain Bandpass Amplifiers

Differential Line Receivers

Differential Line Drivers

Features

- ♦ **275MHz Bandwidth (MAX435)**
- ♦ **850V/µs Slew Rate**
- ♦ **18ns Settling Time to 1%**
- ♦ **53dB CMRR at 10MHz**
- ♦ **Low Noise, 7nV/√Hz at 1kHz**
- ♦ **No Feedback**
- ♦ **True Differential High-Impedance Inputs**
- ♦ **Shutdown Capability: 450µA (MAX435)**

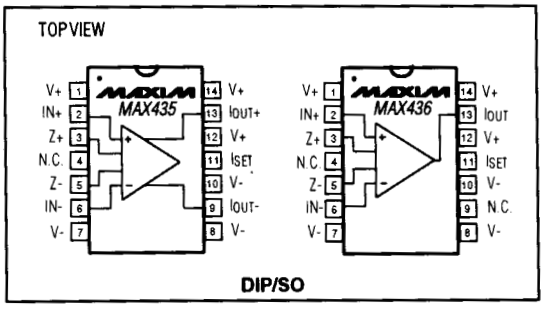
Ordering Information

PART	TEMP. RANGE	PIN-PACKAGE
MAX435CPD	0°C to +70°C	14 Plastic DIP
MAX435CSD	0°C to +70°C	14 SO
MAX435C/D	0°C to +70°C	Dice*
MAX435EPD	-40°C to +85°C	14 Plastic DIP
MAX435ESD	-40°C to +85°C	14 SO
MAX435MJD	-55°C to +125°C	14 CERDIP

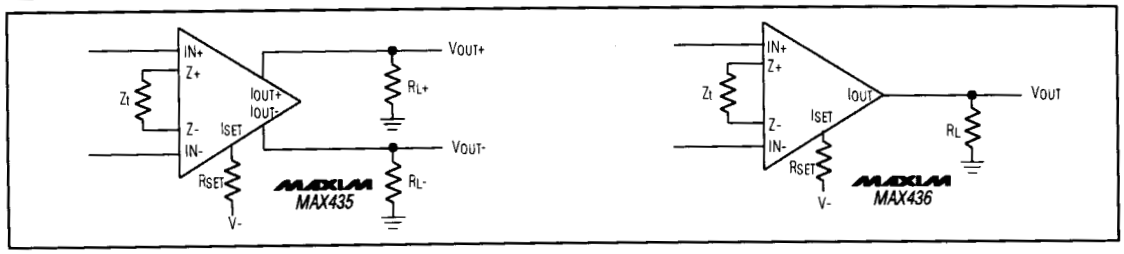
Ordering Information continued on last page.

** Dice are specified at T_A = +25°C, DC parameters only.*

Pin Configurations



Typical Operating Circuits



For free samples & the latest literature: http://www.maxim-ic.com, or phone 1-800-998-8800

Wideband Transconductance Amplifiers

ABSOLUTE MAXIMUM RATINGS

Total Supply Voltage (V+ to V-) 12V
Positive Supply Voltage (V+ to GND) 6.0V
Negative Supply Voltage (V- to GND) ..... -6.0V
Input Voltage ..... (V- - 0.3V) to (V+ + 0.3V)
Current Limit through Z+ or Z- 10mA
Output Short-Circuit Duration ..... Continuous
I_{SET} Short-Circuit Duration (to GND or V_{EE}) ..... 1 sec
Continuous Power Dissipation (T_A = +70°C)
Plastic DIP (derate 10.00mW/°C above +70°C) 800mW
SQ (derate 8.33mW/°C above +70°C) 667mW
CERDIP (derate 9.09mW/°C above +70°C) 727mW
Operating Temperature Ranges:
MAX43_C_ _ ..... 0°C to +70°C
MAX43_E_ _ ..... -40°C to +85°C
MAX43_MJD ..... -55°C to +125°C
Storage Temperature Range ..... -65°C to +160°C
Lead Temperature (soldering, 10 sec) ..... +300°C

Stresses beyond those listed under "Absolute Maximum Ratings" may cause permanent damage to the device. These are stress ratings only, and functional operation of the device at these or any other conditions beyond those indicated in the operational sections of the specifications is not implied. Exposure to absolute maximum rating conditions for extended periods may affect device reliability.

ELECTRICAL CHARACTERISTICS – MAX435

(V+ = 5V, V- = -5V, -2.5V ≤ IN+ ≤ 2.5V, -2.5V ≤ IN- ≤ 2.5V, $Z_{L+} = Z_{L-} = 50\Omega$, $Z_t = 400\Omega$, $R_{SET} = 5.9k\Omega$, T_A = +25°C, T_j = +25°C, unless otherwise noted. Specification is at +25°C junction temperature due to requirements of high-speed automatic testing. Actual values at operating temperatures may exceed the value at T_j = +25°C. No-load operating junction temperature may rise 30°C to 50°C above ambient.)

PARAMETER	SYMBOL	CONDITIONS	MIN	TYP	MAX	UNITS
Operating Supply Voltage	V+, V-	-55°C to +125°C	±4.75	±5.00	±5.25	V
Supply Current	I_S	-55°C to +125°C	30	35	41	mA
Input Voltage Range	IN+, IN-	-55°C to +125°C	-2.5		2.5	V
Input Offset Voltage	V_{OS}			0.3	3.0	mV
Input Offset-Voltage Drift	$\Delta V_{OS}/\Delta T$			3.5		µV/°C
Input Bias Current	I_B	T_A = +25°C		1.0	3.0	µA
		0°C to +70°C			5	
		-40°C to +85°C			10	
		-55°C to +125°			10	
Input Resistance	R_{IN}		200	800		kΩ
DC Resistance Looking into Z+ or Z-	R_Z			0.15	1.50	Ω
Output Voltage Range	V_{OUT+}, V_{OUT-}	$Z_{L+} = Z_{L-} = 500\Omega$	-3.5		3.5	V
Output Impedance	R_{OUT}		2.0	3.5		kΩ
Output Current	I_{OUT}		-10		10	mA
Differential Output Offset Current	$I_{OS, DIFF}$	Z+ and Z- are open, inputs are grounded	-140	0	140	µA
Output Offset Current (either side)	I_{OS}	Z+ and Z- are open, inputs are grounded	-100	-20	100	µA
Output Offset-Current Drift (either side)	$\Delta I_{OS}/\Delta T$	Z+ and Z- are open, inputs are grounded		2.3		µA/°C
Current Gain Ratio	K	T_A = +25°C	3.90	4.00	4.10	A/A
		0°C to +70°C	3.80		4.15	
		-40°C to +85°C	3.70		4.15	
		-55°C to +125°	3.60		4.20	
Common-Mode Rejection Ratio	CMRR	At DC	70	90		dB
		At 10MHz, $Z_t = 200\Omega$, $Z_{L+} = Z_{L-} = 25\Omega$		53		
Power-Supply Rejection Ratio	PSRR	At DC	50	77		dB
Settling Time to 1%	t_s	$Z_t = 200\Omega$, $Z_{L+} = Z_{L-} = 25\Omega$		18		ns
Slew Rate	SR	Z_t shorted		800		V/µs
		Z_t shorted, $Z_{L+} = Z_{L-} = 25\Omega$		550		
-3dB Bandwidth	BW	$Z_t = 200\Omega$, $Z_{L+} = Z_{L-} = 25\Omega$		275		MHz
Input Noise-Voltage Density	e_n	At 1kHz, single-ended input, $Z_t = 50\Omega$		7.0		$nV/\sqrt{Hz}$

ELECTRICAL CHARACTERISTICS – MAX436

(V+ = 5V, V- = -5V, -2.5V ≤ IN+ ≤ 2.5V, -2.5V ≤ IN- ≤ 2.5V, Z_{L+} = 50Ω, Z_t = 400Ω, R_{SET} = 5.9kΩ, T_A = +25°C, T_j = +25°C, unless otherwise noted. Specification is at +25°C junction temperature due to requirements of high-speed automatic testing. Actual values at operating temperatures may exceed the value at T_j = +25°C. No-load operating junction temperature may rise 30°C to 50°C above ambient.)

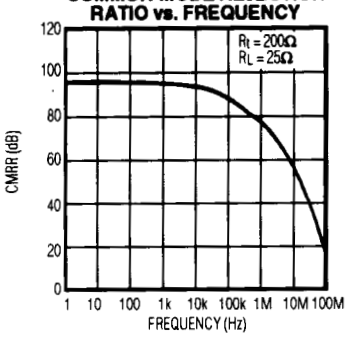
PARAMETER	SYMBOL	CONDITIONS	MIN	TYP	MAX	UNITS
Operating Supply Voltage	V+, V-	-55°C to +125°C	±4.75	±5.00	±5.25	V
Supply Current	I_S	-55°C to +125°C	30	35	41	mA
Input Voltage Range	IN+, IN-	-55°C to +125°C	-2.5		2.5	V
Input Offset Voltage	V_{OS}			0.3	3.0	mV
Input Offset-Voltage Drift	$\Delta V_{OS}/\Delta T$			3.5		µV/°C
Input Bias Current	I_B	T_A = +25°C		1.0	3.0	µA
		0°C to +70°C			5	
		-40°C to +85°C			10	
		-55°C to +125°			10	
Input Resistance	R_{IN}		200	700		kΩ
DC Resistance Looking into Z+ or Z-	R_Z			0.15	1.50	Ω
Output Voltage Range	V_{OUT+}, V_{OUT-}	Z_{L+} = 500Ω	-3.5		3.5	V
Output Impedance	R_{OUT}		2.0	3.3		kΩ
Output Current	I_{OUT}		-20		20	mA
Output Offset Current	I_{OS}	Z+ and Z- are open, inputs are grounded.	-100	6	100	µA
Output Offset-Current Drift	$\Delta I_{OS}/\Delta T$	Z+ and Z- are open, inputs are grounded.		0.75		µA/°C
Current Gain Ratio	K	T_A = +25°C	7.8	8.0	8.2	A/A
		0°C to +70°C	7.7		8.3	
		-40°C to +85°C	7.6		8.3	
		-55°C to +125°	7.4		8.4	
Common-Mode Rejection Ratio	CMRR	At DC	70	84		dB
		At 10MHz, Z_t = 200Ω, Z_{L+} = 25Ω		53		
Power-Supply Rejection Ratio	PSRR	At DC	40	50		dB
Settling Time to 1%	t_s	Z_t = 200Ω, Z_{L+} = 25Ω		18		ns
Slew Rate	SR	Z_t shorted		850		V/µs
		Z_t shorted, Z_{L+} = Z_{L-} = 25Ω		450		
-3dB Bandwidth	BW	Z_t = 200Ω, Z_{L+} = 25Ω		200		MHz
Input Noise-Voltage Density	e_n	At 1kHz, single-ended input, Z_t = 50Ω		7.0		nV/√Hz

Wideband Transconductance Amplifiers

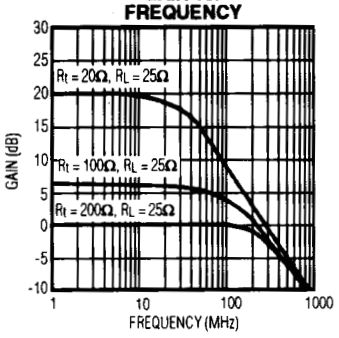
Typical Operating Characteristics

MAX435

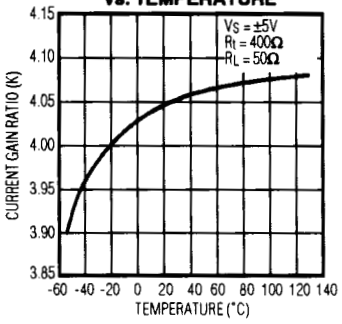
MAX435 COMMON-MODE REJECTION RATIO vs. FREQUENCY



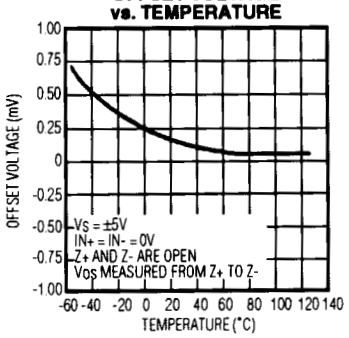
MAX435 GAIN vs. FREQUENCY



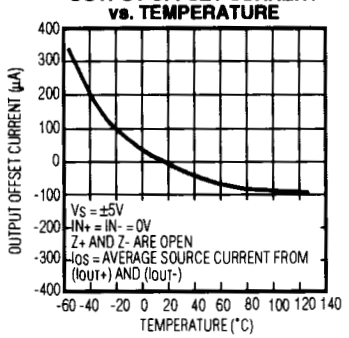
MAX435 CURRENT GAIN RATIO (K) vs. TEMPERATURE



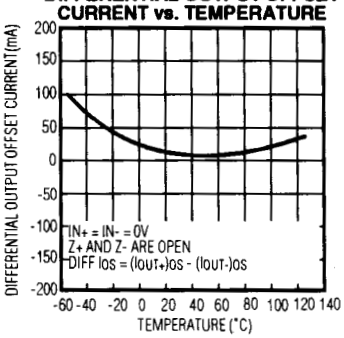
MAX435 OFFSET VOLTAGE vs. TEMPERATURE



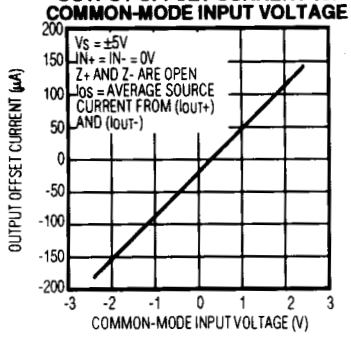
MAX435 OUTPUT OFFSET CURRENT vs. TEMPERATURE



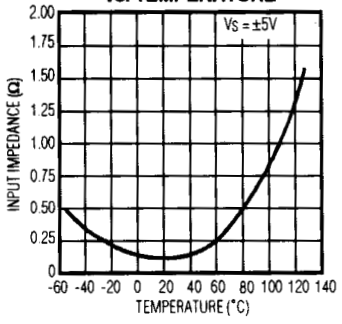
MAX435 DIFFERENTIAL OUTPUT OFFSET CURRENT vs. TEMPERATURE



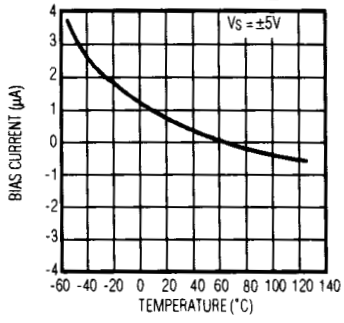
MAX435 OUTPUT OFFSET CURRENT vs. COMMON-MODE INPUT VOLTAGE



MAX435 Z TERMINAL INPUT IMPEDANCE vs. TEMPERATURE

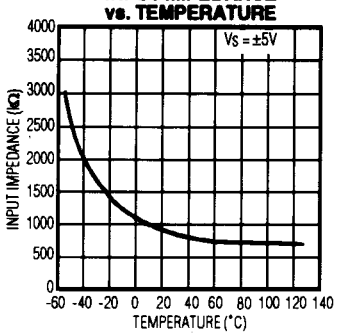


MAX435 INPUT BIAS CURRENT vs. TEMPERATURE

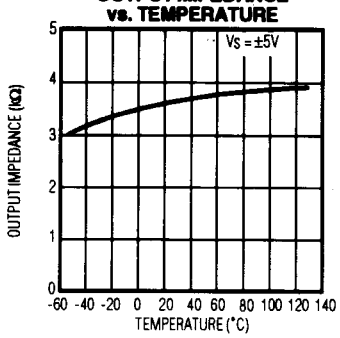


Typical Operating Characteristics (continued)

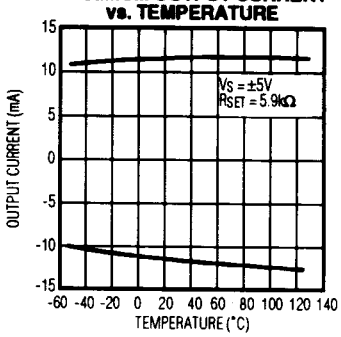
MAX435
INPUT IMPEDANCE
vs. TEMPERATURE



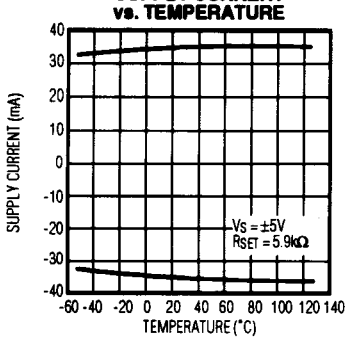
MAX435
OUTPUT IMPEDANCE
vs. TEMPERATURE



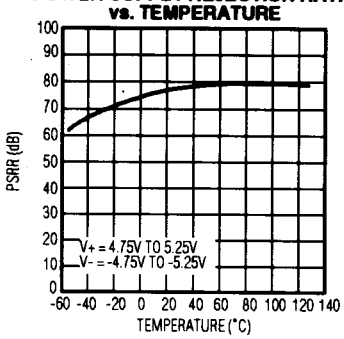
MAX435
MAXIMUM OUTPUT CURRENT
vs. TEMPERATURE



MAX435
SUPPLY CURRENT
vs. TEMPERATURE



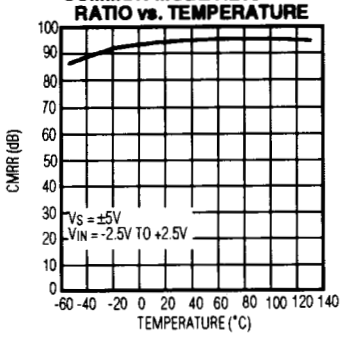
MAX435
POWER-SUPPLY REJECTION RATIO
vs. TEMPERATURE



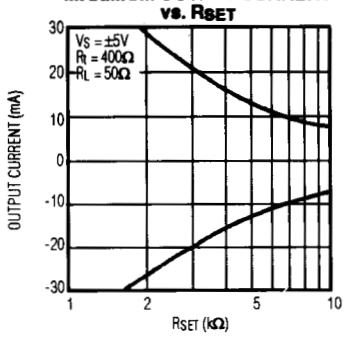
Wideband Transconductance Amplifiers

Typical Operating Characteristics (continued)

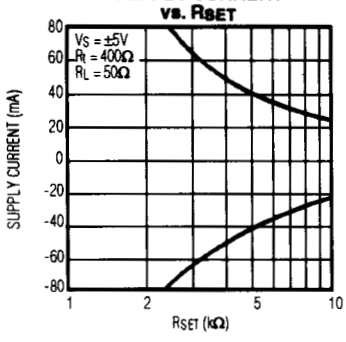
MAX435
COMMON-MODE REJECTION RATIO vs. TEMPERATURE



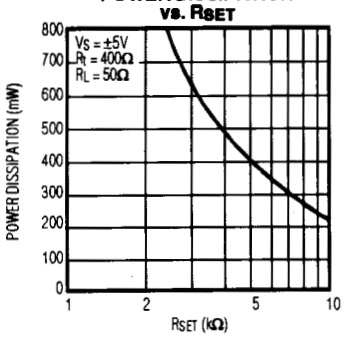
MAX435
MAXIMUM OUTPUT CURRENT vs. RSET



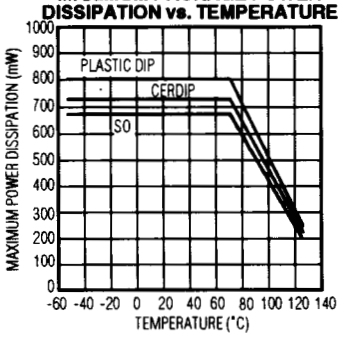
MAX435
SUPPLY CURRENT vs. RSET



MAX435
POWER DISSIPATION vs. RSET



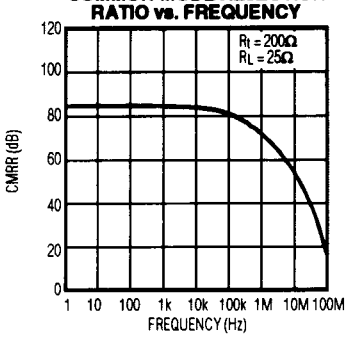
MAX435
MAXIMUM PACKAGE POWER DISSIPATION vs. TEMPERATURE



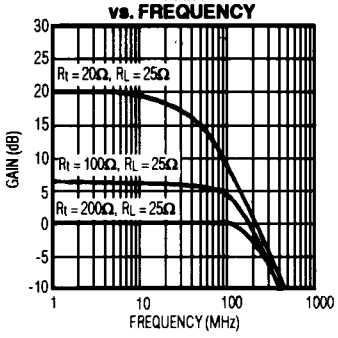
Typical Operating Characteristics (continued)

MAX436

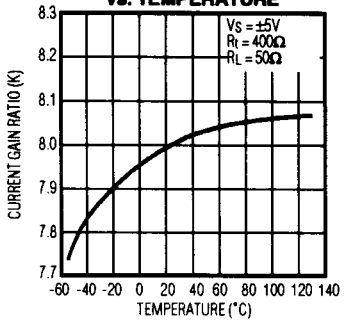
MAX436 COMMON-MODE REJECTION RATIO vs. FREQUENCY



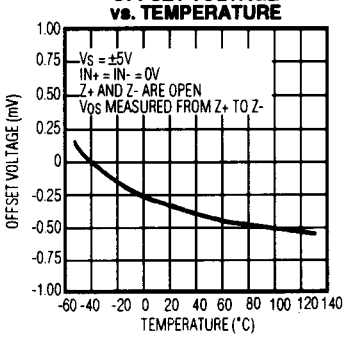
MAX436 GAIN vs. FREQUENCY



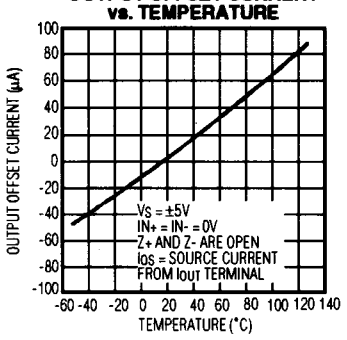
MAX436 CURRENT GAIN RATIO (K) vs. TEMPERATURE



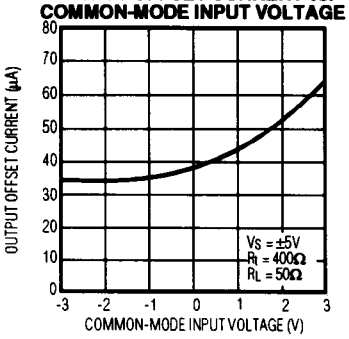
MAX436 OFFSET VOLTAGE vs. TEMPERATURE



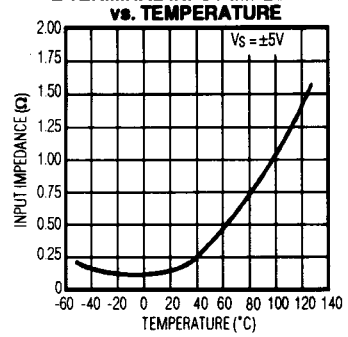
MAX436 OUTPUT OFFSET CURRENT vs. TEMPERATURE



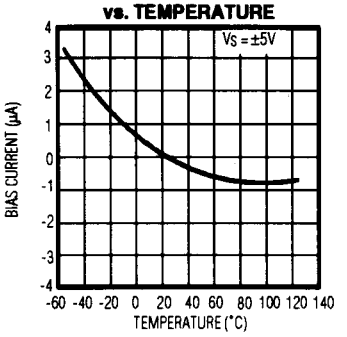
MAX436 OUTPUT OFFSET CURRENT vs. COMMON-MODE INPUT VOLTAGE



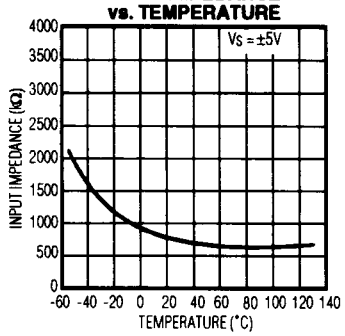
MAX436 Z TERMINAL INPUT IMPEDANCE vs. TEMPERATURE



MAX436 INPUT BIAS CURRENT vs. TEMPERATURE



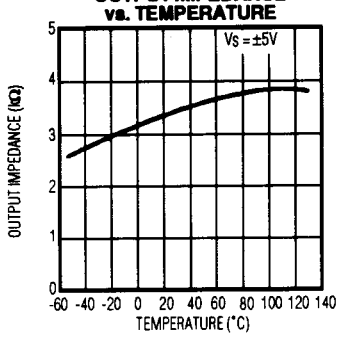
MAX436 INPUT IMPEDANCE vs. TEMPERATURE



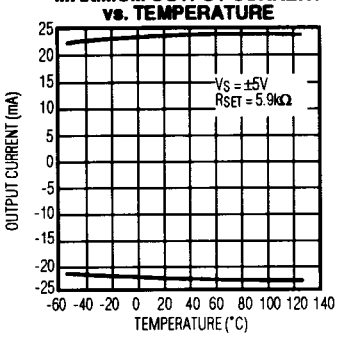
Wideband Transconductance Amplifiers

Typical Operating Characteristics (continued)

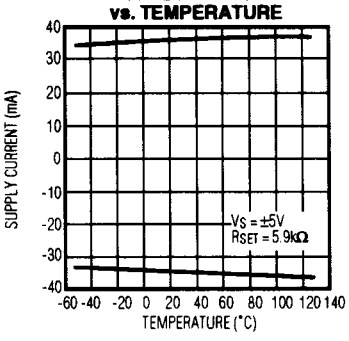
MAX436
OUTPUT IMPEDANCE vs. TEMPERATURE



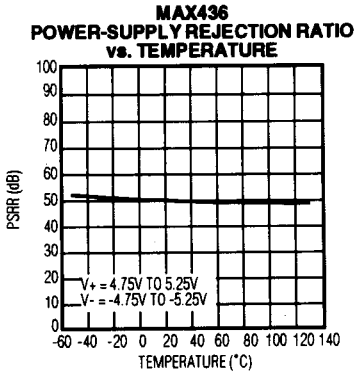
MAX436
MAXIMUM OUTPUT CURRENT vs. TEMPERATURE



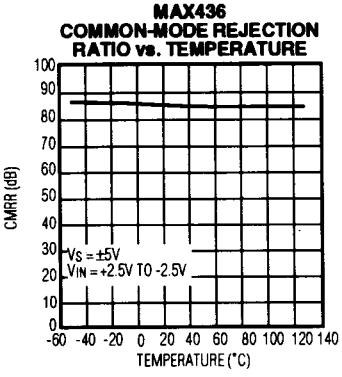
MAX436
SUPPLY CURRENT vs. TEMPERATURE



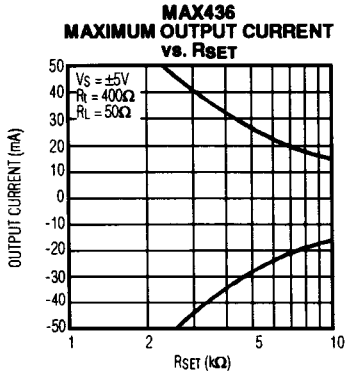
MAX436
POWER-SUPPLY REJECTION RATIO vs. TEMPERATURE



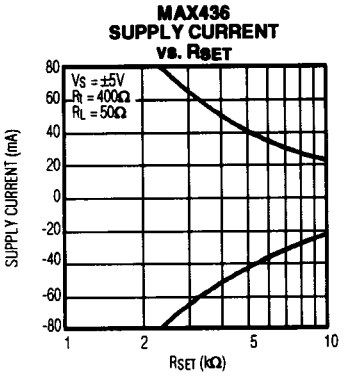
MAX436
COMMON-MODE REJECTION RATIO vs. TEMPERATURE



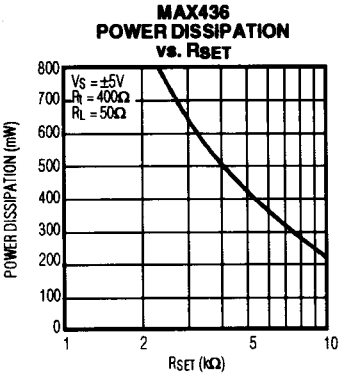
MAX436
MAXIMUM OUTPUT CURRENT vs. RSET



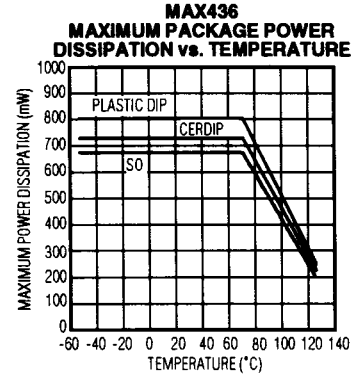
MAX436
SUPPLY CURRENT vs. RSET



MAX436
POWER DISSIPATION vs. RSET



MAX436
MAXIMUM PACKAGE POWER DISSIPATION vs. TEMPERATURE



Pin Description

PIN		NAME	FUNCTION
MAX435	MAX436		
1, 12, 14	1, 12, 14	V+	Positive Power Supply (+5V)
2	2	IN+	Noninverting Amplifier Input
3	3	Z+	Positive Transconductance Terminal
4	4, 9	N.C.	No Connect - Leave this pin open
5	5	Z-	Negative Transconductance Terminal
6	6	IN-	Inverting Amplifier Input Terminal
7, 8, 10	7, 8, 10	V-	Negative Power Supply (-5V)
9	–	I_{OUT-}	Inverting Differential Output
11	11	I_{SET}	Supply-Current Set Terminal
13	–	I_{OUT+}	Noninverting Differential Output
–	13	I_{OUT}	Amplifier Output Terminal

Theory of Operation

The MAX435/MAX436 are wideband transconductance amplifiers (WTA) that operate with no feedback. These amplifiers are intrinsically stable since closed-loop phase shift does not affect amplifier stability. Unlike conventional voltage-mode amplifiers, WTAs need no compensation from an internally set dominant pole. Voltage-mode amplifiers require this pole to roll off open-loop gain and prevent oscillation from high-frequency phase shift.

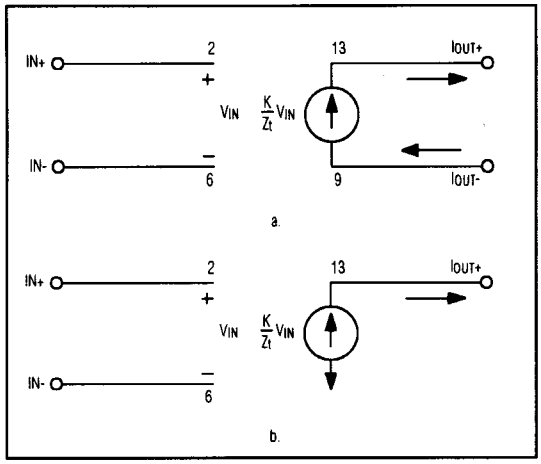


Figure 1. a) MAX435 Operational Model. b) MAX436 Operational Model.

The MAX435/MAX436's unique architecture allows single-ended (MAX436) or differential (MAX435) signal gain to be set by the ratio of two impedances: the user-selected transconductance element or network (Z_t), and the output load impedance (Z_L). The WTAs are essentially voltage-controlled current sources, as shown in Figure 1.

The MAX435/MAX436's output is a current proportional to the differential input voltage, and inversely proportional to the impedance of the user-selected transconductance element or network (Z_t). The current output provides inherent short-circuit protection for the output terminals.

A differential input voltage (V_{IN}) applied between the input terminals causes current to flow in the transconductance element (Z_t). This current is equal to V_{IN}/Z_t. The current in the transconductance element is multiplied by the preset current gain (K) of the WTA, and appears at the output terminal(s) as a current equal to (K) x (V_{IN})/Z_t. This current flows through the load impedance to produce an output voltage according to the following equation:

$$V_{OUT} = K\left(\frac{Z_L}{Z_t}\right)V_{IN}$$

Where: K = WTA Current-Gain Ratio
Z_L = Output Load Impedance
Z_t = Transconductance Element Impedance
V_{IN} = Differential Input Voltage

Unlike current-feedback amplifiers, the MAX435/MAX436 have symmetrical inputs with an input impedance of about 750kΩ. This allows true differential input applications to be realized, as shown in the MAX435 *Typical Operating Circuit*. And, because the input is symmetrical, opposite signal polarity can be obtained by swapping the input terminals.

With proper selection of component values, it is possible to implement an amplifier circuit that accepts differential input voltages covering the WTA's entire input voltage range (-2.5V to +2.5V), without overloading its output stage. This characteristic makes the MAX435/MAX436 ideal for use in wideband instrumentation amplifiers, differential receivers, and settling-time measurement circuits.

Design Procedure

Setting the Circuit Gain

The MAX435/MAX436 produce an output current by multiplying a differential input voltage, V_{IN}, by the transconductance, K/Z_t. The voltage gain (A_V) is set by the impedance of the transconductance network (Z_t) and the output load impedance (Z_L), according to the following formula:

$$A_V = K\left(\frac{Z_L}{Z_t}\right)$$

Wideband Transconductance Amplifiers

The factor K in the gain equation refers to the WTA's current gain. K is factory trimmed to provide a low-drift, stable circuit gain for WTA applications. K is trimmed to 4.0 ±2.5% for the MAX435, to 8.0 ±2.5% for the MAX436.

The factor Z_t in the gain equation is the impedance of a user-selected, 2-terminal transconductance element or network. This network is connected across the WTA's transconductance terminals, labeled as Z+ and Z-. The transconductance network should be selected, along with the output impedance, to provide the desired circuit gain and frequency shaping. To maintain linearity, the transconductance network should also be selected so that the current flowing through it, equal to V_{IN}/Z_t, does not exceed 2.5mA under worst-case conditions of maximum input voltage and minimum transconductance element impedance (Z_t). Output current should not exceed ±10mA per output for the MAX435, or ±20mA for the MAX436.

Supply Current (I_{SET}) Control

An external current source controls the WTA's internal current. Connecting an external resistor (R_{SET}) from the I_{SET} pin to the negative power supply (V-) controls the current source. The typical value for this resistor is 5.9kΩ, which provides an output-current drive capability of ±10mA per output for the MAX435, ±20mA for the MAX436. Connect a 0.22µF ceramic capacitor from the I_{SET} pin to V+, to decouple the current source.

A larger resistor value will reduce the devices' supply current, power dissipation, and output-current drive capability. A smaller resistor value may also be used to increase the output-current capability, but take care that the power dissipation rating for the device package type is not exceeded for the specific circuit operating conditions. It is especially important to consider the maximum load current and ambient operating temperature of the circuit. Refer to the graphs in the *Typical Operating Characteristics* section for typical values of load current and power dissipation for different values of R_{SET}.

Shutdown Mode

The WTA supply current varies with the value of R_{SET}, but it is relatively independent of the differential input voltage and the output load. The WTAs achieve much of their high-speed performance by maintaining a constant level of current within the internal transistors, and steering some of this current to the output when a differential input voltage is applied. To reduce power dissipation when the output load does not need to be actively driven, the I_{SET} pin can be used to place the WTA into a low-power shutdown mode. If the I_{SET} pin is disconnected from the R_{SET} resistor and left open, the supply current of the MAX435 is reduced to approximately 450µA, while the supply current of the MAX436 is reduced to approximately 850µA. To reactivate the WTA, simply reconnect the I_{SET} pin to the R_{SET} resistor.

PIN	R_{SET}	Supply Current (µA Typ, +25°C)	
		MAX435	MAX436
I_{SET} (11)	Open (00)	450	850

DC Accuracy

The DC accuracy of circuits implemented with the MAX435/MAX436 is affected by several parameters, including:

1) Input Offset Voltage
2) Output Offset Current
3) Accuracy of WTA Current Gain Factor (K)
4) Tolerance of External Transconductance and Load Impedances

The input offset voltage of the WTA is caused by a mismatch of V_{BE} voltages between the transistors in the amplifier input stage - the same mechanism that produces input offset voltages in ordinary bipolar amplifiers. The input offset voltage is measured as the voltage between the transconductance terminals (from Z+ to Z-), with each of the inputs (IN+ and IN-) grounded and no transconductance element in the circuit (Z+ and Z- are open).

In an actual application circuit, the input offset voltage causes a current to flow in the transconductance element when no differential voltage is applied at the input terminals. This current is then multiplied by K to produce an error current at the output terminal(s).

A second source of DC error is the output offset current. This parameter is measured with no transconductance element in the circuit (Z+ and Z- are open) and the input terminals (IN+ and IN-) grounded. Output offset current is the current that flows from an output terminal of the WTA under these conditions. The output offset current is the result of imperfect matching of devices in the output current mirror(s) of the WTA. The output current caused by the input offset voltage, as discussed previously, is NOT included in the Output Offset Current specification, since that component of the total output current will vary with the value of the transconductance element.

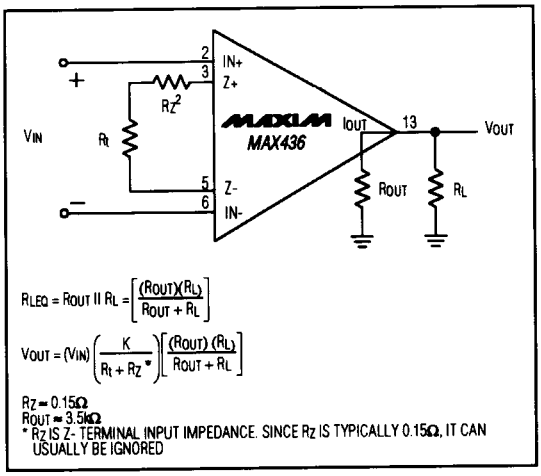


Figure 2. Finite Output-Impedance of WTA

The total DC output voltage error (at each output for the MAX435) due to the input offset voltage and output offset current is calculated with the following formulas:

MAX435:

$$V_{ERR+} = (R_{L+})\left[(I_{OS+}) + (K)\left(\frac{V_{OS}}{R_t}\right)\right]$$

$$V_{ERR-} = (R_{L-})\left[(I_{OS-}) - (K)\left(\frac{V_{OS}}{R_t}\right)\right]$$

$$V_{ERR(DIFF)} = (V_{ERR+}) - (V_{ERR-})$$

MAX436:

$$V_{ERR} = R_L\left(I_{OS} + K\frac{V_{OS}}{R_t}\right)$$

Where: V_{ERR} = Output-Voltage Error
R_L = Output Load Resistance
R_t = Transconductance Element Resistance
I_{OS} = Output Offset Current
V_{OS} = Input Offset Voltage (from Z+ to Z-)
K = WTA Current Gain

Variation of the current gain factor (K) between devices and over operating temperature is also a source of gain error in WTA applications. K for each of the devices is factory trimmed to an initial tolerance of ±2.5%. The variation of K with operating temperature is listed in the *Electrical Characteristics* tables.

The finite output impedance also affects the amplifier gain. The output(s) are voltage-controlled current sources. An ideal current source would have an infinite output impedance, so that the output current would be independent of the load impedance. In practice, the MAX435/MAX436's output impedance is about 3.5kΩ.

As shown in Figure 2, the WTA output impedance (R_{OUT}) is paralleled with the circuit load impedance (R_L), reducing the equivalent load impedance. After accounting for the finite WTA output impedance (R_{OUT}), the actual circuit gain is calculated with the following formula:

$$A_V = \frac{V_{OUT}}{V_{IN}} = \left(\frac{K}{R_t}\right)\left[\frac{(R_{OUT})(R_L)}{R_{OUT} + R_L}\right]$$

The voltage gain error (ΔA_V) with respect to the theoretical gain ($A_V = K \times R_L/R_t$) is equal to:

$$\frac{\Delta A_V}{A_V} = \frac{R_L}{R_L + R_{OUT}}$$

Power-Supply Bypassing and Board Layout

Although the WTA architecture eliminates closed-loop phase shift as a source of circuit oscillation, careful high-frequency circuit design methods should be used to optimize the performance of WTA circuits.

Proper power-supply bypassing and board layout deserve careful attention. It is recommended that a ground plane be used with the MAX435/MAX436. The ground plane should include the entire portion of the PC board that is not dedicated to a specific signal trace.

Sockets are not recommended with the WTAs, because the additional pin-to-pin capacitance they introduce degrades wideband performance. Keep the length of traces connecting to the WTA input terminals as short as possible to minimize signal reflections and/or inductive coupling of high-frequency signals to the WTA. If the input signals must travel more than a few inches, use controlled impedance lines or coaxial cables; all signals should be properly terminated. Minimize the PCB pad

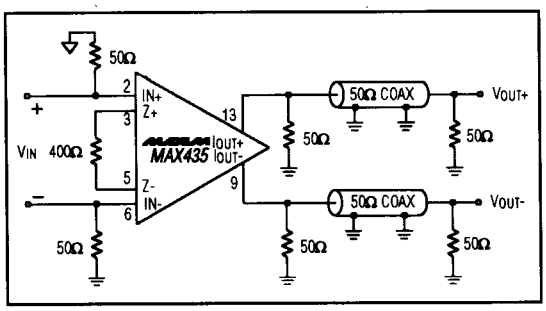


Figure 3. MAX435 Coaxial Cable Driving Circuit

Wideband Transconductance Amplifiers

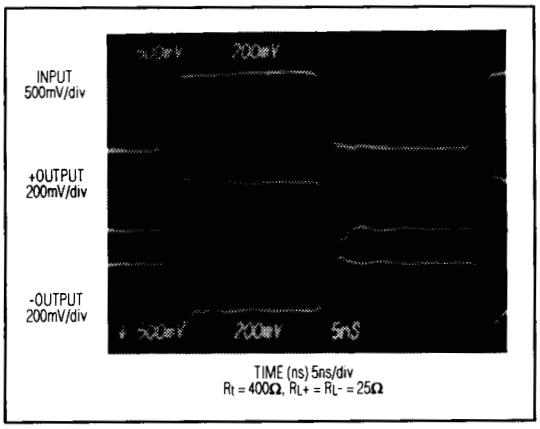


Figure 4. MAX435 Pulse Response

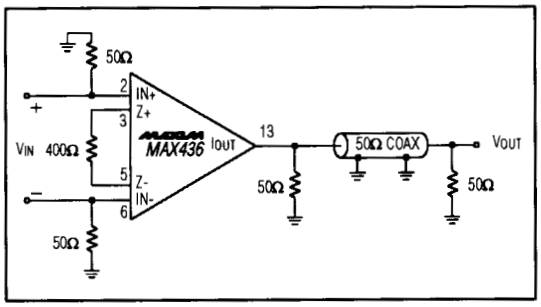


Figure 5. MAX436 Coaxial Cable Driving Circuit

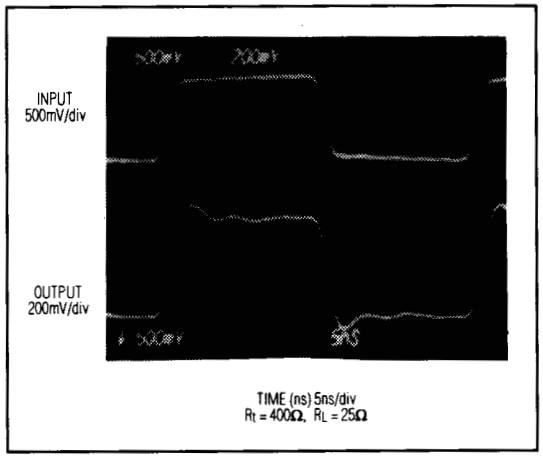


Figure 6. MAX436 Pulse Response

area for input connections, to prevent capacitive coupling of stray high-frequency signals.

Passive components used with the WTA should preferably be surface mount, to minimize stray inductance. If surface-mount components are not used, component lead lengths should be kept to an absolute minimum.

Bypass each power supply directly to the ground plane with a 0.22µF ceramic capacitor, placed as close to the supply pins as possible. Bypass the ISET pin with a 0.22µF ceramic capacitor to the V+ pin. Keep capacitor lead lengths as short as possible to minimize series inductance; surface-mount (chip) capacitors are ideal for this application.

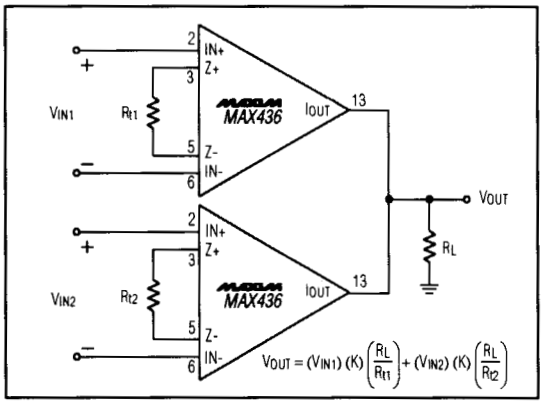


Figure 7. Summing Amplifier

Capacitive Load Driving

Since the WTA requires no feedback, phase shift due to capacitive loading of the output will not degrade the circuit stability. The primary effect of capacitive loading is a reduction in the output slew rate and bandwidth, which is limited by the rate at which the WTA output current can charge the capacitive load. Avoid capacitive coupling from the WTA output terminals to the input or transconductance terminals, since it introduces high- frequency feedback and possible oscillations.

Application Circuits

The WTA's unique architecture allows the implementation of many unique application circuits, some of which have been included here.

For the sake of clarity, bypass capacitors and the RSET resistor have not been shown in the following applications circuit schematics. The value of RSET is 5.9kΩ for every application circuit in Figures 3-16. For every application circuit in this data sheet, each power supply was bypassed to GND with a 0.22µF ceramic capacitor. The ISET pin was bypassed to V+ with a 0.22µF ceramic capacitor.

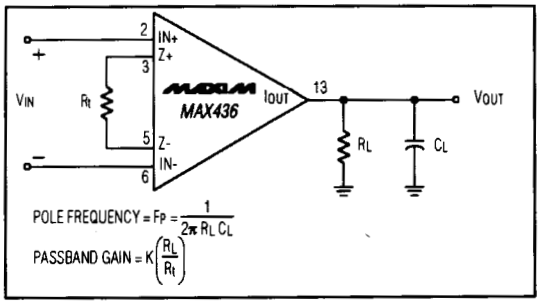


Figure 8. Lowpass Amplifier

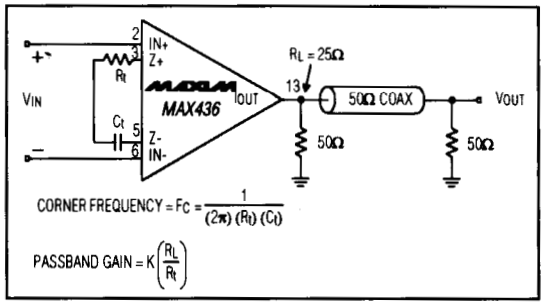


Figure 10. Highpass Amplifier

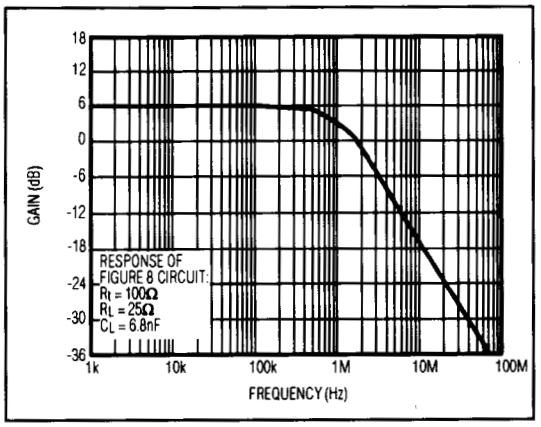


Figure 9. Lowpass Amplifier Gain vs. Frequency

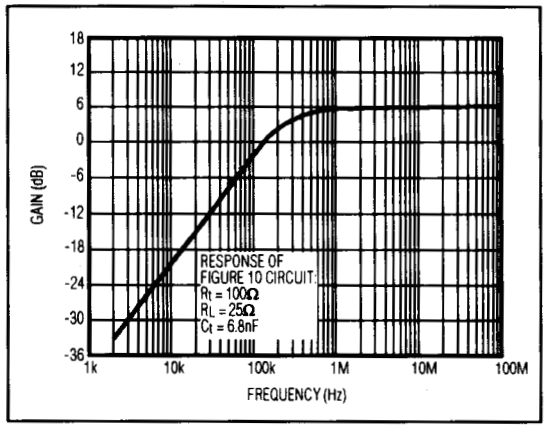


Figure 11. Highpass Amplifier Gain vs. Frequency

Coaxial Cable Drivers

To maximize power transfer and minimize distortion of high-speed signals, transmission lines must be terminated at the source and receiving ends with the characteristic impedance of the line itself. A 50Ω coaxial cable requires an impedance of 50Ω at each end for optimum performance.

With voltage-mode amplifiers, the transmitting end of a coaxial cable is typically back-terminated with a resistor in series with the amplifier output, since the voltage-mode amplifier's output has a very low impedance.

The WTA output is a current source, so its output impedance is relatively high (approximately 3.5kΩ). Therefore, when driving coaxial cables with the WTA, the back-termination resistor should be placed in parallel with the output impedance of the WTA, as shown in Figures 3 and 5. Note that back terminating cables in this manner will reduce the effective voltage gain of the circuit by a factor of 2.

The pulse response of the circuit in Figure 3 is shown in Figure 4, while Figure 6 is the pulse response of Figure 5's circuit.

Summing Amplifier

Two or more signals can be summed together by simply tying the output terminals of WTA's together as shown in Figure 7.

Lowpass Amplifier

A parallel RC network at the amplifier output will result in the lowpass amplifier circuit of Figure 8, with a -3dB corner frequency of:

$$FC = \frac{1}{(2\pi)\ (R_L)\ (C_L)}$$

The response of Figure 8's circuit with R_t = 100Ω, C_L = 6.8nF and R_L = 25Ω is plotted in Figure 9.

Highpass Amplifier

A series RC network between the transconductance terminals, as shown in the circuit of Figure 10, will produce a highpass amplifier. The response of this circuit is plotted in Figure 11, for R_t = 100Ω, C_t = 6.8nF and R_L = 25Ω.

Wideband Transconductance Amplifiers

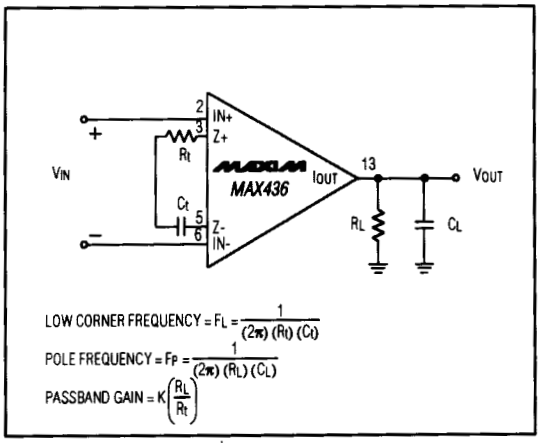


Figure 12. Bandpass Amplifier

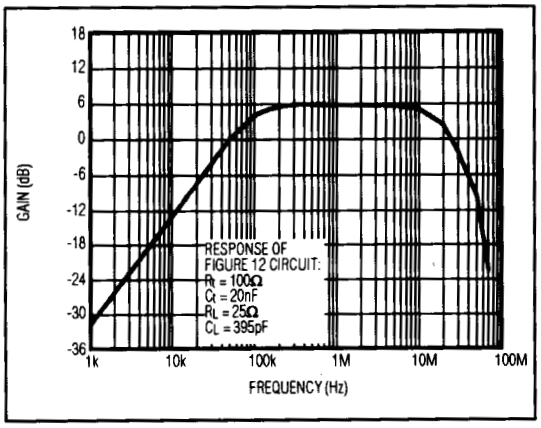


Figure 13. Bandpass Amplifier Gain vs. Frequency

Bandpass Amplifier

Since there is no interaction between the transconductance network impedance and the output load impedance, poles and zeros in the WTA transfer function can be independently set by the two impedance networks. The circuit in Figure 12 is a bandpass amplifier, with the low corner frequency set by the impedance of the transconductance network. The high corner frequency is set by the impedance of the RC network at the amplifier output. The passband gain is (K) x (RL/Rt).

Figure 13 is a plot of the response of the circuit in Figure 12, with R_t = 100Ω, C_t = 20nF, R_L = 25Ω, and C_L = 395pF.

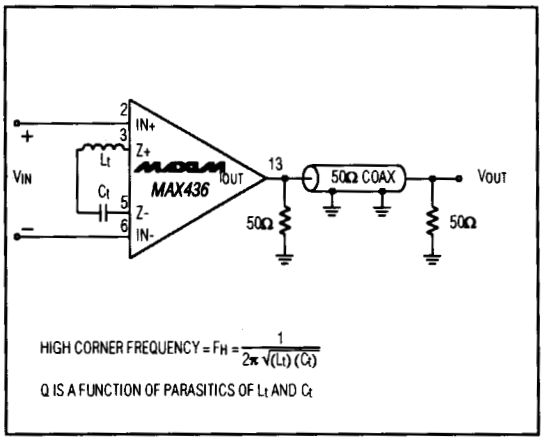


Figure 14. Tuned Amplifier

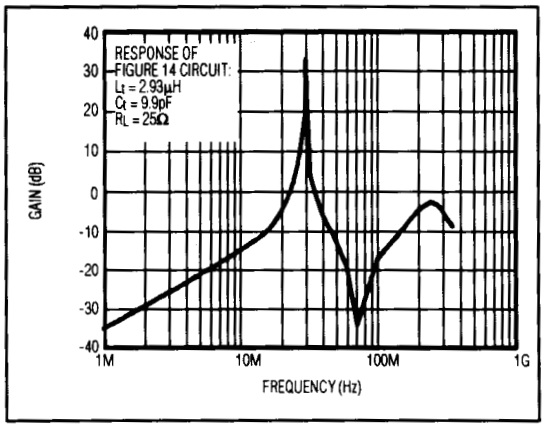


Figure 15. Tuned Amplifier Gain vs. Frequency

Tuned Amplifier

The circuit in Figure 14 is a tuned amplifier circuit, tuned to the resonant frequency of the LC transconductance network, or

$$FC = \frac{1}{2\pi\sqrt{Lt\ Ct}}$$

The impedance of the transconductance network is a minimum at the resonant frequency, providing maximum amplifier gain at that frequency. The Q of the amplifier is a function of the parasitic components associated with the LC network. Figure 15 is the frequency response of the circuit in Figure 14, with L_t = 2.93µH and C_t = 9.9pF.

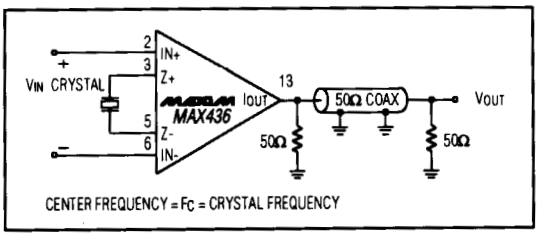


Figure 16. Crystal Tuned Amplifier

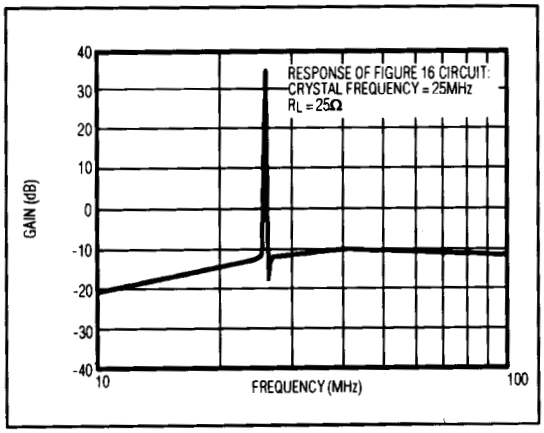


Figure 17. Crystal Tuned Amplifier Gain vs. Frequency

The high-frequency gain of this circuit (beyond the resonant frequency) can be further reduced by increasing the load capacitance.

Crystal Tuned Amplifier

If a higher Q and more accurate control of the tuned amplifier frequency is required, a crystal can replace the tuned LC network as the WTA transconductance element, as shown in the "Crystal Tuned Amplifier" circuit of Figure 16. The crystal's impedance is a minimum at its resonant frequency, resulting in maximum gain for the amplifier circuit at that frequency.

The frequency response of Figure 16's circuit with a 25MHz crystal is shown in Figure 17.

Video Twisted-Pair Driver/Receiver Circuit

For distances less than about 5000 ft. (1500m), a single channel of baseband (composite) video can be transmitted with surprisingly high quality across twisted-pair cabling, saving significant cost over traditional coaxial cabling. When using twisted pair as a transmission medium, two things are of highest importance: balanced (differential) transmission, to minimize common-mode noise, and proper termination to minimize reflections. The MAX435/MAX436, with 53dB CMRR at 10MHz and high input and output impedance, are well suited for this application.

Figure 18's circuit illustrates the MAX435 and MAX436 used as a driver/receiver circuit for twisted-pair video transmissions. One differential output MAX435 drives the balanced twisted pair from a ground-referred input signal, eliminating the need for a balun transformer or two single-ended output drivers.

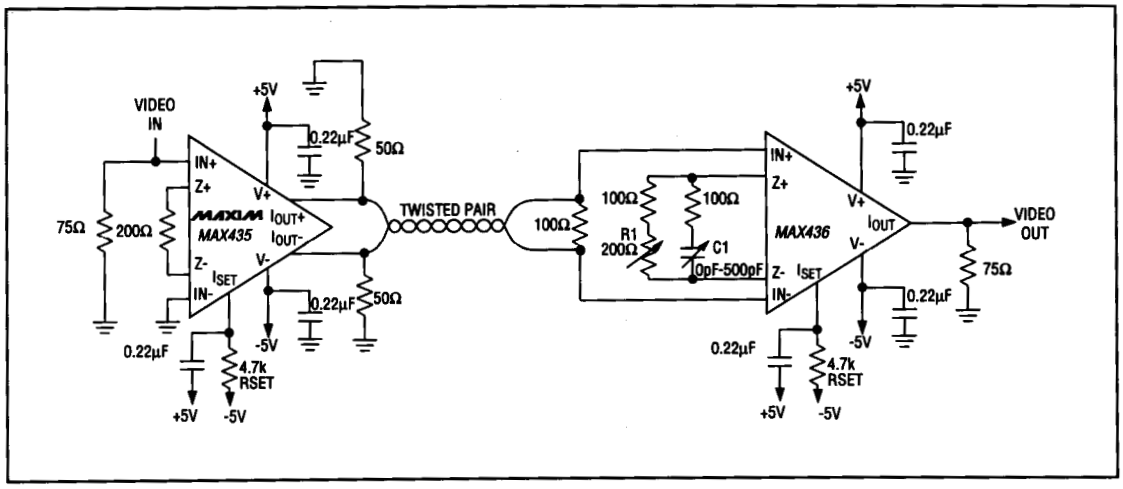


Figure 18. Video Twisted Pair Driver/Receiver Circuit

Wideband Transconductance Amplifiers

The receiver circuit uses the single-ended output MAX436 for balanced to single-ended line conversion. The 100Ω resistor from IN+ to IN- provides proper line termination. The transconductance network (from Z+ to Z-) performs the gain adjustment (+6dB), as well as line equalization. Line equalization is sometimes required to boost the high-frequency gain of the receiver to account for the limited bandwidth of the twisted pair.

In Figure 18's circuit, compensation is achieved by adjusting R1 for proper **brightness** (to boost overall gain to compensate for ohmic losses), and C1 for best **color** (to add a zero/pole pair to extend the bandwidth). Since this equalization is done on the receiver end, the end user simply views the screen and adjusts the controls for the best picture. For many NTSC applications, this line equalization may not be necessary, since the NTSC monitor performs a fair amount of loss equalization by calibrating to the test patterns in the vertical interval test signal (VITS).

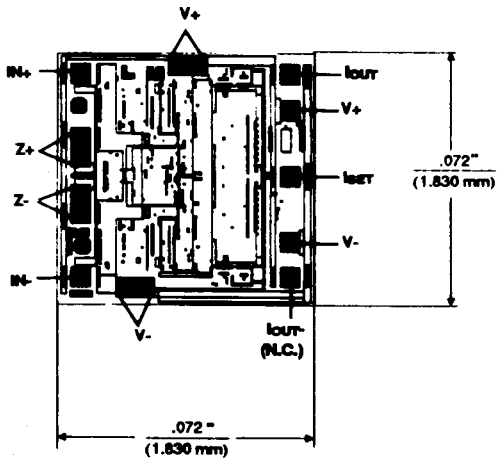
Ordering Information (continued)

PART	TEMP. RANGE	PIN-PACKAGE
MAX436CPD	0°C to +70°C	14 Plastic DIP
MAX436CSD	0°C to +70°C	14 SO
MAX436C/D	0°C to +70°C	Dice*
MAX436EPD	-40°C to +85°C	14 Plastic DIP
MAX436ESD	-40°C to +85°C	14 SO
MAX436MJD	-55°C to +125°C	14 CERDIP

** Dice are specified at T_A = +25°C, DC parameters only.*

Chip Topography

MAX435/MAX436

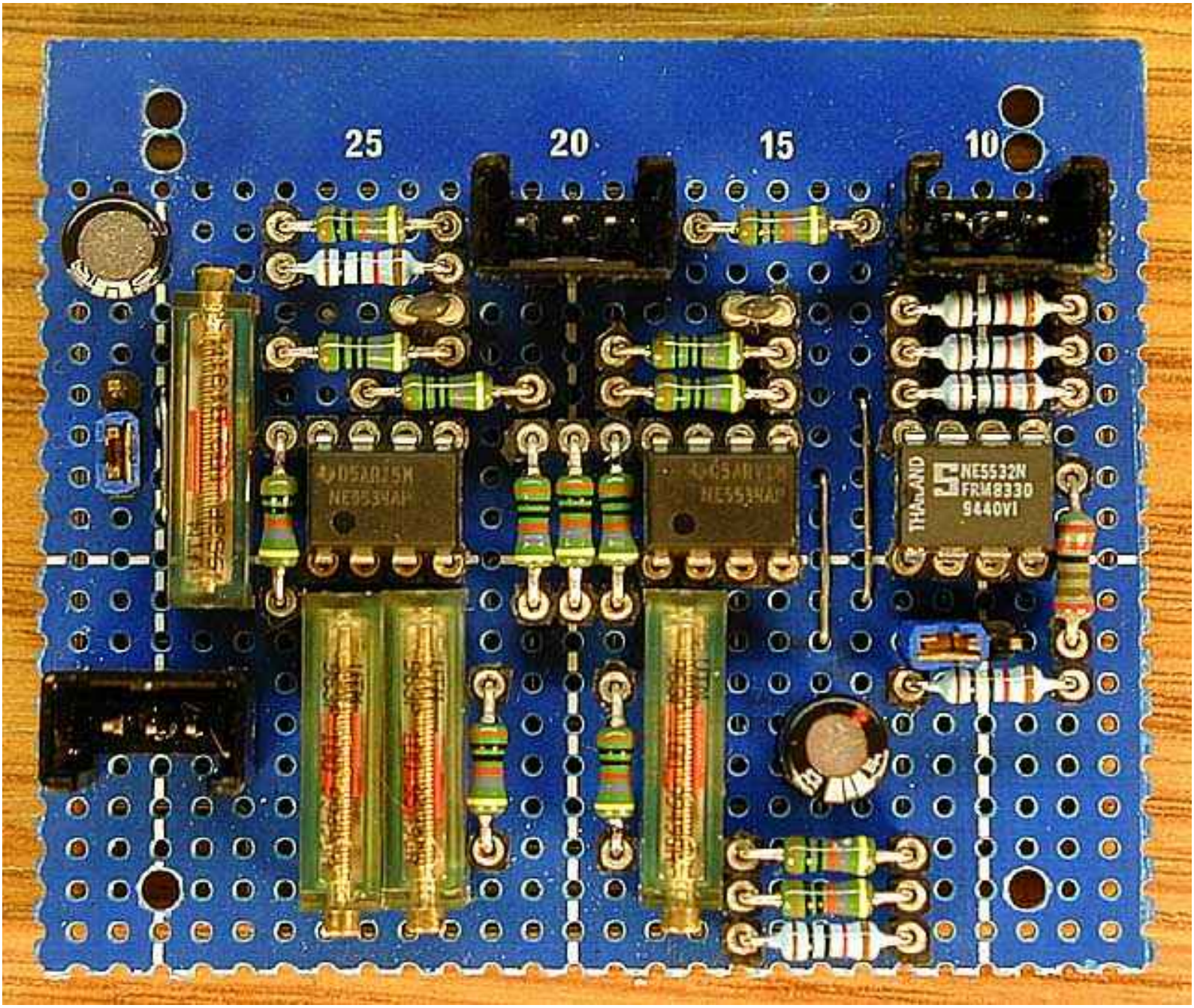


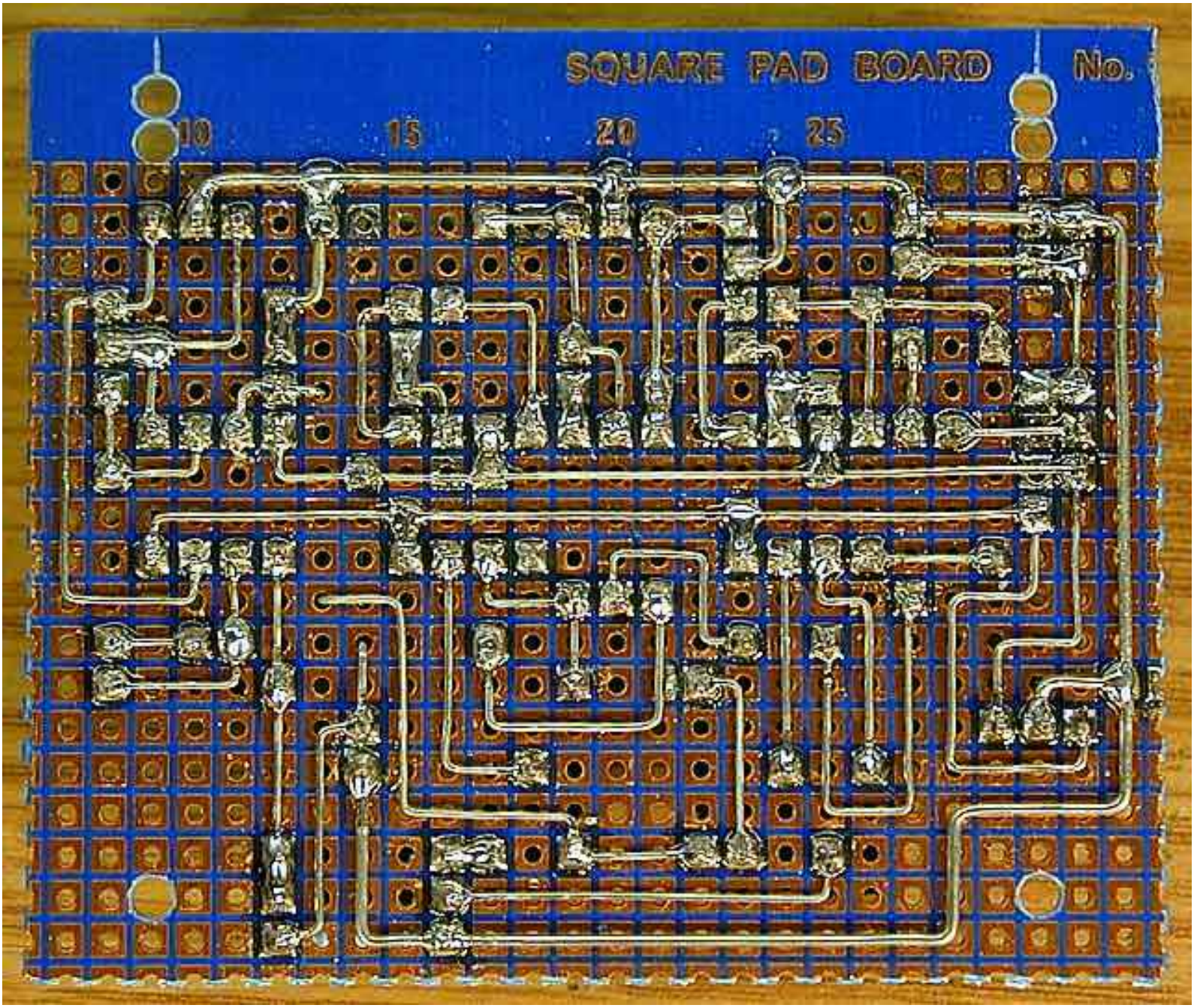
() are for MAX436 only.

TRANSISTOR COUNT: 230;
SUBSTRATE CONNECTED TO V-.

Maxim cannot assume responsibility for use of any circuitry other than circuitry entirely embodied in a Maxim product. No circuit patent licenses are implied. Maxim reserves the right to change the circuitry and specifications without notice at any time.

© 1993 Maxim Integrated Products Printed USA MAXIM is a registered trademark of Maxim Integrated Products.


25
20
15
10
NE5532N
FRM8330
9440VI
THAILAND


SQUARE PAD BOARD No.
10
15
20
25

The Audio Pages

Elliott Sound Products — Download Page

Page Last Updated - 31 Dec 2004

This page has links for all download files available from The Audio Pages, as well as some other useful resources. There are not a great many, but the ones here are all extremely useful (IMHO). Feel free to suggest others, or submit your own (submissions must be freeware, and not crippled or restricted in any way). Ad-ware (with embedded advertising) will not be accepted under any circumstances.

Main Index

Index

Executables

- Stores - Small Business & Hobbyist Stock Management
- Cron - Windows Job Scheduler
- Discrete transistor data
- Linkwitz-Riley crossover calculator
- Multiple feedback bandpass filter calculator
- World Time Clock
- Reminder - Memory jogger for the terminally forgetful
- LM3815 - Nifty little calculator to determine the values for the LED meter IC
- My "BOGUS" fake e-mail generator program. Clog the spammers lists until they choke!

Spreadsheets

- Linkwitz transform calculator
- Theile-Small Parameters Spreadsheet

- Heatsink thermal rating calculator
- Passive Crossover Design Spreadsheet

Application Notes

- See Application Notes Index

Miscellaneous

- SIMetrix - An excellent freeware circuit simulator

Icon Legend

	ESP original executable (Windoze only - sorry)		Microsoft Excel Spreadsheet
	Compressed Archive (including self extracting)		Adobe Portable Document Format

Windows DLL Files

If you get an error message at startup, it is probable that the Visual Basic runtime library is not installed on your computer. The error message will be along the lines of 'ERROR starting Program. A required DLL file, VB40032.DLL was not found'.

For Windows 98 and ME, the DLL (Dynamically Linked Library) should be in the ***\windows\system*** folder, or ***\windows\system32*** for XP, NT or 2000. The executable programs will not run if this file is missing, or is in the wrong location.

For those souls who have tried in vain to get a copy of the Visual Basic DLL, visit ... http://support.microsoft.com/support/kb/articles/q196/2/86.asp and select VB4run.exe as a download. Alternatively, download the VB40032.DLL file from here.

Programs

STORES - StoreSetup.zip (Compressed setup file - 291,866 bytes) The Stores program was written to replace my old stock management system, which was past its use-by date. Make sure you read the readme file (readme.txt).

Stores is a basic but extremely useful stock management program, that allows small manufacturers, repair shops and hobbyists to maintain a full database of all the parts they normally use. It is not limited to electronics, and can be used with anything, from tools, car parts to spares and finished goods in any industry. The version here is limited to a single database of 250 entries in a single database file, and has no reports (although the menu items are still there).

The setup file is set up to use a directory C:\Program Files\Stores which will be created during the installation. Simply run SetupStores to extract the files. If you are upgrading, don't allow "parts.txt" to be overwritten! This is the default database file. By default, the SetupStores program will not overwrite the database or config files.

The program may be registered, which will get you a CD-ROM and unlimited downloads of updates and additional functionality. The full version supports unlimited database size, 5 separate database files, stock value and re-order reports, job files, and more to come.

Both versions of the program have extensive help, and use a simple text database format. Even without indexing, a search on a 32,000 record database takes about 2 seconds on a Pentium 666MHz machine.

Registration is AU$40.00 (US$29.00), and includes a CD-ROM with the latest version and as a bonus also includes the ESP construction articles. The CD-ROM also contains the key to unlock future downloads.

CRON - cron.exe (14,428 bytes - self extracting archive) Very simple but useful job scheduler for Windows™ machines, and it has nothing to do with audio. Ideal for PCs or servers to schedule network backups (via a simple batch file), or to run other programs at the selected time(s). Can control up to 10 scheduled jobs.

ESP-SEMI - esp-semi.exe (37,735 Bytes - self extracting archive) This is a small program to find transistor data. Not everything is listed (1442 different devices). Download a copy, place it into the directory of choice and run the program, which is a self extracting archive. There are two extracted files - ESP-TRAN.EXE and BIPOLAR.TXT, being the executable and database respectively. ESP-SEMI.EXE may be deleted after extraction unless you want to give a copy to someone else. New stuff can be added as you find the data. If you add a lot of stuff, feel free to e-mail me a copy of the new database (compressed, *please*!), and I will add it to the database file.

For users who eschew Windows, the text file is still useful, as it can be read with any text editor or even a spreadsheet (TAB delimited format). Be careful with spreadsheets, as they like to try to convert some data into dates (really useful - not!).

ESP-LR - esp_lr.exe (53,760 bytes) Linkwitz-Riley crossover network calculator program. Shows the component values needed for the selected frequency, or will show the frequency for given component values. The help screen also has the standard E12 and E24 component value range for reference.

ESP-LR12 - esp-lr12.exe (66,560 bytes) Version 1.2 supports both 24dB/Octave and 12dB/Octave Linkwitz-Riley filters for greater flexibility and more options for the constructor.

MFB-FILTER - mfb-filter.exe (48,128 bytes) This program is designed to take the tedium from designing multiple feedback bandpass filters. These filters are commonly used in graphic equalisers, analysers and for special applications. A set of help screens are provided to assist with component value selection, and for determining the optimum frequency and Q of the filters for various applications.

WTIME - wtime.exe (180,224 bytes) WorldTime95 - An interesting and useful world time clock, submitted by Cyril Martin. Has 160 different cities, so time zone selection has never been easier. Requires Microsoft Visual Basic 6 runtime libraries.

REMINDER - reminder.exe (8,844 bytes) Reminder is an interesting and very useful little program to help you remember those important dates, such as birthdays, anniversaries, car repayments, and almost anything you dare not forget. In various forms, I have used this for over 9 years, and it has continued to prove itself (since I have been known to forget even my own birthday, the need for this program was fairly obvious :-)

Make sure that you create a shortcut in your startup folder, and you will be reminded each time you log in or start your computer. When you first run the program, click on "Edit" quickly (the program will exit by itself in 10 seconds if there is nothing for that day). Read the help info in the supplied demo file, delete the things you don't want, and enjoy.

LM3915 - lm3915.zip (12,583 bytes) If you have checked the data sheet for the LM3915 LED bargraph display, you know just how irksome it is to calculate the resistor values to get the correct sensitivity and LED current. Well fret no more, as this tiny program will do the work for you. (See Project 60 for the circuit details)

BOGUS - bogus.zip (192,900 bytes) This program is designed for anyone who has a website - small or large. By generating complete web pages full of bogus random e-mail addresses, the idea is to choke the spam-bots to death, and pollute the spammers' databases to the extent that they become unusable. Contains it's own dictionary, and a custom wordlist of words and phrases to make interesting (and sometimes amusing) web pages that you can upload to your site. BOGUS is freeware, and redistribution is encouraged. For more info, see the article Death to all Spammers. Note - requires VB6 runtime library to operate.

Note: Updated - there was an error in the program that would stop it from working on any system other than mine. My apologies, and it has now been fixed.

TERMS AND CONDITIONS: The ESP programs shown here are distributed as Freeware unless noted otherwise, and as such may be freely given away. The software must not be modified or changed in any way and no fee is to be charged for redistribution. Software is believed to be bug and virus free, but it is the user's responsibility absolutely to use the software and accept all or any consequences from the use thereof. ESP accepts no liability or responsibility for data or other loss howsoever caused. It is the user's responsibility to scan for viruses before using any program.

Please note that the executables are 32 bit, and need VB40032.DLL to run. This should be installed as described above. Sorry, but the programs are only available for Windows95™ or above.

Spreadsheets

linkxfrm.zip An updated version of the excellent Linkwitz Transform spreadsheet from True Audio (www.trueaudio.com). This has had additions from Dean Canafranca (one of my readers) and I added the ability to use litres or cubic feet in the new section. This is reproduced with the kind permission of True Audio. The spreadsheet is in Microsoft Excel format, and is zipped to reduce the download time.

This spreadsheet is essential if you plan on building the Linkwitz transform circuit (Project 71), since it supplies the component values to achieve the desired response.

TERMS AND CONDITIONS: The Linkwitz Transform spreadsheet is the intellectual property of True Audio, and permission to re-publish or otherwise distribute the program may be granted only by True Audio.

---

ls-param.zip Theile-Small loudspeaker parameters made easy. Use this spreadsheet to determine all the Theile Small parameters, with a few simple measurements (See the article Measuring Loudspeaker Driver Parameters for the details of the tests.

---

heatsink.zip (Zipped archive) This is a calculator to allow you to determine the thermal rating of a heatsink, based on the size of the fins and base. It is passably accurate, although slightly pessimistic compared to manufacturer ratings (either that, or it is more realistic). Heatsink dimensions can be in inches or millimetres, and the total heatsink thermal resistance is in degrees C/W. The spreadsheet is in Microsoft Excel format, and is zipped to reduce download time.

---

xover.zip Design passive crossover networks with ease - includes zobel network for woofer inductance compensation, notch filter for tweeter resonance supression, and 6dB/Octave and 12dB/Octave Linkwitz-Riley aligned passive networks. Refer to the article Design of Passive Crossovers for full details.

Miscellaneous

SIMetrix Intro - Circuit Simulator - SIMetrix is a low cost SPICE analog circuit simulation package and schematic editor for Windows 95 and NT. A free ("intro") version of the software may be downloaded from this site. This is an excellent simulator - and especially so as freeware! Highly recommended.

Warning, this is a big download, at over 6MB

Main Index

Copyright Notice. All material described, including but not limited to all text and diagrams, are the intellectual property of Rod Elliott unless otherwise stated, and are Copyright © 1999-2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author / editor (Rod Elliott) grants the reader the right to use this information for personal use only. Commercial use in whole or in part is prohibited without express written authorisation from Rod Elliott or the identified copyright owner.

Update Information: 17 Jun 05 - moved application notes./ Apr 03 - added BOGUS./ 14 Sept - LM3915 calculator./ 13 Nov - reminder./ Nov 2001 - updated format./ Jan 2001 - added world time clock./ 08 Apr - heatsink calc./ 11 Mar - Download page created.

HEATSINKS

by Harry Lythall - SM0VPO

Introduction
Power and heat
Heatsink characteristics
Size of heatsink
Making heatsinks
Use existing heatsinks
How big do we need it?
Mounting devices
And finally ...

INTRODUCTION

Projects often require a lot of heat to be conducted away from a device generating it. Constructors often guess at the size of heatsink then build. I used to be that way myself, but in the late 70s I began using power transistors and the tears began. I had to educate myself. After a few experiments I developed 'rough' formulas for estimating the size of heatsink needed.

POWER AND HEAT

A silicon transistor should not normally be allowed to become much hotter than about 100°C. Most transistors will happily operate hotter, but do not count upon it.

Let us take the case of the simple PSU described earlier. The PSU may deliver 3 amperes at up to 15 volts. W=VA (Watts = Volts x Amperes). The load is therefore dissipating 3 x 15 = 45 watts. With a pre-regulated voltage of 24 volts, the regulator transistor has the remaining 9 volts accross it, also at 3 amperes. The transistor is therefore turning 27 watts (9v x 3A) into heat.

If you now turn the output voltage of the PSU DOWN to 3 volts, the problem gets much worse. The load is now taking 3 x 3 = 9 watts and the transistor is dissipating (24v - 3v) x 3A = 63 watts as heat. What do we do with all this heat? - use a heatsink, of course, but how big?

HEATSINK CHARACTERISTICS

Heatsinks are given a characteristic ' °C/Watt '. A TO5 transistor (for example) without a heatsink has a case heatsink characteristic of somewhere in the region of 220°C per watt. So for every watt you make it dissipate it will increase its temperature by 220°C, or about 240 - 250°C at normal room temperatures. That's hot enough to melt solder!

A clip-on heatsink with a characteristic of 40°C/W will limit that device to 40°C + 30°C room temperature = 70°C. With the heatsink the transistor will therefore survive. The heatsink simply conducts heat away from the device and dissipates it into the room, so the larger the surface area, the lower (better) the °C/W rating that heatsink has.

SIZE OF HEATSINK

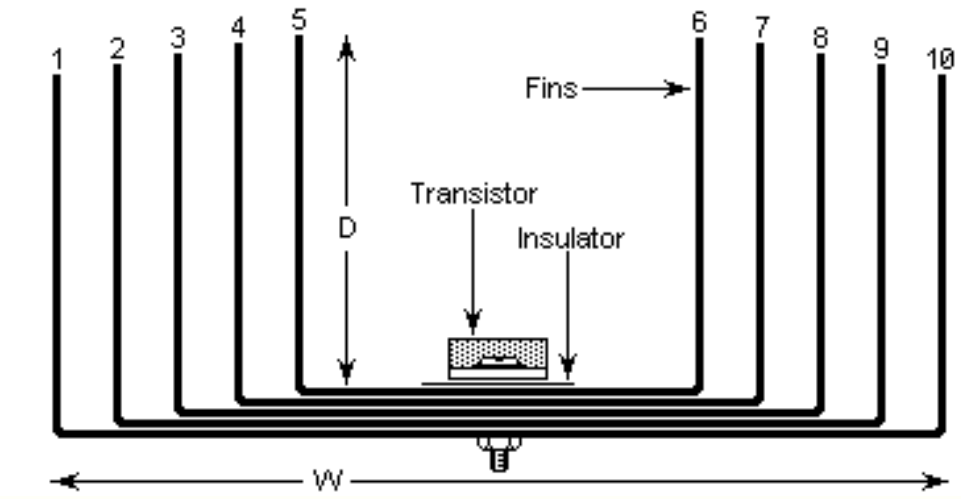
Our transistor in the PSU is dissipating 62 watts, so to keep the temperature below 100°C we should calculate the smallest heatsink required. The room temperature is assumed to be 30°C, so the transistor can increase its temperature by only 70°C. 70°C/62W = 1.13 so we will need a heatsink rated at 1.13°C/Watt, or less. Most component suppliers catalogues give a heatsinks °C/W characteristic.

MAKING HEATSINKS

You can evaluate home-made heatsinks by using the following 'ROUGH' formula:

$$\frac{50}{\sqrt{A}} = °C/watt \qquad A = \text{area (square CM)}$$

As an example, let us make a heatsink with 18 SWG aluminium sheet, folded as shown (viewed from above):



The heatsink is going to be 20cm wide (W), 10cm deep (D) and 12cm high (not shown). Each 'fin' is 10cm x 12cm = 120cm. Each fin also has two sides, area = 240 square cm. There are 10 fins so we have a total of 2400 square cm. The rear plate is also 2 sides x 20cm x 12cm = 480 square cm. Total area = 480 + 2400 = 2880 square cm.

Since $\frac{50}{\sqrt{A}} = °C/watt$ A = area (square CM) and the square root of 2880 = 53.66 we therefore have 50/53.66 = 0.932°C/watt. That would be an expensive heatsink to buy! It may not look pretty home-made, but it works just as well.

USE EXISTING HEAT SINKS

It is common practice to use the chassis of a unit as the heatsink, or part of it. An aluminium box 5cm x 10cm x 20cm would therefore have an OUTSIDE surface area of 5 x 10 x 20 = 1000 square cm. Square root of 1000 = 31.623 so our equipment case has a thermal dissipation of (50/31.623)°C/W = 1.58°C/W Since we need 1.13 we do not have enough, but we can use the case to cut down the size of the heatsink we need. Notice that the inside of the case was not taken into account. The case is a closed box so without ventilation the inside of the case will dissipate nothing.

HOW BIG DO WE NEED IT?

Ok, now let us work backwards using our PSU regulator transistor that is still dissipating over 60 watts without a heatsink. Let us mount it in the 5cm x 10cm x 20cm cabinet we found above. Transposing the formula we have:

$$A = \left(\frac{50}{°C/watt}\right)^2$$

We know we need a heatsink of 1.13°C/W so putting this into the formula we get 1957.9 square cm of heatsink required. We already have 1000 in the case, so we must provide the other 960 as fins on the back. 10 fins of 96 square cm (48 per side) will do the job. Each fin can therefore be 5cm x 10cm. Forget the backplane, that is part of the case. 5cm x 10cm is still quite big, but the case is only 5cm deep, so it is reasonable. Including heatsink, the case will now be 5cm x 20cm x 20cm. Note that only the fins were considered.

In reality, a 5cm tall case would be a little small so the case would not be able to accomodate the mains transformer, even if you used three 6.3v 3A fillament transformers. Increasing the height of the case by 1cm would reduce the size of the fins by 2cm; 6cm x 8cm (as opposed to 5cm x 10cm earlier).

MOUNTING DEVICES

Almost all devices that we need to mount on a heatsink have a metalic surface from which the heat is to be conducted. This surface is usually electrically connected to one of the terminals of the device. We therefore need to place an insulator between the device and the heat sink. In the case of transistors this is in the form of a mica washer or disk. The disk is lubricated with a heat conductive oil to help the heat to be transfered. The oil and the washer should never be omitted if the device is to be effectively protected.

Since aluminium is not the worlds best conductor of heat, I usually lay a strip of copper the full length of the heatsink. The device is mounted on the copper and the copper distributes the heat to the rest of the heatsink.

AND FINALLY ...

There are such things a temperature gradients, junction to case thermal resistance and many other parameters that could be investigated. If the subject of heatsinks were an English country mansion set in 110 acres of garden maintained by over 50 servants, I have just shown you the doorbell, or just enough to get you in the door (and out of trouble).

So, that is about all I want to say about heatsinking. Do have fun, but don't get your fingers burned! Best regards Harry, Lunda, Sweden.

Return to INFO page

International Rectifier

THE POWER MANAGEMENT LEADER

PRODUCT LINE | SALES SUPPORT | TECHNICAL LIBRARY | DESIGN SUPPORT | LOGIN/MY PAGE

Three Phase PWM controller with DirectFET MOSFET pair deliver up to 80A in small size.

80A POL regulator that takes up less than two square inches of board space, ideal for space-constrained applications like DDR memory in rack servers.
more...

IR Introduces Mid-Voltage MOSFETs Optimized for Switching Converter Applications in Netcom Systems more...

IR's New 30V HEXFET® MOSFETs Deliver 1.5% Improvement in Light-Load DC-DC Buck Converter Applications more...

IR Trench IGBTs Reduce Power Dissipation Up to 60% in Motion Control Applications more...

FEATURED PRODUCT

Mid-Voltage MOSFETs Optimized for Switching Converter Applications in Netcom Systems

More News

SEARCH

GO

Part Search

Site Search

APPLICATIONS

AC-DC
Appliances
Audio
Automotive
Die Products
DC-DC
Desktop/Server
HiRel
Lighting
Motion Control
NetCom
Portables

COMPANY INFO

Press Room
Tradeshows
Webcasts
E-Mail News
RSS Feeds
Site Index
Careers
Contact Us
Environmental
Epi Services
Investor Relations

PRODUCT LINE

Product Tree
Parametric Search
HEXFET Power MOSFETs
Integrated Circuits
High Voltage Gate Drivers
G5 HVIC
IGBTs
Diodes, Bridges, SCRs
Die Products
Modules
HiRel

SALES SUPPORT

How to Buy Products
Part Search
Inventory Availability
Cross Reference
Product Change Notices
Manufacturing Sites
Sales Representative
Distributors

TECHNICAL LIBRARY

Application Training Guide
Application Notes
Technical Papers
Technical Support
Search FAQs
Ask a Question Online
Track your Questions

DESIGN SUPPORT

Design Services
myPower
Reference Designs
Models
Design Tips
Packaging Index
Lead Free
Quality and Reliability

IR: International Sites | China | Japan |

About International Rectifier | Contact Us | Privacy

©1995-2006 International Rectifier

Application Note AN-1000

Mounting Guidelines for the SUPER-220

By Andrew Sawle and Arthur Woodworth

Table of Contents

Page

Thermal Resistance and Its Effects on System Performance 2
Clip Mounting of the Super-220™ Package 3
Why and how does contact force affect thermal resistance? 3
What is the minimum contact force that should be applied to gain good thermal contact? 4
Maximum Contact Force 4
Contact Conditions 4
Interface Materials 5
Heatsink Compound 5
Interface Pads 6
Phase Change Materials 6
Heatsinks and Clips 7
Saddle Clips 7
'U' Clips 8
Extrusion Mounted Clips 8
Typical Clip Solution Suppliers 9

This Application Note is intended to outline the thermal options available to designers with regard to interface materials, clip types and the contact forces required to give good thermal contact and the performance that they might expect from each. The following topics will be covered:

- A brief revision of thermal resistance and its effects on system performance.
- Clip mounting of SUPER-220 devices to heat sinks, suppliers and interface materials

Mounting Guidelines for the SUPER-220

by Andrew Sawle and Arthur Woodworth

The impressive gains in power density born from the development of the 'SUPER' type power packages, *i.e. the SUPER-247*, has lead to a similar development of the TO-220 package. The removal of the hole and incorporation of new innovative ideas has lead to the SUPER-220. This new power package can now house areas of silicon that previously would have called for the use of a TO-247. These increases in power density allow designers to realise cost savings through reducing the number of paralleled devices required for a desired current rating and also reducing the areas of costly circuit board needed for their designs. Hence reductions in both the size and cost of designs are possible through the use of the SUPER-220 package.

However, increasing the power density within a design will mean that the designer needs to make careful consideration of the thermal aspects of the design aswell as the electrical. As with the SUPER-247 (see Application Note AN997) the SUPER-220 does not have a screw hole to facilitate bolting to a heatsink and therefore the package will need to be clip mounted. Clip mounting facilitates mass production techniques and good thermal performance but the latter will only be the case if the designer has made careful use of the thermal options available.

This Application Note is intended to outline the thermal options available to designers with regard to interface materials, clip types and the contact forces required to give good thermal contact and the performance that they might expect from each. The following topics will be covered:

- ➜ A brief revision of thermal resistance and its effects on system perfomance.
- ➜ Clip mounting of SUPER-220 devices to heatsinks, suppliers and interface materials.

Thermal Resistance and its Effects on System Performance

The thermal resistance ($Rth_{(junction\text{-}ambient)}$) of a system (system = package through to heatsink) is made up from a number of component parts as shown in Figure 1.1 below.

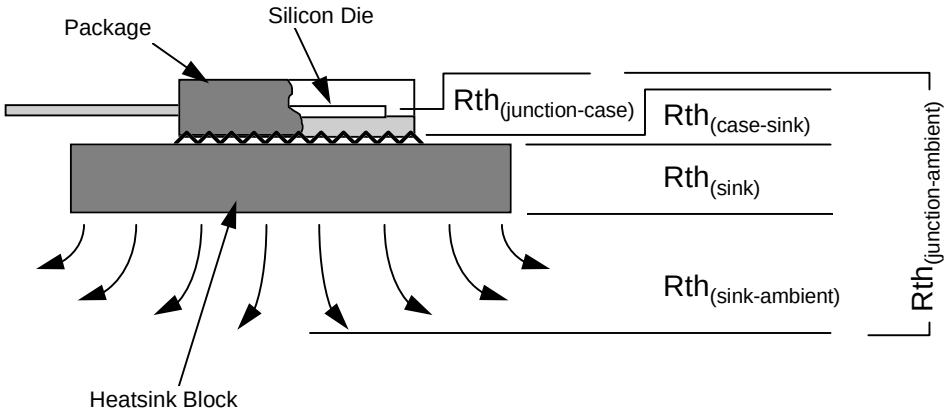


Figure 1.1 - Build-up of Thermal Resistance in a System

The designer of a system has varying amounts of influence over the component parts of the overall thermal resistance of his design:

- **$Rth_{(junction-case)}$** - this has been determined during the design and manufacture of the product. The system designer has no direct influence.
- **$Rth_{(case-sink)}$ (or Contact Thermal Resistance)** - determined by the size and quality of the contact areas between the package and the sink, the use of intermediate materials and the contact pressure. Hence, the system designer can have a large influence over this parameter.
- **$Rth_{(sink)}$ and $Rth_{(sink-ambient)}$** - determined by heatsink design, *i.e. material and shape.* The selection of a heatsink is done by trading off performance and cost requirements.

The thermal resistance of a system will determine the amount of power that can be dissipated by a device without exceeding the maximum junction temperature. Operating at temperatures above $T_{j(max)}$ will reduce the reliability of the device. Therefore by optimising the thermal performance of a system, a greater amount of power can be dissipated by each device thereby minimising the number of devices that need to be used.

Clip Mounting of the SUPER-220 Package

The removal of the hole from the TO-220 package to facilitate the greater area of silicon in the SUPER-220 package means that the new package can **NOT** be bolted to a heatsink. However the package lends itself to clip mounting, a technology that is quickly becoming popular as it holds a number of advantages over the older 'bolting' technology:

- Clips are quick to apply and therefore lend themselves to mass production assembly techniques.
- Clips maintain a constant assembly force even if the intermediate material deteriorates, unlike bolt assembly where the force diminishes.
- Clips impart their force on the central part of the package over the silicon, distributing the pressure evenly which creates a good thermal contact with the heatsink. Bolts however, impart the force at one end of the package and the pressure is uneven, see Figure 1.2.

Application of force off-centre (i.e. bolting a device to heatsink) leads to uneven themal contact. Using clip mounting ensures that the force is applied above the silicon and that the thermal contact is good.

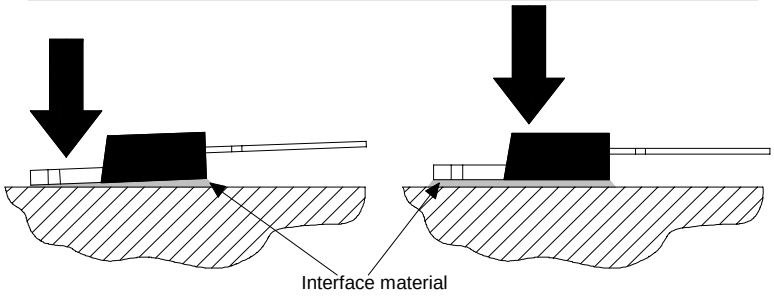


Figure 1.2 - Position of Force Application

The contact force and the way in which it is applied is therefore very important to the thermal performance of a system.

Why and How Does Contact Force Affect Thermal Resistance?

Package cases and heatsink surfaces are never perfectly flat. Hence contact between the two will only occur at several points allowing an air gap between the surfaces (as illustrated in Figure 1.3). Since air is a very good thermal insulator this means that the contact thermal resistance is much greater than it would be if the two surfaces were in perfect contact (no air gap). However, as the contact force (pushing the two surfaces together) increases then so will the number of points at which the two surfaces contact one another and the air-gap will be reduced, in turn reducing the contact thermal resistance.

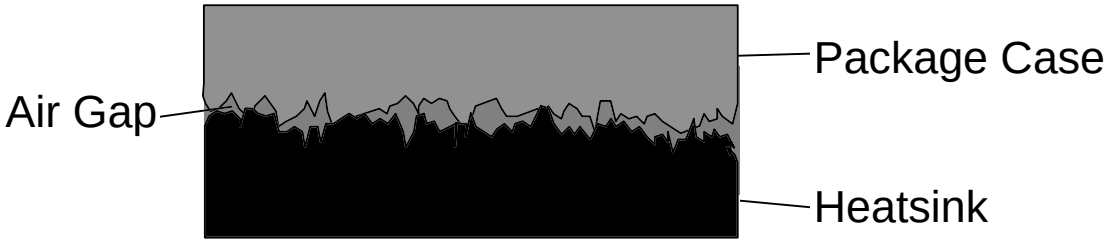


Figure 1.3 - Diagram Showing the Effect (Under High Magnification) When Two Non-perfect Surfaces Meet.

What is the minimum contact force that should be applied to gain good thermal contact?
The minimum force required for a good thermal contact will vary from package to package as it will depend on the surface finish, flatness of contact surface and area of contact. Contact thermal resistance reduces with increased contact force but the relationship is non-linear. In essence, diminishing returns in reduced thermal resistance are seen for increases in force. This relationship for the SUPER-220 on a heatsink with NO interface materials, *i.e. 'dry' conditions*, is shown in Figure 1.4. A minimum contact force of 20N is recommended.

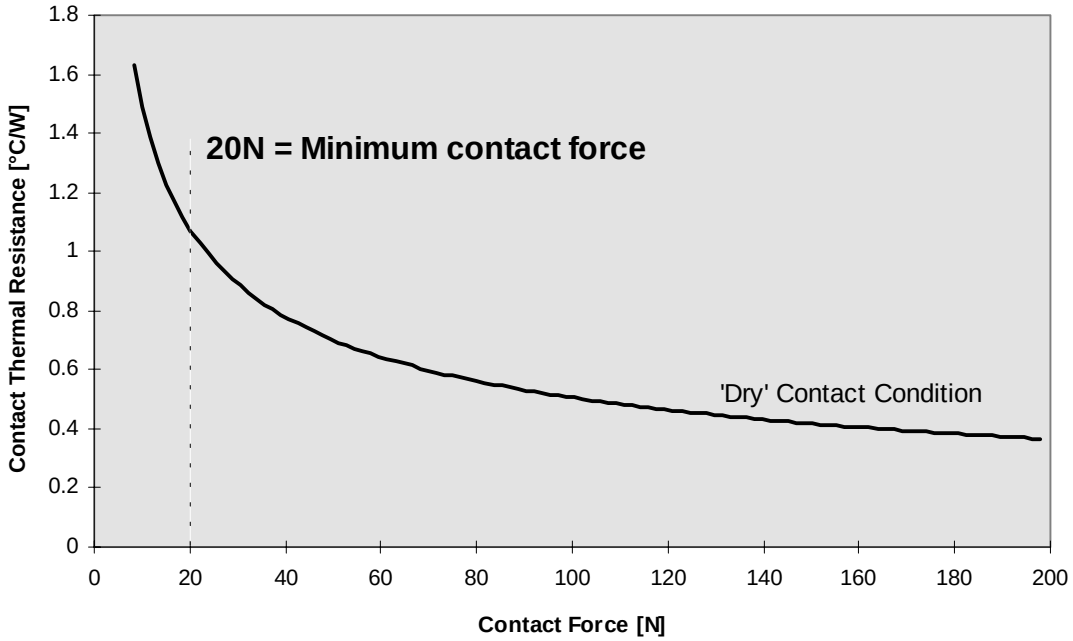


Figure 1.4 - Minimum Contact Force

Maximum Contact Force
The minimum contact force of 20N mentioned above is purely that, the MINIMUM force. Any force applied above that figure will still show gains in reduced contact thermal resistance until the maximum force that the package can withstand before the device characteristics are altered or the package is destroyed. This maximum limit figure has been measured to be 200N TYP. However, these gains are not free, for in general terms a greater contact force means a larger, more expensive clamping system. A contact force should therefore be chosen that optimises both the thermal and the cost requirements of the system.

Contact Conditions
As mentioned previously, the contact conditions between the package and the heatsink will affect the contact thermal resistance. Contact conditions encompass a number of areas including: surface roughness, surface cleanliness, paint finishes and intermediate materials. The surface roughness of the heatsinking material should be no greater than 0.02mm over the area where the device is to be mounted. Surface cleanliness during assembly of package and heatsink is imperative, even if a thermal grease or other material is subsequently added. Unclean

surfaces can be held apart by dirt or grease thus increasing the thermal resistance. However, normal paint finishes (up to 50 mm thick) have been shown to have little effect on thermal resistance, this therefore leaves intermediate materials as an area for discussion.

Interface Materials

Figure 1.4 above shows the force vs Rth relationship for ‘dry’ conditions, however, the use of no interface materials in an assembly is rare and usually one of two conditions is required:

- An interface material to improve the thermal performace of the contact area between device and heatsink.
- An interface material providing electrical isolation between the device and heatsink whilst minimising the adverse effects that it will have on the the thermal performance of the contact area. Isolation is usually required where more than one device is going to share the same heatsink.

Heatsink compound

A number of different companies offer heatsinking compound, these usually consist of silicon grease loaded with some electrically insulating, good thermally conducting material such as alumina. Thinly applied, these compounds are advantageous as they fill the air gaps and do not further increase the distances between the surfaces. Thickly applied they can hold the two surfaces apart and increase the contact thermal resistance. The following graph displayed in Figure 1.5 shows the contact thermal resistance for ‘dry’ conditions (no compound) and the thermal resistance using compound, both are plotted against contact force.

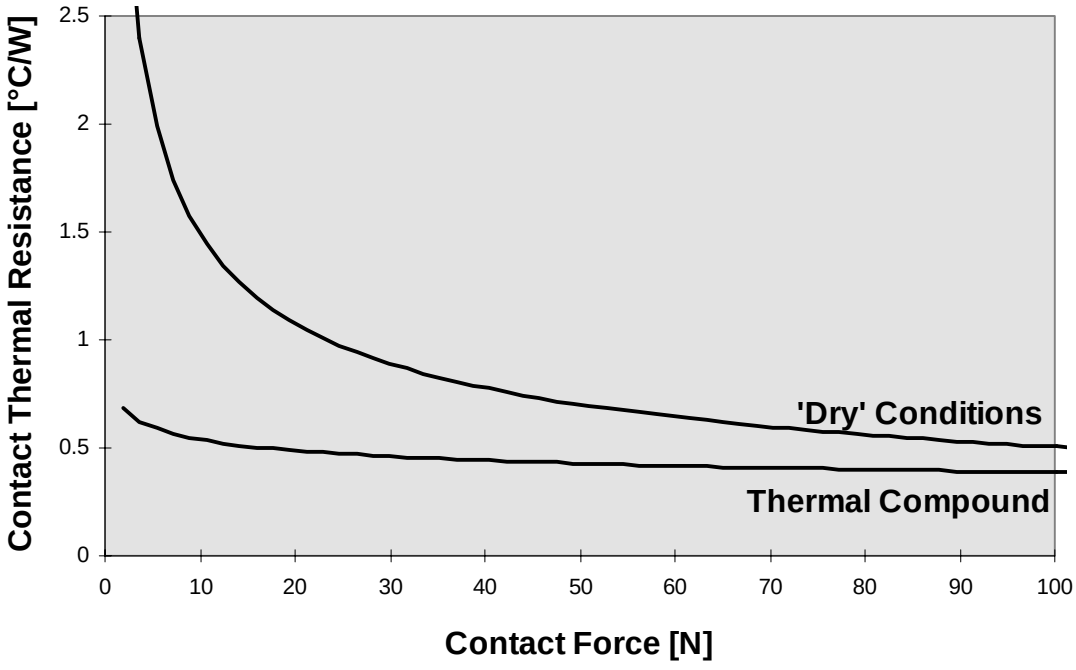


Figure 1.5 - Contact Thermal Resistance for ‘Dry’ and Compound Conditions, Plotted Against Contact Force

Interface Pads

The metal heatspreader on the back of the SUPER-220 package is non-electrically isolated from the pinouts of the device within, *i.e. the case is the drain contact for a MOSFET.* Hence in cases where devices are not electrically paralleled but share the same heatsink, it is necessary to insert an electrically isolating material between the package and the heatsink block. The isolator usually takes the form of a pad and many companies offer a range of pad material types and sizes dependant on requirements.

The pad obviously has a direct and detremental effect on the contact resistance as insertion adds an extra resistance into the build-up. Again, the contact thermal resistance is dependant on contact pressure. The following graph, shown in Figure 1.6, illustrates the higher thermal resistance when using an isolator pad. The isolator used for the line plot was a typical silicone loaded pad. Therefore, it should be noted that when electrically isolating a device from a heatsink the thermal resistance of the system will increase.

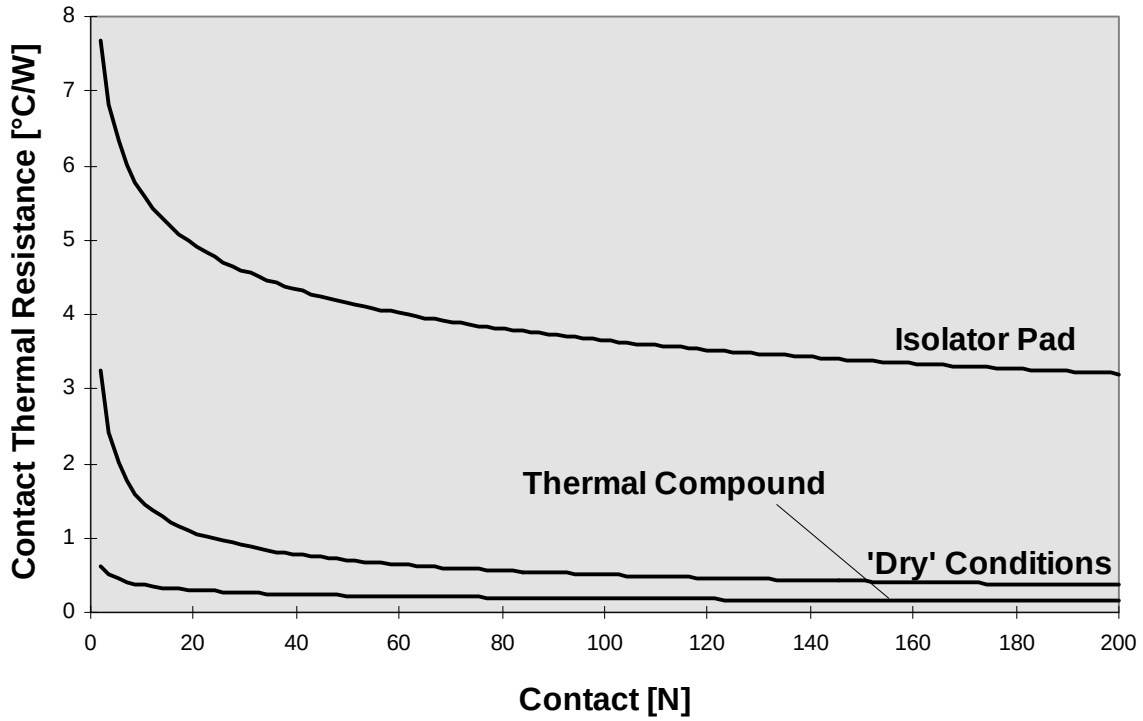


Figure 1.6 - Contact Thermal Resistance Plotted Against Contact Force

Phase Change Materials

One of the problems that interface pads have is that the forces required for good thermal contact are very high. This is because it is difficult to gain the amount of pad deformation necessary to fill the air-gaps in the heatspreader/pad and pad/heatsink interfaces. One of the solutions being offered by a number of suppliers is that of PHASE-CHANGE materials.

Phase change materials are a solid in nature until raised above a certain temperature (~50-65°C) at which point they become a viscous fluid, this change from solid to viscous occurs every time the temperature goes to this level. This fluidity means that with an adequate contact force the air-gaps can easily be filled and good thermal contact can be made. This good contact is maintained once the temperature is reduced as the material solidifies in place, filling the gaps.

The following chart in Figure 1.7 shows the contact thermal resistance using a typical phase-change isolating pad and a non-isolating pad.

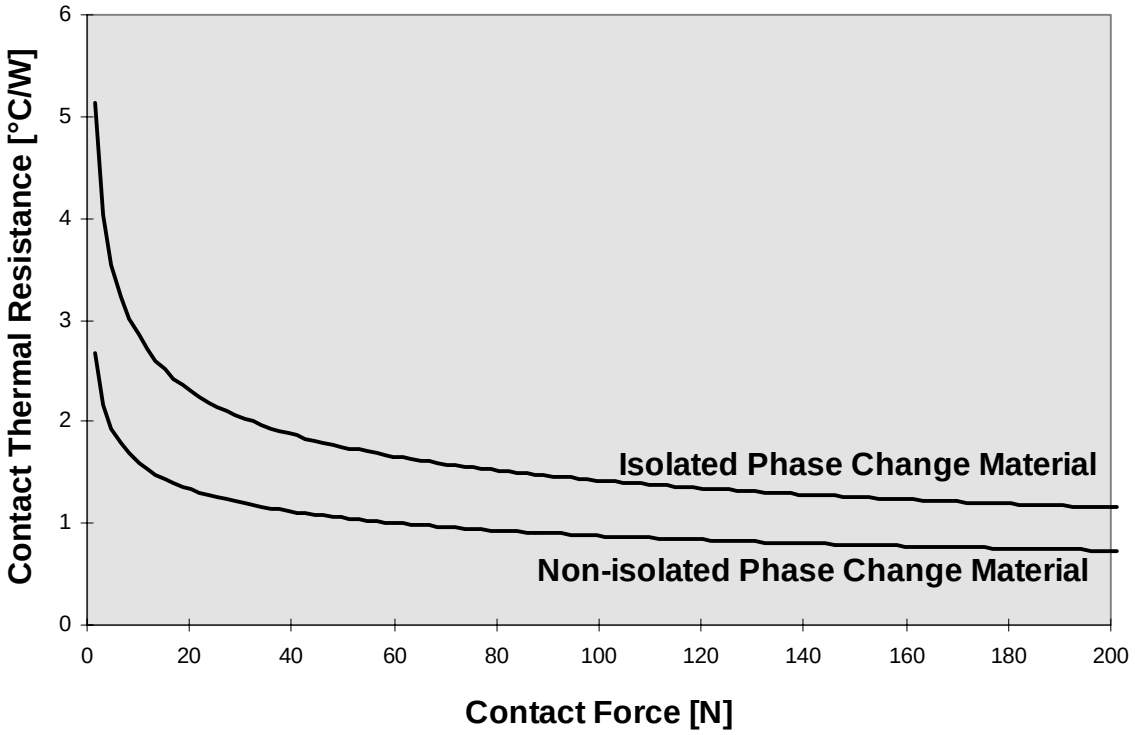


Figure 1.7 - Contact Thermal Resistance for Isolated and Non-isolated Phase Change Materials

It can be seen that phase change materials can offer lower contact thermal resistances when isolation is required over regular isolation pads. Two current manufacturers/suppliers of phase change materials are **Berquist** and **Orcus Inc**.

Heatsinks and Clips
A range of heatsinks and clips have been developed by a number of suppliers for the SUPER-220 package type. A number of these clip types are shown in the following diagrams. However, the clips shown are only a small selection and by no means encompass all available types and solutions.

Saddle Clips
An example of a SUPER-220 mounted to a heatsink using a saddle clip is shown in Figure 1.8.

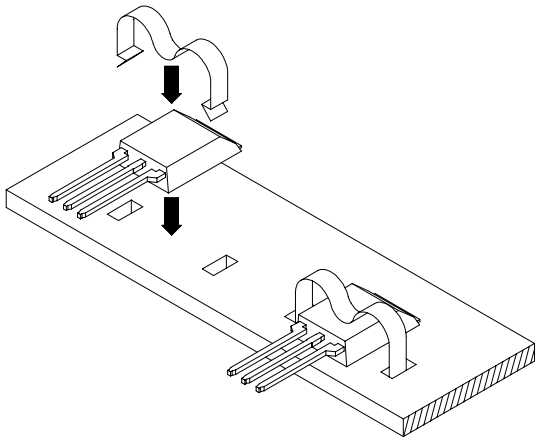


Figure 1.8 - SUPER-220 Mounted to Heatsink Using Saddle Clip

When using saddle clip type solutions, the heatsink materials are thin in cross-section, typically less than 5mm in thickness. The clips push into holes cut into heatsink material and lock against the back face of the heatsink. These clips produce contact forces in the range 15-50N.

'U' Clips
An example of a heatsink assembly utilising a 'U' clip is shown in Figure 1.9.

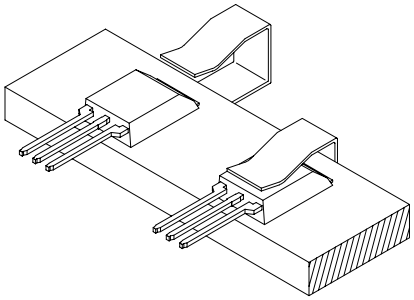


Figure 1.9 - SUPER-220 and a 'U' Clip Assembly

This clip type clamps the device and the heatsink material together. Variations on this clip type allow devices to be clamped to the front and back of the heatsink block using the same clip. 'U' clips can offer contact forces in the range 15-50N.

Extrusion Mounted Clips
There are a number of proprietary clip solutions offered where the clip is anchored in a feature in an extruded heatsink. One such solution is to use rails on extruded sinks for clip anchorage. This method is shown in Figure 1.10.

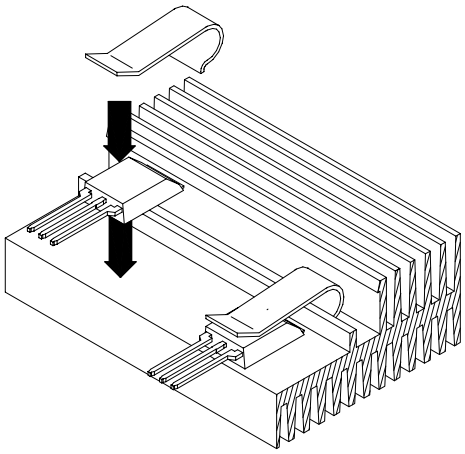


Figure 1.10 - Extrusion Mounted Clip (Rail Anchorage)

Clips on extruded sinks can offer contact forces in the range 25-50N

Typical Clip Solution Suppliers

Redpoint Thermalloy
Cheney Manor,
Swindon,
Wiltshire. SN2 2QN
England

Avvid Thermal Technologies
Corporate Headquarters,
One Kool Path,
Laconia,
New Hampshire, USA

Austerlitz Electronic gmbh
Ludwig Feuerbach-Straße 38, 90961
Nüenburg,
Postfach 1048, GERMANY

Thermalloy, Inc.
2021 W. Valley View Ln.,
Dallas, TX 75234
USA

AN1040/D

Mounting Considerations For Power Semiconductors

ON Semiconductor™

http://onsemi.com

Prepared by: Bill Roehr

APPLICATION NOTE

INTRODUCTION

Current and power ratings of semiconductors are inseparably linked to their thermal environment. Except for lead–mounted parts used at low currents, a heat exchanger is required to prevent the junction temperature from exceeding its rated limit, thereby running the risk of a high failure rate. Furthermore, the semiconductor industry's field history indicated that the failure rate of most silicon semiconductors decreases approximately by one–half for a decrease in junction temperature from 160°C to 135°C.(1) Guidelines for designers of military power supplies impose a 110°C limit upon junction temperature.(2) Proper mounting minimizes the temperature gradient between the semiconductor case and the heat exchanger.

Most early life field failures of power semiconductors can be traced to faulty mounting procedures. With metal packaged devices, faulty mounting generally causes unnecessarily high junction temperature, resulting in reduced component lifetime, although mechanical damage has occurred on occasion from improperly mounting to a warped surface. With the widespread use of various plastic–packaged semiconductors, the prospect of mechanical damage is very significant. Mechanical damage can impair the case moisture resistance or crack the semiconductor die.

Figure 1 shows an example of doing nearly everything wrong. A tab mount TO–220 package is shown being used as a replacement for a TO–213AA (TO–66) part which was socket mounted. To use the socket, the leads are bent – an operation which, if not properly done, can crack the package, break the internal bonding wires, or crack the die. The package is fastened with a sheet–metal screw through a 1/4″ hole containing a fiber–insulating sleeve. The force used to tighten the screw tends to pull the package into the hole, possibly causing enough distortion to crack the die. In addition, the contact area is small because of the area consumed by the large hole and the bowing of the package; the result is a much higher junction temperature than expected. If a rough heatsink surface and/or burrs around the hole were displayed in the illustration, most but not all, poor mounting practices would be covered.

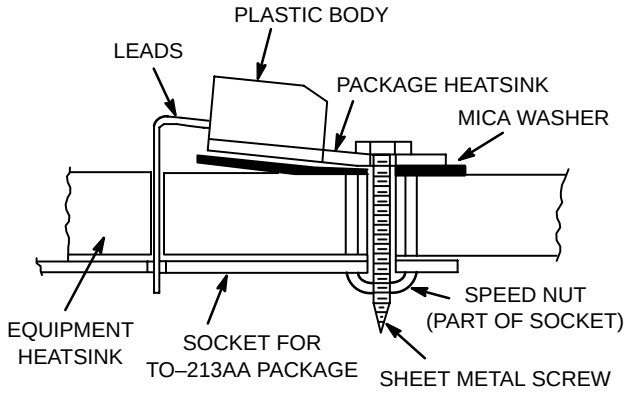


Figure 1. Extreme Case of Improperly Mounting a Semiconductor (Distortion Exaggerated)

In many situations, the case of the semiconductor must be electrically isolated from its mounting surface. The isolation material is, to some extent, a thermal isolator as well, which raises junction operating temperatures. In addition, the possibility of arc–over problems is introduced if high voltages are present. Various regulating agencies also impose creepage distance specifications which further complicates design. Electrical isolation thus places additional demands upon the mounting procedure.

Proper mounting procedures usually necessitate orderly attention to the following:

1. Preparing the mounting surface
2. Applying a thermal grease (if required)
3. Installing the insulator (if electrical isolation is desired)
4. Fastening the assembly
5. Connecting the terminals to the circuit

© Semiconductor Components Industries, LLC, 2001

In this note, mounting procedures are discussed in general terms for several generic classes of packages. As newer packages are developed, it is probable that they will fit into the generic classes discussed in this note. Unique requirements are given on data sheets pertaining to the particular package. The following classes are defined:

Stud Mount
Flange Mount
Pressfit
Plastic Body Mount
Tab Mount
Surface Mount

Appendix A contains a brief review of thermal resistance concepts. Appendix B discusses measurement difficulties with interface thermal resistance tests. Appendix C indicates the type of accessories supplied by a number of manufacturers.

MOUNTING SURFACE PREPARATION

In general, the heatsink mounting surface should have a flatness and finish comparable to that of the semiconductor package. In lower power applications, the heatsink surface is satisfactory if it appears flat against a straight edge and is free from deep scratches. In high–power applications, a more detailed examination of the surface is required. Mounting holes and surface treatment must also be considered.

Surface Flatness

Surface flatness is determined by comparing the variance in height (Δh) of the test specimen to that of a reference standard as indicated in Figure 2. Flatness is normally specified as a fraction of the Total Indicator Reading (TIR). The mounting surface flatness, i.e, Δh/TlR, if less than 4 mils per inch, normal for extruded aluminum, is satisfactory in most cases.

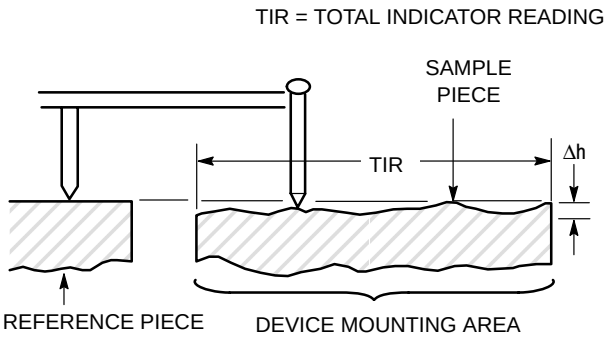


Figure 2. Surface Flatness Measurement

Surface Finish

Surface finish is the average of the deviations both above and below the mean value of surface height. For minimum interface resistance, a finish in the range of 50 to 60 microinches is satisfactory; a finer finish is costly to achieve and does not significantly lower contact resistance. Tests conducted by Thermalloy, Inc., using a copper TO–204 (TO–3) package with a typical 32–microinch finish, showed that heatsink finishes between 16 and 64 μ–in caused less than ± 2.5% difference in interface thermal resistance when the voids and scratches were filled with a thermal joint compound.[3] Most commercially available cast or extruded heatsinks will require spotfacing when used in high–power applications. In general, milled or machined surfaces are satisfactory if prepared with tools in good working condition.

Mounting Holes

Mounting holes generally should only be large enough to allow clearance of the fastener. The larger thick flange type packages having mounting holes removed from the semiconductor die location, such as the TO–3, may successfully be used with larger holes to accommodate an insulating bushing, but many plastic encapsulated packages are intolerant of this condition. For these packages, a smaller screw size must be used such that the hole for the bushing does not exceed the hole in the package.

Punched mounting holes have been a source of trouble because if not properly done, the area around a punched hole is depressed in the process. This “crater” in the heatsink around the mounting hole can cause two problems. The device can be damaged by distortion of the package as the mounting pressure attempts to conform it to the shape of the heatsink indentation, or the device may only bridge the crater and leave a significant percentage of its heat–dissipating surface out of contact with the heatsink. The first effect may often be detected immediately by visual cracks in the package (if plastic), but usually an unnatural stress is imposed, which results in an early–life failure. The second effect results in hotter operation and is not manifested until much later.

Although punched holes are seldom acceptable in the relatively thick material used for extruded aluminum heatsinks, several manufacturers are capable of properly utilizing the capabilities inherent in both fine–edge blanking or sheared–through holes when applied to sheet metal as commonly used for stamped heatsinks. The holes are pierced using Class A progressive dies mounted on four–post die sets equipped with proper pressure pads and holding fixtures.

When mounting holes are drilled, a general practice with extruded aluminum, surface cleanup is important. Chamfers must be avoided because they reduce heat transfer surface and increase mounting stress. However, the edges must be broken to remove burrs which cause poor contact between device and heatsink and may puncture isolation material.

Surface Treatment

Many aluminum heatsinks are black–anodized to improve radiation ability and prevent corrosion. Anodizing results in significant electrical, but negligible thermal insulation; it need only be removed from the mounting area when electrical contact is required. Heatsinks are also available which have a nickel plated copper insert under the semiconductor mounting area. No treatment of this surface is necessary.

Another treated aluminum finish is iridite, or chromateacid dip, which offers low resistance because of its thin surface, yet has good electrical properties because it resists oxidation. It need only be cleaned of the oils and films that collect in the manufacture and storage of the sinks, a practice which should be applied to all heatsinks.

For economy, paint is sometimes used for sinks; removal of the paint where the semiconductor is attached is usually required because of paint's high thermal resistance. However, when it is necessary to insulate the semiconductor package from the heatsink, hard anodized or painted surfaces allow an easy installation for low voltage applications. Some manufacturers will provide anodized or painted surfaces meeting specific insulation voltage requirements, usually up to 400 volts.

It is also necessary that the surface be free from all foreign material, film, and oxide (freshly bared aluminum forms an oxide layer in a few seconds). Immediately prior to assembly, it is a good practice to polish the mounting area with No. 000 steel wool, followed by an acetone or alcohol rinse.

INTERFACE DECISIONS

When any significant amount of power is being dissipated, something must be done to fill the air voids between mating surfaces in the thermal path. Otherwise, the interface thermal resistance will be unnecessarily high and quite dependent upon the surface finishes.

For several years, thermal joint compounds, often called grease, have been used in the interface. They have a resistivity of approximately 60°C/W/in whereas air has 1200°C/W/in. Since surfaces are highly pock–marked with minute voids, use of a compound makes a significant reduction in the interface thermal resistance of the joint. However, the grease causes a number of problems, as discussed in the following section.

To avoid using grease, manufacturers have developed dry conductive and insulating pads to replace the more traditional materials. These pads are conformal, therefore, they partially fill voids when under pressure.

Thermal Compounds (Grease)

Joint compounds are a formulation of fine zinc or other conductive particles in a silicone oil or other synthetic base fluid which maintains a grease–like consistency with time and temperature. Since some of these compounds do not spread well, they should be evenly applied in a very thin layer using a spatula or lintless brush, and wiped lightly to remove excess material. Some cyclic rotation of the package will help the compound spread evenly over the entire contact area. Some experimentation is necessary to determine the correct quantity; too little will not fill all the voids, while too much may permit some compound to remain between well–mated metal surfaces where it will substantially increase the thermal resistance of the joint.

To determine the correct amount, several semiconductor samples and heatsinks should be assembled with different amounts of grease applied evenly to one side of each mating surface. When the amount is correct, a very small amount of grease should appear around the perimeter of each mating surface as the assembly is slowly torqued to the recommended value. Examination of a dismantled assembly should reveal even wetting across each mating surface. In production, assemblers should be trained to slowly apply the specified torque even though an excessive amount of grease appears at the edges of mating surfaces. Insufficient torque causes a significant increase in the thermal resistance of the interface.

To prevent accumulation of airborne particulate matter, excess compound should be wiped away using a cloth moistened with acetone or alcohol. These solvents should not contact plastic–encapsulated devices, as they may enter the package and cause a leakage path or carry in substances which might attack the semiconductor chip.

The silicone oil used in most greases has been found to evaporate from hot surfaces with time and become deposited on other cooler surfaces. Consequently, manufacturers must determine whether a microscopically thin coating of silicone oil on the entire assembly will pose any problems. It may be necessary to enclose components using grease. The newer synthetic base greases show far less tendency to migrate or creep than those made with a silicone oil base. However, their currently observed working temperature range is less, they are slightly poorer on thermal conductivity and dielectric strength and their cost is higher.

Data showing the effect of compounds on several package types under different mounting conditions is shown in Table 1. The rougher the surface, the more valuable the grease becomes in lowering contact resistance; therefore, when mica insulating washers are used, use of grease is generally mandatory. The joint compound also improves the breakdown rating of the insulator.

Conductive Pads

Because of the difficulty of assembly using grease and the evaporation problem, some equipment manufacturers will not, or cannot, use grease. To minimize the need for grease, several vendors offer dry conductive pads which approximate performance obtained with grease. Data for a greased bare joint and a joint using Grafoil®, a dry graphite compound, is shown in the data of Figure 3 through Figure 6. Grafoil is claimed to be a replacement for grease when no electrical isolation is required; the data indicates it does indeed perform as well as grease. Another conductive pad available from Aavid is called Kon–Dux™. It is made with a unique, grain oriented, flake–like structure (patent pending). Highly compressible, it becomes formed to the surface roughness of both the heatsink and semiconductor. Manufacturer's data shows it to provide an interface thermal resistance better than a metal interface with filled silicone grease. Similar dry conductive pads are available from other manufacturers. They are a fairly recent development; long term problems, if they exist, have not yet become evident.

Table 1. Approximate Values for Interface Thermal Resistance Data from Measurements Performed in ON Semiconductor Applications Engineering Laboratory

Dry interface values are subject to wide variation because of extreme dependence upon surface conditions. Unless otherwise noted, the case temperature is monitored by a thermocouple located directly under the die reached through a hole in the heatsink. (See Appendix B for a discussion of Interface Thermal Resistance Measurements.)

Package Type and Data		Interface Thermal Resistance (°C/W)						
			Metal–to–Metal		With Insulator			
JEDEC Outlines	Description	Test Torque In–Lb	Dry	Lubed	Dry	Lubed	Type	See Note
DO–203AA, TO–210AA TO–208AB	10–32 Stud 7/16″ Hex	15	0.3	0.2	1.6	0.8	3 mil Mica	
DO–203AB, TO–210AC TO–208	1/4–28 Stud 11/16″ Hex	25	0.2	0.1	0.8	0.6	5 mil Mica	
DO–208AA	Pressfit, 1/2″	–	0.15	0.1	–	–	–	
TO–204AA (TO–3)	Diamond Flange	6	0.5	0.1	1.3	0.36	3 mil Mica	1
TO–213AA (TO–66)	Diamond Flange	6	1.5	0.5	2.3	0.9	2 mil Mica	
TO–126	Thermopad 1/4″ x 3/8″	6	2.0	1.3	4.3	3.3	2 mil Mica	
TO–220AB	Thermowatt	8	1.2	1.0	3.4	1.6	2 mil Mica	1, 2

NOTES: 1. See Figure 3 through Figure 7 for additional data on TO–3 and TO–220 packages.
2. Screw not insulated. See Figure 20.

INSULATION CONSIDERATIONS

Since most power semiconductors use vertical device construction, it is common to manufacture power semiconductors with the output electrode (anode, collector or drain) electrically common to the case; the problem of isolating this terminal from ground is a common one. For lowest overall thermal resistance, which is quite important when high power must be dissipated, it is best to isolate the entire heatsink/semiconductor structure from ground, rather than to use an insulator between the semiconductor and the heatsink. Heatsink isolation is not always possible, however, because of EMI requirements, safety reasons, instances where a chassis serves as a heatsink or where a heatsink is common to several non–isolated packages. In these situations, insulators are used to isolate the individual components from the heatsink. Newer packages, such as the ON Semiconductor FULLPAK™ and EMS modules, contain the electrical isolation material within, thereby saving the equipment manufacturer the burden of addressing the isolation problem.

Insulator Thermal Resistance

When an insulator is used, thermal grease is of greater importance than with a metal–to–metal contact, because two interfaces exist instead of one and some materials, such as mica, have a hard, markedly uneven surface. With many isolation materials, reduction of interface thermal resistance of between 2 to 1 and 3 to 1 are typical when grease is used.

Data obtained by Thermalloy, showing interface resistance for different insulators and torques applied to TO–204 (TO–3) and TO–220 packages, are shown in Figure 3 through Figure 6, for bare and greased surfaces. Similar materials to those shown are available from several manufacturers. It is obvious that with some arrangements, the interface thermal resistance exceeds that of the semiconductor (junction to case).

Referring to Figure 3 through Figure 6, one may conclude that when high power is handled, beryllium oxide is unquestionably the best. However, it is an expensive choice. (It should not be cut or abraded, as the dust is highly toxic.) Thermafilm® is a filled polymide material which is used for isolation (variation of Kapton®). It is a popular material for low power applications because of its low cost ability to withstand high temperatures, and ease of handling in contrast to mica which chips and flakes easily.

A number of other insulating materials are also shown. They cover a wide range of insulation resistance, thermal resistance and ease of handling. Mica has been widely used in the past because it offers high breakdown voltage and fairly low thermal resistance at a low cost, but it certainly should be used with grease.

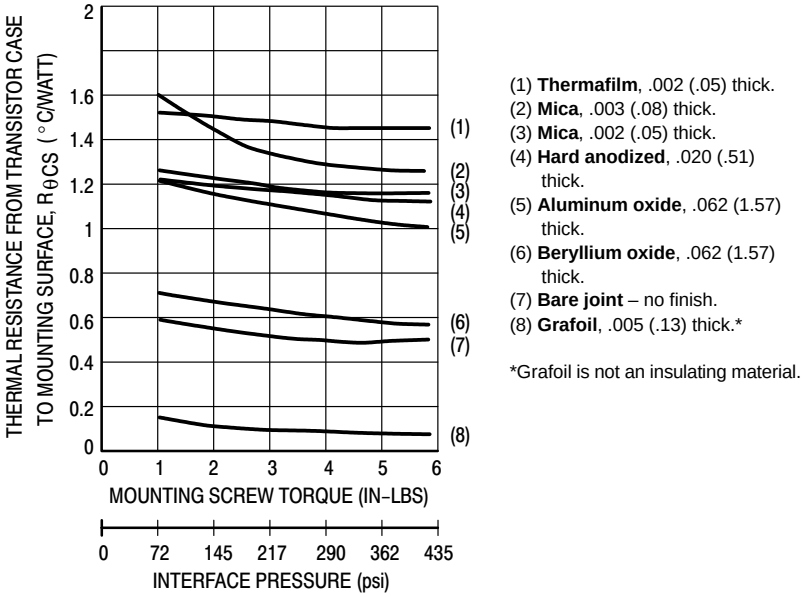


Figure 3. TO–204AA (TO–3) Without Thermal Grease

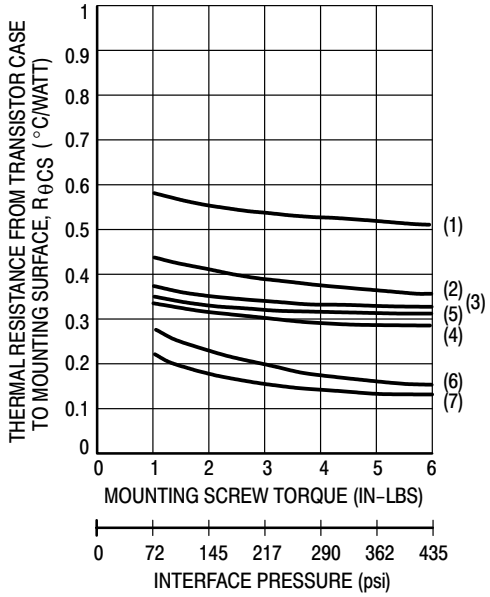


Figure 4. TO–204AA (TO–3) With Thermal Grease

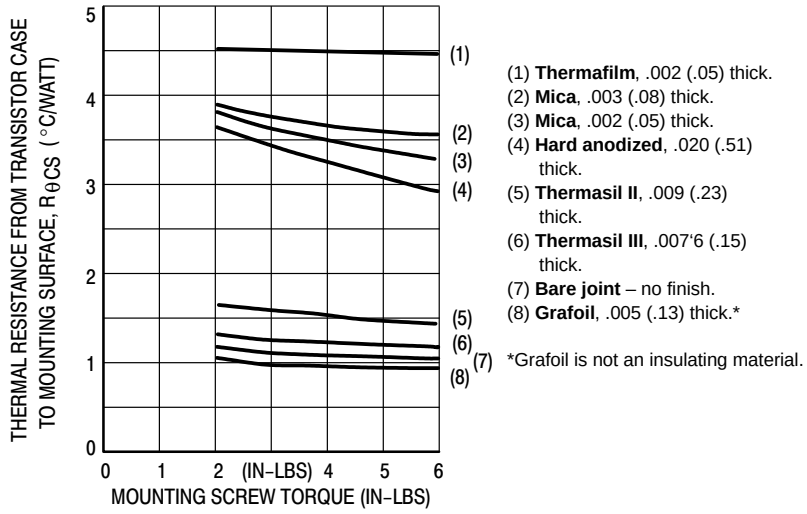


Figure 5. TO–220 Without Thermal Grease

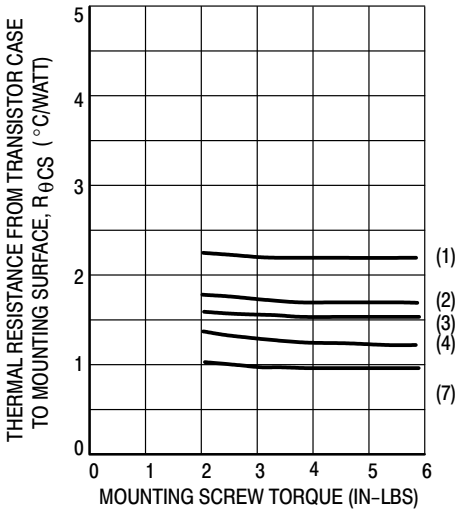


Figure 6. TO–220 With Thermal Grease

INTERFACE THERMAL RESISTANCE FOR TO–204, TO–3, AND TO–220 PACKAGES USING DIFFERENT INSULATING MATERIALS AS A FUNCTION OF MOUNTING SCREW TORQUE (DATA COURTESY THERMALLOY)

Silicone rubber insulators have gained favor because they are somewhat conformal under pressure. Their ability to fill in most of the metal voids at the interface reduces the need for thermal grease. When first introduced, they suffered from cut–through after a few years in service. The ones presently available have solved this problem by having imbedded pads of Kapton or fiberglass. By comparing Figure 5 and Figure 6, it can be noted that Thermasil™, a filled silicone rubber, without grease, has about the same interface thermal resistance as greased mica for the TO–220 package.

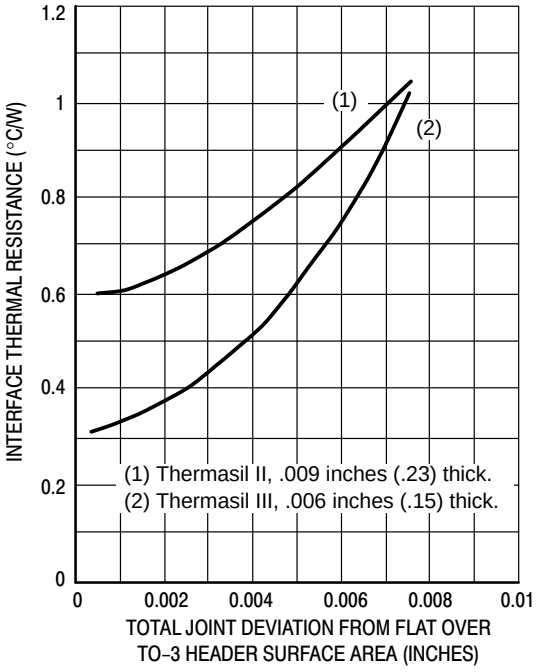
A number of manufacturers offer silicone rubber insulators. Table 2 shows measured performance of a number of these insulators under carefully controlled, nearly identical conditions. The interface thermal resistance extremes are over 2:1 for the various materials. It is also clear that some of the insulators are much more tolerant than others of out–of–flat surfaces. Since the tests were performed, newer products have been introduced. The Bergquist K–10® pad, for example, is described as having about 2/3 the interface resistance of the Sil–Pad® 1000 which would place its performance close to the Chomerics 1671 pad. Aavid also offers an isolated pad called Rubber–Duc™, however, it is only available vulcanized to a heatsink and, therefore, was not included in the comparison. Published data from Aavid shows $R_{\theta CS}$ below 0.3°C/W for pressures above 500 psi. However, surface flatness and other details are not specified, so a comparison cannot be made with other data in this note.

Table 2. Thermal Resistance of Silicone Rubber Pads

Manufacturer	Product	$R_{\theta CS}$ @ 3 Mils*	$R_{\theta CS}$ @ 7.5 Mils*
Wakefield	Delta Pad 173–7	.790	1.175
Bergquist	Sil–Pad K–4®	.752	1.470
Stockwell Rubber	1867	.742	1.015
Bergquist	Sil–Pad 400–9®	.735	1.205
Thermalloy	Thermasil II	.680	1.045
Shin–Etsu	TC–30AG	.664	1.260
Bergquist	Sil–Pad 400–7®	.633	1.060
Chomerics	1674	.592	1.190
Wakefield	Delta Pad 174–9	.574	.755
Bergquist	Sil–Pad 1000®	.529	.935
Ablestik	Thermal Wafers	.500	.990
Thermalloy	Thermasil III	.440	1.035
Chomerics	1671	.367	.655

*Test Fixture Deviation from flat from Thermalloy EIR86–1010.

The thermal resistance of some silicone rubber insulators is sensitive to surface flatness when used under a fairly rigid base package. Data for a TO–204AA (TO–3) package insulated with Thermasil is shown in Figure 7. Observe that the “worst case” encountered (7.5 mils) yields results having about twice the thermal resistance of the “typical case” (3 mils), for the more conductive insulator. In order for Thermasil III to exceed the performance of greased mica, total surface flatness must be under 2 mils, a situation that requires spot finishing.



Data courtesy of Thermalloy

Figure 7. Effect of Total Surface Flatness on Interface Resistance Using Silicon Rubber Insulators

Silicon rubber insulators have a number of unusual characteristics. Besides being affected by surface flatness and initial contact pressure, time is a factor. For example, in a study of the CHO–THERM® 1688 pad thermal interface impedance dropped from 0.90°C/W to 0.70°C/W at the end of 1000 hours. Most of the change occurred during the first 200 hours where $R_{\theta CS}$ measured 0.74°C/W. The torque on the conventional mounting hardware had decreased to 3 in–lb from an initial 6 in–lb. With nonconformal materials, a reduction in torque would have increased the interface thermal resistance.

Because of the difficulties in controlling all variables affecting tests of interface thermal resistance, data from different manufacturers is not in good agreement. Table 3 shows data obtained from two sources. The relative performance is the same, except for mica which varies widely in thickness. Appendix B discusses the variables which need to be controlled. At the time of this writing, ASTM Committee D9 is developing a standard for interface measurements.

The conclusions to be drawn from all this data is that some types of silicon rubber pads, mounted dry, will outperform the commonly used mica with grease. Cost may be a determining factor in making a selection.

Table 3. Performance of Silicon Rubber Insulators Tested Per MIL–I–49456

Material	Measured Thermal Resistance (°C/W)	
	Thermalloy Data(1)	Bergquist Data(2)
Bare Joint, greased	0.033	0.008
BeO, greased	0.082	–
CHO–THERM, 1617	0.233	–
Q Pad (non–insulated)	–	0.009
Sil–Pad, K–10	0.263	0.200
Thermasil III	0.267	–
Mica, greased	0.329	0.400
Sil–Pad 1000	0.400	0.300
CHO–THERM 1674	0.433	–
Thermasil II	0.500	–
Sil–Pad 400	0.533	0.440
Sil–Pad K–4	0.583	0.440

(1) From Thermalloy EIR 87–1030
(2) From Bergquist Data Sheet

Insulation Resistance

When using insulators, care must be taken to keep the matting surfaces clean. Small particles of foreign matter can puncture the insulation, rendering it useless or seriously lowering its dielectric strength. In addition, particularly when voltages higher than 300 V are encountered, problems with creepage may occur. Dust and other foreign material can shorten creepage distances significantly; so having a clean assembly area is important. Surface roughness and humidity also lower insulation resistance. Use of thermal grease usually raises the withstand voltage of the insulation system, but excess must be removed to avoid collecting dust. Because of these factors, which are not amenable to analysis, hi–pot testing should be done on prototypes and a large margin of safety employed.

Insulated Electrode Packages

Because of the nuisance of handling and installing the accessories needed for an insulated semiconductor mounting, equipment manufacturers have longed for cost–effective insulated packages since the 1950's. The first to appear were stud mount types which usually have a layer of beryllium oxide between the stud hex and the can. Although effective, the assembly is costly and requires manual mounting and lead wire soldering to terminals on top of the case. In the late 80's, a number of electrically isolated parts became available from various semiconductor manufacturers. These offerings presently consist of multiple chips and integrated circuits as well as the more conventional single chip devices.

The newer insulated packages can be grouped into two categories. The first has insulation between the semiconductor chips and the mounting base; an exposed area of the mounting base is used to secure the part. The Energy Management Series (EMS) modules, shown in Figure 16, Case 806 (ICePAK™) and Case 388A (TO–258AA) (see Figure 16) are examples of parts in this category. The second category contains parts which have a plastic overmold covering the metal mounting base. The isolated, Case 221C, illustrated in 21, is an example of parts in the second category.

Parts in the first category (those with an exposed metal flange or tab) are mounted the same as their non–insulated counterparts. However, as with any mounting system where pressure is bearing on plastic, the overmolded type should be used with a conical compression washer, described later in this note.

FASTENER AND HARDWARE CHARACTERISTICS

Characteristics of fasteners, associated hardware, and the tools to secure them determine their suitability for use in mounting the various packages. Since many problems have arisen because of improper choices, the basic characteristics of several types of hardware are discussed next.

Compression Hardware

Normal split ring lock washers are not the best choice for mounting power semiconductors. A typical #6 washer flattens at about 50 pounds, whereas 150 to 300 pounds is needed for good heat transfer at the interface. A very useful piece of hardware is the conical, sometimes called a Belleville washer, compression washer. As shown in Figure 8, it has the ability to maintain a fairly constant pressure over a wide range of its physical deflection – generally 20% to 80%. When installing, the assembler applies torque until the washer depresses to half its original height. (Tests should be run prior to setting up the assembly line to determine the proper torque for the fastener used to achieve 50% deflection.) The washer will absorb any cyclic expansion of the package, insulating washer or other materials caused by temperature changes. Conical washers are the key to successful mounting of devices requiring strict control of the mounting force or when plastic hardware is used in the mounting scheme. They are used with the large face contacting the packages. A new variation of the conical washer includes it as part of a nut assembly. Called a "sync nut," the patented device can be soldered to a PC board and the semiconductor mounted with a 6–32 machine screw.(4)

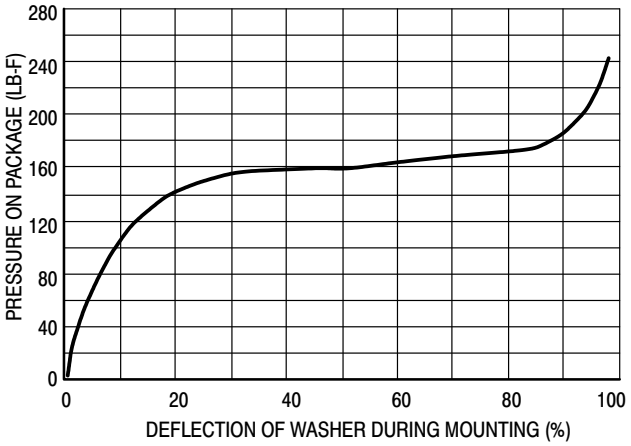


Figure 8. Characteristics of the Conical Compression Washers Designed for Use with Plastic Body Mounted Semiconductors

Clips

Fast assembly is accomplished with clips. When only a few watts are being dissipated, the small board–mounted or free–standing heat dissipaters with an integral clip, offered by several manufacturers, result in a low cost assembly. When higher power is being handled, a separate clip may be used with larger heatsinks. In order to provide proper pressure, the clip must be specially designed for a particular heatsink thickness and semiconductor package.

Clips are especially popular with plastic packages such as the TO–220 and TO–126. In addition to fast assembly, the clip provides lower interface thermal resistance than other assembly methods when it is designed for proper pressure to bear on the top of the plastic over the die. The TO–220 package usually is lifted up under the die location when mounted with a single fastener through the hole in the tab because of the high pressure at one end.

Machine Screws

Machine screws, conical washers, and nuts (or sync nuts) can form a trouble–free fastener system for all types of packages which have mounting holes. However, proper torque is necessary. Torque ratings apply when dry; therefore, care must be exercised when using thermal grease to prevent it from getting on the threads as inconsistent torque readings result. Machine screw heads should not directly contact the surface of plastic packages types as the screw heads are not sufficiently flat to provide properly distributed force. Without a washer, cracking of the plastic case could occur.

Self–Tapping Screws

Under carefully controlled conditions, sheet–metal screws are acceptable. However, during the tapping process with a standard screw, a volcano–like protrusion will develop in the metal being threaded; an unacceptable surface that could increase the thermal resistance may result. When standard sheet metal screws are used, they must be used in a clearance hole to engage a speednut. If a self–tapping process is desired, the screw type must be used which roll–forms machine screw threads.

Rivets

Rivets are not a recommended fastener for any of the plastic packages. When a rugged metal flange–mount package or EMS module is being mounted directly to a heatsink, rivets can be used provided press–riveting is used. Crimping force must be applied slowly and evenly. Pop–riveting should never be used because the high crimping force could cause deformation of most semiconductor packages. Aluminum rivets are much preferred over steel because less pressure is required to set the rivet and thermal conductivity is improved.

The hollow rivet, or eyelet, is preferred over solid rivets. An adjustable, regulated pressure press is used such that a gradually increasing pressure is used to pan the eyelet. Use of sharp blows could damage the semiconductor die.

Solder

Until the advent of the surface mount assembly technique, solder was not considered a suitable fastener for power semiconductors. However, user demand has led to the development of new packages for this application. Acceptable soldering methods include conventional belt–furnace, irons, vapor–phase reflow, and infrared reflow. It is important that the semiconductor temperature not exceed the specified maximum (usually 260°C) or the die bond to the case could be damaged. A degraded die bond has excessive thermal resistance which often leads to a failure under power cycling.

Adhesives

Adhesives are available which have coefficients of expansion compatible with copper and aluminum.[5] Highly conductive types are available; a 10 mil layer has approximately 0.3°C/W interface thermal resistance. Different types are offered: high strength types for non–field serviceable systems or low strength types for field serviceable systems. Adhesive bonding is attractive when case–mounted parts are used in wave soldering assembly because thermal greases are not compatible with the conformal coatings used and the greases foul the solder process.

Plastic Hardware

Most plastic materials will flow, but differ widely in this characteristic. When plastic materials form parts of the fastening system, compression washers are highly valuable to assure that the assembly will not loosen with time and temperature cycling. As previously discussed, loss of contact pressure will increase interface thermal resistance.

FASTENING TECHNIQUES

Each of the various classes of packages in use requires different fastening techniques. Details pertaining to each type are discussed in the following sections. Some general considerations follow.

To prevent galvanic action from occurring when devices are used on aluminum heatsinks in a corrosive atmosphere, many devices are nickel– or gold–plated. Consequently, precautions must be taken not to mar the finish.

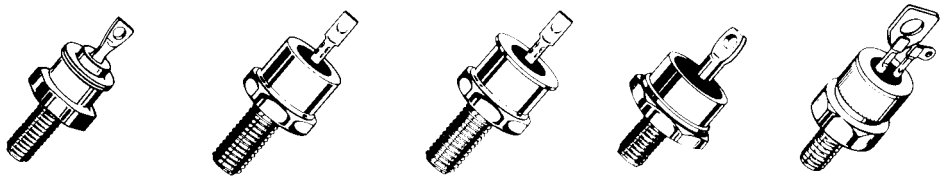
Another factor to be considered is that when a copper–based part is rigidly mounted to an aluminum heatsink, a bi–metallic system results which will bend with temperature changes. Not only is the thermal coefficient of expansion different for copper and aluminum, but the temperature gradient through each metal also causes each component to bend. If bending is excessive and the package is mounted by two or more screws, the semiconductor chip could be damaged.

Bending can be minimized by:

1. Mounting the component parallel to the heatsink fins to provide increased stiffness.
2. Allowing the heatsink holes to be a bit oversized so that some can slip between surfaces as the temperature changes.
3. Using a highly conductive thermal grease or mounting pad between the heatsink and semiconductor to minimize the temperature gradient and allow for movement.

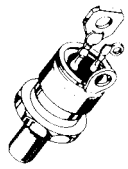
Stud Mount

Parts which fall into the stud–mount classification are shown in Figure 8 through Figure 11. Mounting errors with non–insulated stud–mounted parts are generally confined to application of excessive torque or tapping the stud into a threaded heatsink hole. Both of these practices may cause a warpage of the hex base which may crack the semiconductor die. The only recommended fastening method is to use a nut and washer. The details are shown in Figure 12.



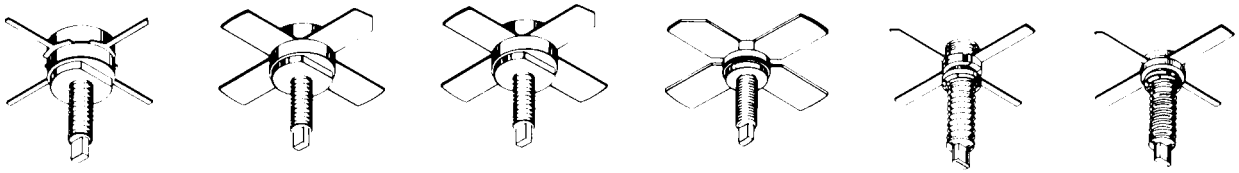
CASE 42A (DO–5) | **CASE 56–03 DO–203AA (DO–4)** | **CASE 245 (DO–4)** | **CASE 257 DO–203AB (DO–5)** | **CASE 263–04**

Figure 9. Standard Non–Isolated Types



CASE 311–02

Figure 10. Isolated Type



CASE 144B–05 (.380″ STUD) | **CASE 145A–09 (.380″ STUD)** | **CASE 145A–10 (.500″ STUD)** | **CASE 244–04 (.280″ STUD)** | **CASE 305–01 (.204″ STUD)** | **CASE 332–04 (.380″ STUD)**

Figure 11. RF Stripline Opposed Emitter (SOE) Series

A VARIETY OF STUD–MOUNT PARTS

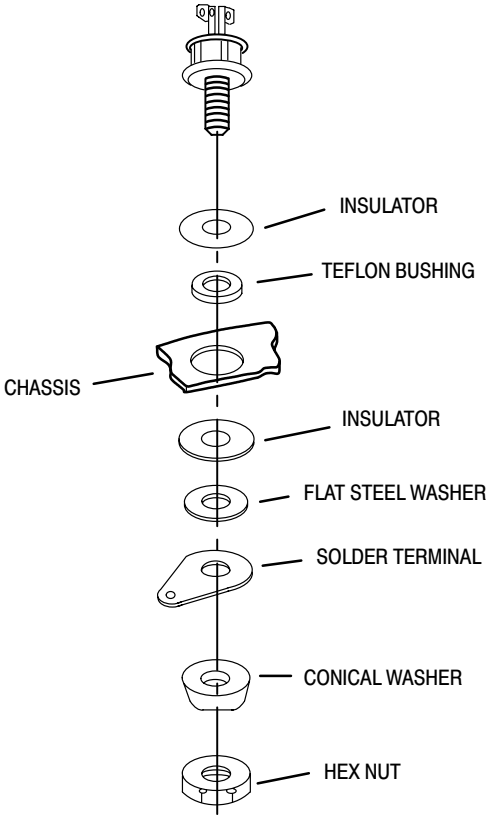


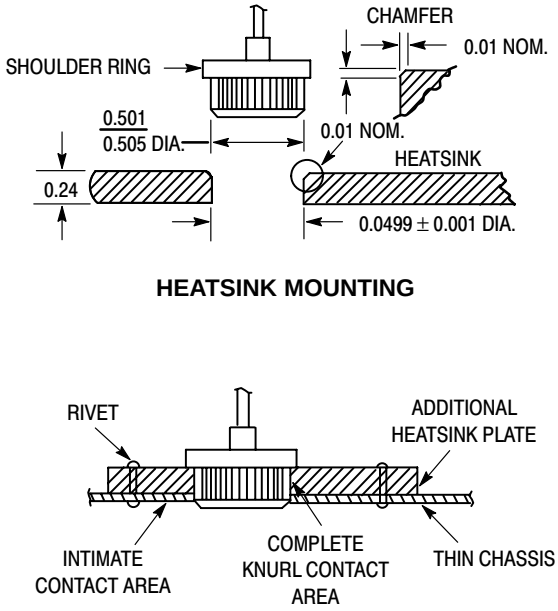
Figure 12. Isolating Hardware Used for a Non–Isolated Stud–Mount Package

Insulated electrode packages on a stud mount base require less hardware. They are mounted the same as their non–insulated counterparts, but care must be exercised to avoid applying a shear or tension stress to the insulation layer, usually a beryllium oxide (BeO) ceramic. This requirement dictates that the leads must be attached to the circuit with flexible wire. In addition, the stud hex should be used to hold the part while the nut is torqued.

RF transistors in the stud–mount Stripline Opposed Emitter (SOE) package impose some additional constraints because of the unique construction of the package. Special techniques to make connections to the stripline leads and to mount the part so no tension or shear forces are applied to any ceramic – metal interface are discussed in the section entitled, “Connecting and Handling Terminals.”

Press Fit

For most applications, the press–fit case should be mounted according to the instructions shown in Figure 13. A special fixture, meeting the necessary requirements, must be used.



The hole edge must be chamfered as shown to prevent shearing off the knurled edge of the case during press–in. The pressing force should be applied evenly on the shoulder ring to avoid tilting or canting of the case in the hole during the pressing operation. Also, the use of a thermal joint compound will be of considerable aid. The pressing force will vary from 250 to 1000 pounds, depending upon the heatsink material. Recommended hardnesses are: copper–less than 50 on the Rockwell F scale; aluminum–less than 65 on the Brinell scale. A heatsink as thin as 1/8″ may be used, but the interface thermal resistance will increase in direct proportion to the contact area. A thin chassis requires the addition of a backup plate.

Figure 13. Press–Fit Package

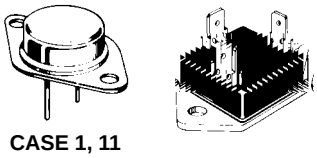
Flange Mount

A large variety of parts fit into the flange mount category as shown in Figure 14 through Figure 17. Few known mounting difficulties exist with the smaller flange mount packages, such as the TO–204 (TO–3). The rugged base and distance between die and mounting holes combine to make it extremely difficult to cause any warpage unless mounted on a surface which is badly bowed or unless one side is tightened excessively before the other screw is started. It is, therefore, good practice to alternate tightening of the screws so that pressure is evenly applied. After the screws are finger–tight, the hardware should be torqued to its final specification in at least two sequential steps. A typical mounting installation for a popular flange type part is shown in Figure 18. Machine screws (preferred), self–tapping screws, eyelets, or rivets may be used to secure the package using guidelines in the previous section, “Fastener and Hardware Characteristics.”

The copper flange of the Energy Management Series (EMS) modules is very thick. Consequently, the parts are rugged and indestructible for all practical purposes. No special precautions are necessary when fastening these parts to a heatsink.

Some packages specify a tightening procedure. For example, with the Power Tap package, 15, final torque should be applied first to the center position.

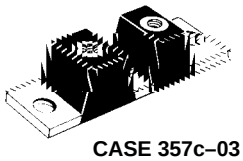
The RF power modules (MHW series) are more sensitive to the flatness of the heatsink than other packages because a ceramic (BeO) substrate is attached to a relatively thin, fairly long, flange. The maximum allowable flange bending to avoid mechanical damage has been determined and presented in detail in Engineering Bulletin EB107/D, “Mounting Considerations for ON Semiconductor RF Power Modules.” Many of the parts can handle a combined heatsink and flange deviation from flat of 7 to 8 mils which is commonly available. Others must be held to 1.5 mils, which requires that the heatsink have nearly perfect flatness.



CASE 1, 11
TO–204AA
(TO–3)

CASE 383–01

Figure 14. TO–3 Variations



CASE 357c–03

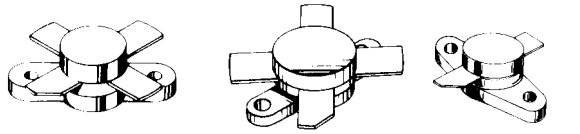
Figure 15. Plastic Power Tap

CASE 373–01 CASE 807–03 CASE 807A–03 CASE 808–01

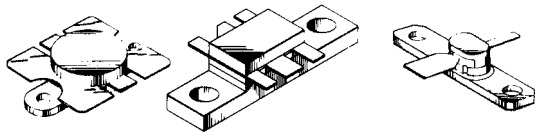
CASE 809–02 CASE 812–02 CASE 813–02

CASE 814–02 CASE 816–02 CASE 819–02

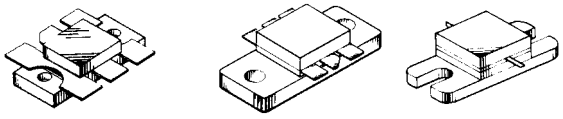
Figure 16. Energy Management Series (Isolated Base Plate)



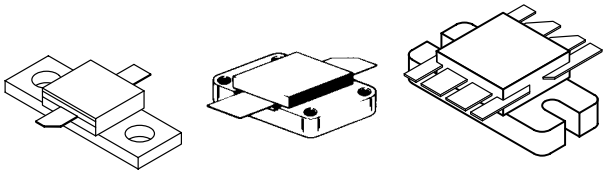
CASE 211–07 CASE 211–11 CASE 215–02



CASE 316–01 CASE 319–07 (CS–12) CASE 328A–03



CASE 333–04 CASE 333A–02 (MAAC PAC) CASE 336–03



CASE 337–02 CASE 368–03 (HOG PAC) CASE 744A–01

Figure 17. RF Stripline Isolated Output Opposed Emitter (SOE) Series

A LARGE ARRAY OF PARTS FIT INTO THE FLANGE–MOUNT CLASSIFICATION

Specific mounting recommendations are critical to RF devices in isolated packages because of the internal ceramic substrate. The large area, Case 368–03 (HOG PAC), will be used to illustrate problem areas. It is more sensitive to proper mounting techniques than most other RF power devices.

Although the data sheets contain information on recommended mounting procedures, experience indicates that they are often ignored. For example, the recommended maximum torque on the 4–40 mounting screws is 5 in./lbs. Spring and flat washers are recommended. Over–torquing is a common problem. In some parts returned for failure analysis, indentions up to 10 mils deep in the mounting screw areas, have been observed.

Calculations indicate that the length of the flange increases in excess of two mils with a temperature change of 75°C. In such cases, if the mounting screw torque is excessive, the flange is prevented from expanding in length; instead, it bends upward in the mid–section, cracking the BeO and the die. A similar result can also occur during the initial mounting of the device if an excessive amount of thermal compound is applied. With

sufficient torque, the thermal compound will squeeze out of the mounting hole areas, but will remain under the center of the flange, deforming it. Deformations of 2 – 3 mils have been measured between the center and the ends under such conditions (enough to crack internal ceramic).

Another problem arises because the thickness of the flange changes with temperature. For the 75°C temperature excursion mentioned, the increased amount is around 0.25 mils which results in further tightening of the mounting screws, thus increasing the effective torque from the initial value. With a decrease in temperature, the opposite effect occurs. Therefore, thermal cycling not only causes risk of structural damage but often causes the assembly to loosen which raises the interface resistance. Use of compression hardware can eliminate this problem.

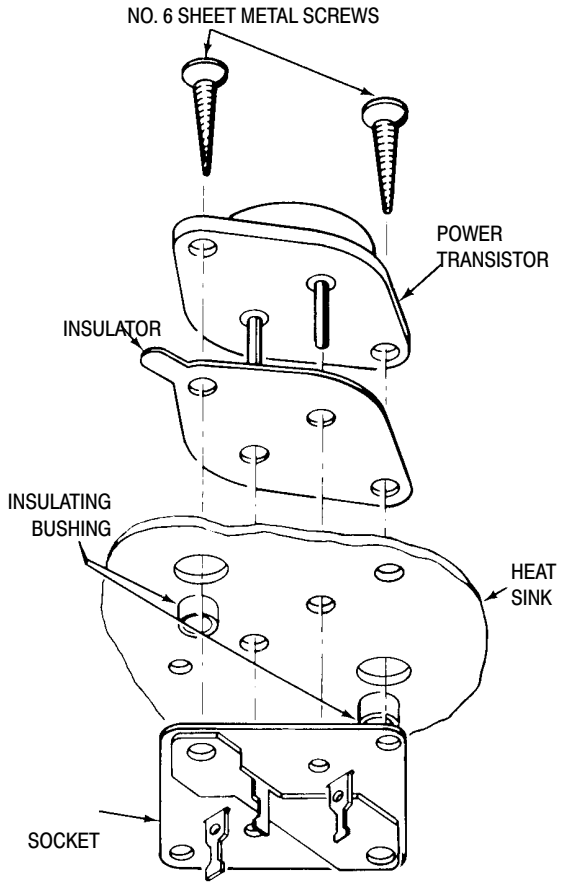


Figure 18. Hardware Used for a TO–204AA (TO–3) Flange Mount Part

Tab Mount

The tab mount class is composed of a wide array of packages as illustrated in Figure 19. Mounting considerations for all varieties are similar to that for the popular TO–220 package, whose suggested mounting arrangements and hardware are shown in Figure 20. The rectangular washer shown in Figure 20a is used to minimize distortion of the mounting flange; excessive distortion could cause damage to the semiconductor chip. Use of the washer is only important when the size of the mounting hole exceeds 0.140 in. (6–32 clearance). Larger holes are needed to accommodate the lower insulating bushing when the screw is electrically connected to the case; however, the holes should not be larger than necessary to provide hardware clearance and should never exceed a diameter of 0.250 inch. Flange distortion is also possible if excessive torque is used during mounting. A maximum torque of 8 in./lbs is suggested when using a 6–32 screw.

Care should be exercised to assure that the tool used to drive the mounting screw never comes in contact with the plastic body during the driving operation. Such contact can result in damage to the plastic body and internal device connections. To minimize this problem, ON Semiconductor TO–220 packages have a chamfer on one end. TO–220 packages of other manufacturers may need a spacer or combination spacer and isolation bushing to raise the screw head above the top surface of the plastic.

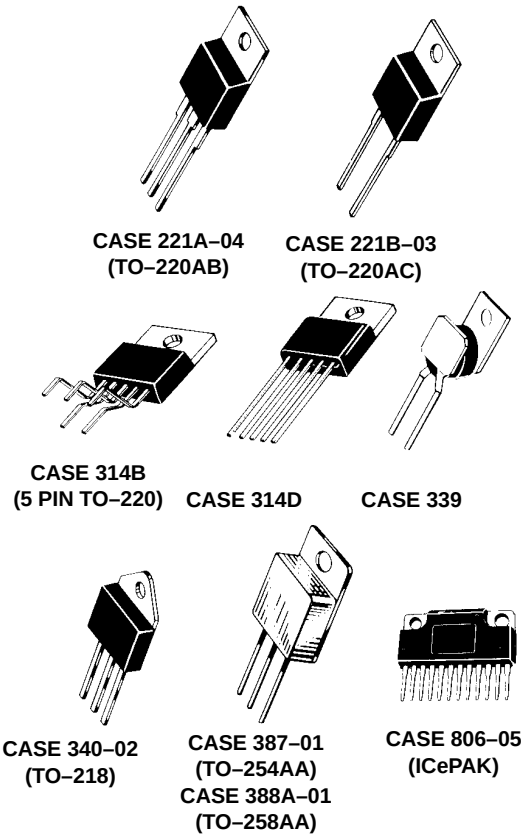


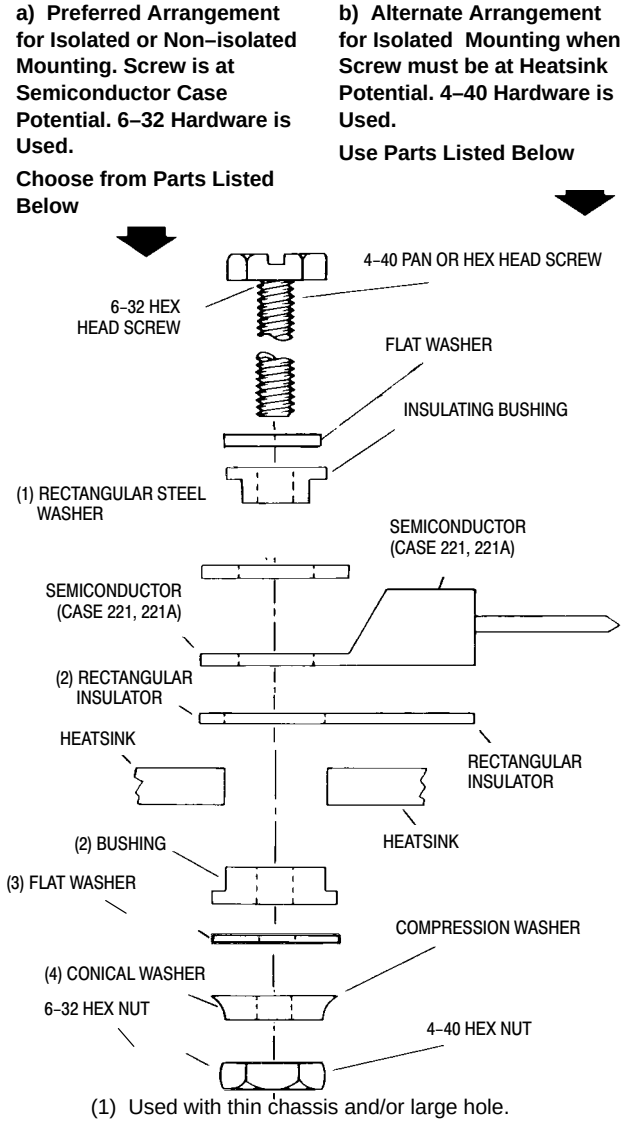
Figure 19. Several Types of Tab–Mount Parts

The popular TO–220 Package and others of similar construction lift off the mounting surface as pressure is applied to one end. (See Appendix B, Figure 36.) To counter this tendency, at least one hardware manufacturer offers a hard plastic cantilever beam which applies more even pressure on the tab.[6] In addition, it separates the mounting screw from the metal tab. Tab mount parts may also be effectively mounted with clips as shown in Figure 27.

To obtain high pressure without cracking the case, a pressure spreader bar should be used under the clip. Interface thermal resistance with the cantilever beam or clips can be lower than with screw mounting.

The ICePAK (Case 806–05) is basically an elongated TO–220 package with isolated chips. The mounting precautions for the TO–220 consequently apply. In addition, since two mounting screws are required, the alternate tightening procedure described for the flange mount package should be used.

In situations where a tab mount package is making direct contact with the heatsink, an eyelet may be used, provided sharp blows or impact shock is avoided.



(1) Used with thin chassis and/or large hole.
(2) Used when isolation is required.
(3) Required when nylon bushing is used.

Figure 20. Mounting Arrangements for Tab Mount TO–220

Plastic Body Mount

The Thermopad™ and isolated plastic power packages shown in Figure 21 are typical of packages in this group. They have been designed to feature minimum size with no compromise in thermal resistance. For the Thermopad (Case 77) parts, this is accomplished by die–bonding the silicon chip on one side of a thin copper sheet; the opposite side is exposed as a mounting surface. The copper sheet has a hole for mounting; plastic is molded enveloping the chip but leaving the mounting hole open. The low thermal resistance of this construction is obtained at the expense of a requirement that strict attention be paid to the mounting procedure.

The isolated (Case 221C–02) is similar to a TO–220 except that the tab is encased in plastic. Because the mounting force is applied to plastic, the mounting procedure differs from a standard TO–220 and is similar to that of the Thermopad.

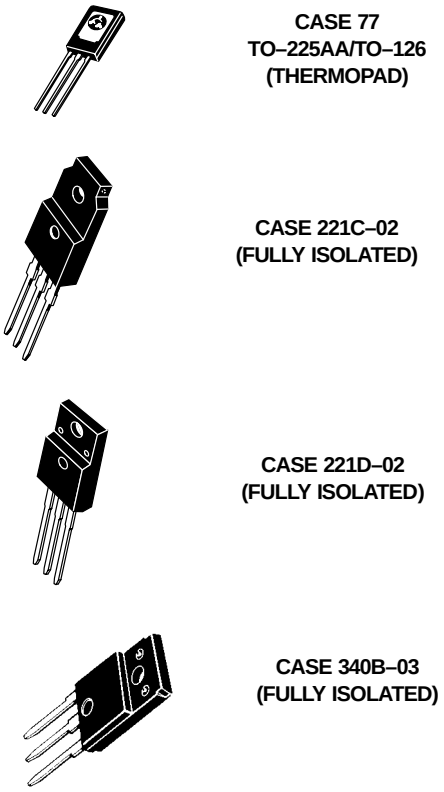


Figure 21. Plastic Body–Mount Packages

Several types of fasteners may be used to secure these packages; machine screws, eyelets, or clips are preferred. With screws or eyelets, a conical washer should be used which applies the proper force to the package over a fairly wide range of deflection and distributes the force over a fairly large surface area. Screws should not be tightened with any type of air–driven torque gun or equipment which may cause high impact. Characteristics of a suitable conical washer is shown in Figure 8.

Figure 22 through Figure 24 shows details of mounting Case 77 devices. Clip mounting is fast and requires minimum hardware, however, the clip must be properly chosen to insure that the proper mounting force is applied. When electrical isolation is required with screw mounting, a bushing inside the mounting hole will insure that the screw threads do not contact the metal base.

The isolated, (Case 221C, 221D, and 340B) permits the mounting procedure to be greatly simplified over that of a standard TO–220. As shown in Figure 27, one properly chosen clip, inserted into two slotted holes in the heatsink, is all the hardware needed. Even though clip pressure is much lower than obtained with a screw, the thermal resistance is about the same for either method. This occurs because the clip bears directly on top of the die and holds the package flat while the screw causes the package to lift up somewhat under the die. (See Figure 36 of Appendix B.) The interface should consist of a layer of thermal grease or a highly conductive thermal pad. Of course, screw mounting shown in Figure 26 may also be used, but a conical compression washer should be included. Both methods afford a major reduction in hardware as compared to the conventional mounting method with a TO–220 package which is shown in Figure 25.

Surface Mount

Although many of the tab mount parts have been surface mounted, special small footprint packages for mounting power semiconductors using surface mount assembly techniques have been developed. The DPAK, shown in Figure 28, for example, will accommodate a die up to 112 mils x 112 mils, and has a typical thermal resistance around 2°C/W junction to case. The thermal resistance values of the solder interface is well under 1°C/W. The printed circuit board also serves as the heatsink.

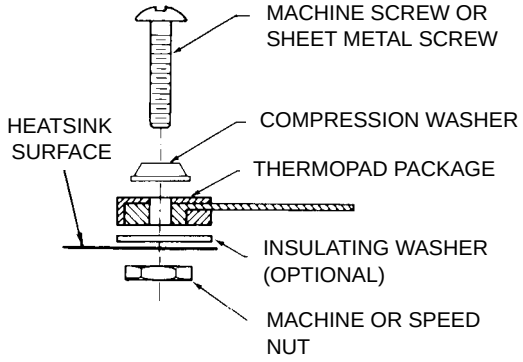
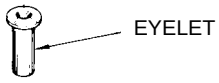


Figure 22. Machine Screw Mounting



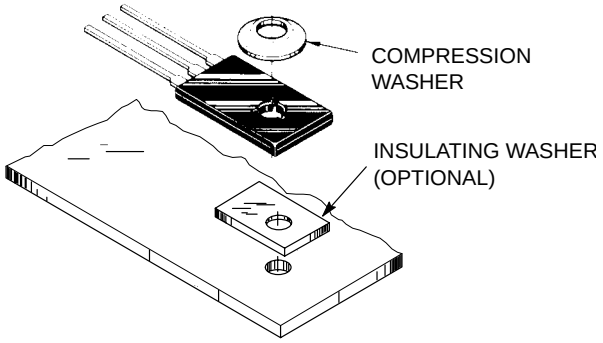


Figure 23. Eyelet Mounting

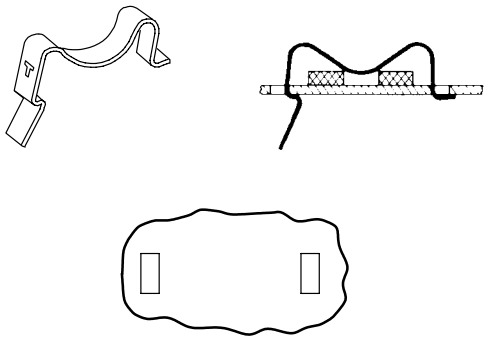


Figure 24. Clips

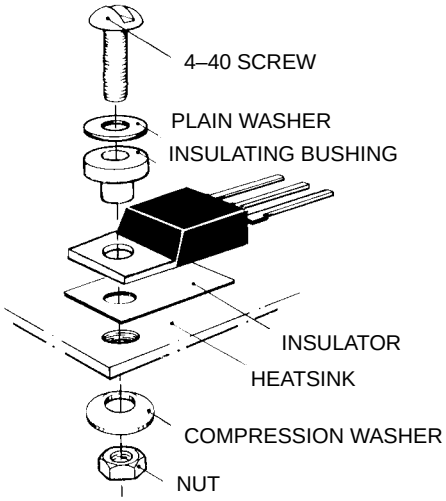


Figure 25. Screw–Mounted TO–220

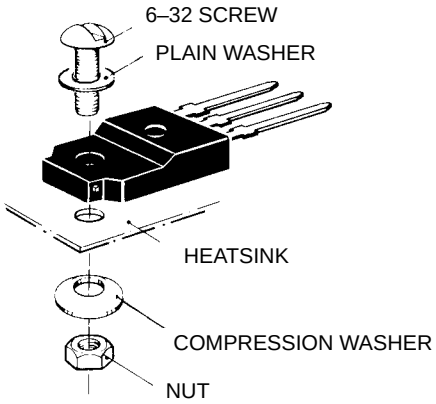


Figure 26. Screw–Mounted Isolated Package

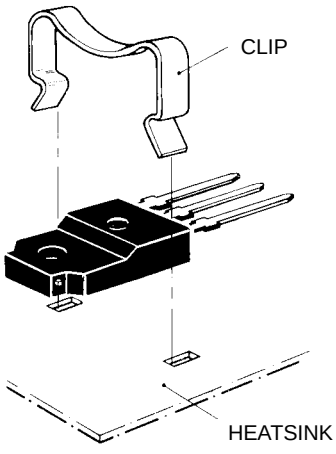


Figure 27. Clip–Mounted Isolated Package

RECOMMENDED MOUNTING ARRANGEMENTS FOR TO–225AA (TO–126) THERMOPAD PACKAGES

MOUNTING ARRANGEMENTS FOR THE ISOLATED PACKAGE AS COMPARED TO A CONVENTIONAL

CASE 369–07 CASE 369A–13

Figure 28. Surface Mount D–PAK Parts

Standard Glass–Epoxy 2–ounce boards do not make very good heatsinks because the thin foil has a high thermal resistance. As Figure 29 shows, thermal resistance asymptotes to about 20°C/W at 10 square inches of board area, although a point of diminishing returns occurs at about 3 square inches.

Boards are offered that have thick aluminum or copper substrates. A dielectric coating designed for low thermal resistance is overlaid with one– or two–ounce copper foil for the preparation of printed conductor traces. Tests run on such a product indicate that case to substrate thermal resistance is in the vicinity of 1°C/W, exact values depending upon board type.[7] The substrate may be an effective heatsink itself, or it can be attached to a conventional finned heatsink for improved performance.

Since DPAK and other surface mount packages are designed to be compatible with surface mount assembly techniques, no special precautions are needed other than to insure that maximum temperature/time profiles are not exceeded.

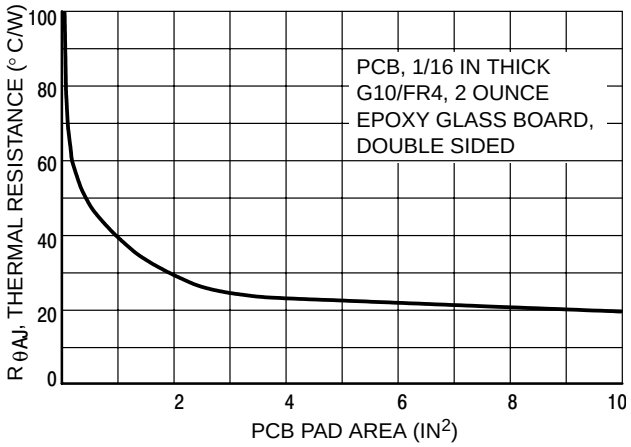


Figure 29. Effect of Footprint Area on Thermal Resistance of DPAK Mounted on a Glass–Epoxy Board

FREE AIR AND SOCKET MOUNTING

In applications where average power dissipation is on the order of a watt or so, most power semiconductors may be mounted with little or no heatsinking. The leads of the various metal power packages are not designed to support the packages; their cases must be firmly supported to avoid the possibility of cracked seals around the leads. Many plastic packages may be supported by their leads in applications where high shock and vibration stresses are not encountered and where no heatsink is used. The leads should be as short as possible to increase vibration resistance and reduce thermal resistance. As a general practice, however, it is better to support the package. A plastic support for the TO–220 Package and other similar types is offered by heatsink accessory vendors.

In many situations, because its leads are fairly heavy, the Case 77 (TO–225AA) (TO–127) package has supported a small heatsink; however, no definitive data is available. When using a small heatsink, it is good practice to have the sink rigidly mounted such that the sink or the board is providing total support for the semiconductor. Two possible arrangements are shown in Figure 30 and Figure 31. The arrangement of Figure 30 could be used with any plastic package, but the scheme of Figure 31 is more practical with Case 77 Thermopad devices. With the other package types, mounting the transistor on top of the heatsink is more practical.

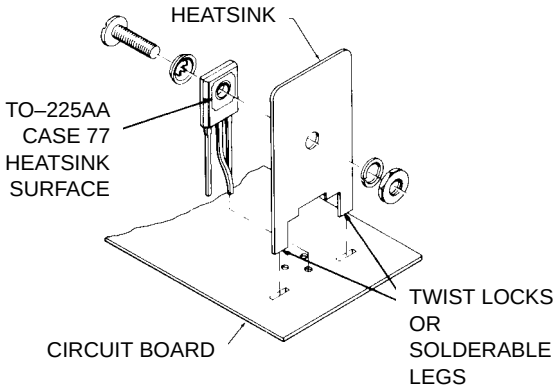


Figure 30. Simple Plate, Vertically Mounted

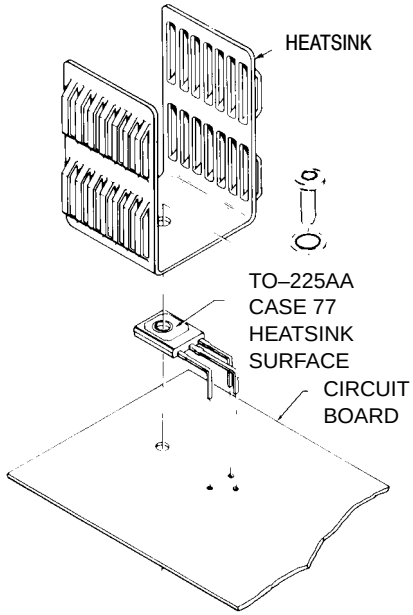


Figure 31. Commercial Sink, Horizontally Mounted

METHODS OF USING SMALL HEATSINKS WITH PLASTIC SEMICONDUCTOR PACKAGES

In certain situations, in particular where semiconductor testing is required or prototypes are being developed, sockets are desirable. Manufacturers have provided sockets for many of the packages available from ON Semiconductor. The user is urged to consult manufacturers' catalogs for specific details. Sockets with Kelvin connections are necessary to obtain accurate voltage readings across semiconductor terminals.

CONNECTING AND HANDLING TERMINALS

Pins, leads, and tabs must be handled and connected properly to avoid undue mechanical stress which could cause semiconductor failure. Change in mechanical dimensions, as a result of thermal cycling over operating temperature extremes, must be considered. Standard metal, plastic, and RF stripline packages each have some special considerations.

Metal Packages

The pins and lugs of metal packaged devices using glass to metal seals are not designed to handle any significant bending or stress. If abused, the seals could crack. Wires may be attached using sockets, crimp connectors or solder, provided the data sheet ratings are observed. When wires are attached directly to the pins, flexible or braided leads are recommended in order to provide strain relief.

EMS Modules

The screw terminals of the EMS modules look deceptively rugged. Since the flange base is mounted to a rigid heatsink, the connection to the terminals must allow some flexibility. A rigid buss bar should not be bolted to terminals. Lugs with braid are preferred.

Plastic Packages

The leads of the plastic packages are somewhat flexible and can be reshaped, although this is not a recommended procedure. In many cases, a heatsink can be chosen which makes lead–bending unnecessary. Numerous lead and tab–forming options are available from ON Semiconductor on large quantity orders. Preformed leads remove the users' risk of device damage caused by bending.

If, however, lead–bending is done by the user, several basic considerations should be observed. When bending the lead, support must be placed between the point of bending and the package. For forming small quantities of units, a pair of pliers may be used to clamp the leads at the case, while bending with the fingers or another pair of pliers. For production quantities, a suitable fixture should be made.

The following rules should be observed to avoid damage to the package.

1. A leadbend radius greater than 1/16 inch is advisable for TO–225AA (CASE 77) and 1/32 inch for TO–220.
2. No twisting of leads should be done at the case.
3. No axial motion of the lead should be allowed with respect to the case.

The leads of plastic packages are not designed to withstand excessive axial pull. Force in this direction greater than 4 pounds may result in permanent damage to the device. If the mounting arrangement imposes axial stress on the leads, a condition which may be caused by thermal cycling, some method of strain relief should be devised. When wires are used for connections, care should be exercised to assure that movement of the wire does not cause movement of the lead at the lead–to–plastic junctions. Highly flexible or braided wires are good for providing strain relief.

Wire–wrapping of the leads is permissible, provided that the lead is restrained between the plastic case and the point of the wrapping. The leads may be soldered; the maximum soldering temperature, however, must not exceed 260°C and must be applied for not more than 5 seconds at a distance greater than 1/8 inch from the plastic case.

Stripline Packages

The leads of stripline packages normally are soldered into a board while the case is recessed to contact a heatsink as shown in Figure 32 through Figure 34. The following rules should be observed:

1. The device should never be mounted in such a manner as to place ceramic–to–metal joints in tension.
2. The device should never be mounted in such a manner as to apply force on the strip leads in a vertical direction towards the cap.
3. When the device is mounted in a printed circuit board with the copper stud and BeO portion of the header passing through a hole in the circuit boards, adequate clearance must be provided for the BeO to prevent shear forces from being applied to the leads.
4. Some clearance must be allowed between the leads and the circuit board when the device is secured to the heatsink.
5. The device should be properly secured into the heatsinks before its leads are attached into the circuit.
6. The leads on stud type devices must not be used to prevent device rotation during stud torque application. A wrench flat is provided for this purpose.

Figure 33 shows a cross–section of a printed circuit board and heatsink assembly for mounting a stud–type stripline device. H is the distance from the top surface of the printed circuit board to the D–flat heatsink surface. If H is less than the minimum distance from the bottom of the lead material to the mounting surface of the package, there is no possibility of tensile forces in the copper stud – BeO ceramic joint. If, however, H is greater than the package dimension, considerable force is applied to the cap to BeO joint and the BeO to stud joint. Two occurrences are possible at this point. The first is a cap joint failure when the structure is heated, as might occur during the lead–soldering operation; while the second is BeO to stud failure if the force generated is high enough. Lack of contact between the device and the heatsink surface will

occur as the differences between H and the package dimension become larger; this may result in device failure as power is applied.

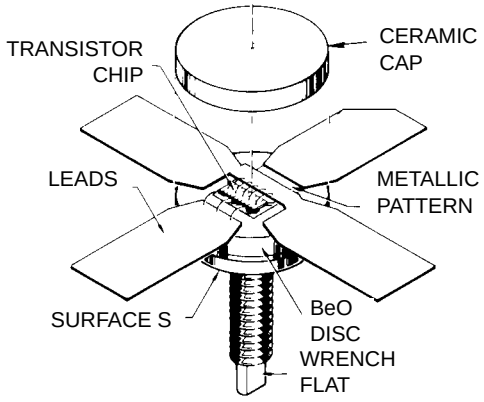


Figure 32. Component Parts of a Stud Mount Stripline Package. Flange Mounted Packages are Similarly Constructed

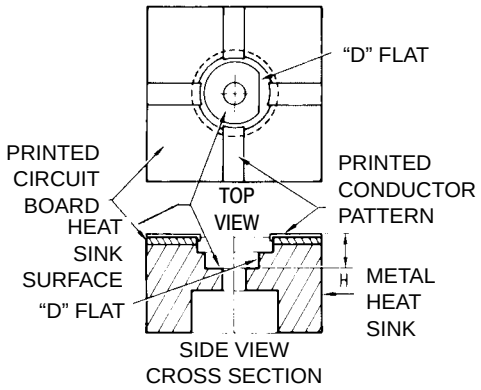


Figure 33. Typical Stud Type SOE Transistor Mounting Method

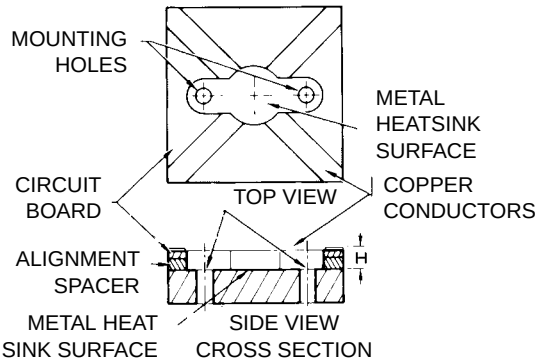


Figure 34. Flange Type SOE Transistor Mounting Method

MOUNTING DETAILS FOR SOE TRANSISTORS

Figure 34 shows a typical mounting technique for flange–type stripline transistors. Again, H is defined as the distance from the top of the printed circuit board to the heatsink surface. If distance H is less than the minimum distance from the bottom of transistor lead to the bottom surface of the flange, tensile forces at the various joints in the package are avoided. However, if distance H exceeds the package dimension, problems similar to those discussed for the stud type devices can occur.

CLEANING CIRCUIT BOARDS

It is important that any solvents or cleaning chemicals used in the process of degreasing or flux removal do not affect the reliability of the devices. Alcohol and unchlorinated Freon solvents are generally satisfactory for use with plastic devices, since they do not damage the package. Hydrocarbons such as gasoline and chlorinated Freon may cause the encapsulant to swell, possibly damaging the transistor die.

When using an ultrasonic cleaner for cleaning circuit boards, care should be taken with regard to ultrasonic energy and time of application. This is particularly true if any packages are free–standing without support.

THERMAL SYSTEM EVALUATION

Assuming that a suitable method of mounting the semiconductor without incurring damage has been achieved, it is important to ascertain whether the junction temperature is within bounds.

In applications where the power dissipated in the semiconductor consists of pulses at a low duty cycle, the instantaneous or peak junction temperature, not average temperature, may be the limiting condition. In this case, use must be made of transient thermal resistance data. For a full explanation of its use, see ON Semiconductor Application Note, AN569/D.

Other applications, notably RF power amplifiers or switches driving highly reactive loads, may create severe current crowding conditions which render the traditional concepts of thermal resistance or transient thermal impedance invalid. In this case, transistor safe operating area, thyristor di/dt limits, or equivalent ratings as applicable, must be observed.

Fortunately, in many applications, a calculation of the average junction temperature is sufficient. It is based on the concept of thermal resistance between the junction and a temperature reference point on the case. (See Appendix A.) A fine wire thermocouple should be used, such as #36 AWG, to determine case temperature. Average operating junction temperature can be computed from the following equation:

$$T_J = T_C + R_{\theta JC} \times P_D$$

where T_J = junction temperature (°C)
T_C = case temperature (°C)
$R_{\theta JC}$ = thermal resistance junction–to case as specified on the data sheet (°C/W)
P_D = power dissipated in the device (W)

The difficulty in applying the equation often lies in determining the power dissipation. Two commonly used empirical methods are graphical integration and substitution.

Graphical Integration

Graphical integration may be performed by taking oscilloscope pictures of a complete cycle of the voltage and current waveforms, using a limit device. The pictures should be taken with the temperature stabilized. Corresponding points are then read from each photo at a suitable number of time increments. Each pair of voltage and current values are multiplied together to give instantaneous values of power. The results are plotted on linear graph paper, the number of squares within the curve counted, and the total divided by the number of squares along the time axis. The quotient is the average power dissipation. Oscilloscopes are available to perform these measurements and make the necessary calculations.

Substitution

This method is based upon substituting an easily measurable, smooth dc source for a complex waveform. A switching arrangement is provided which allows operating the load with the device under test, until it stabilizes in temperature. Case temperature is monitored. By throwing the switch to the “test” position, the device under test is connected to a dc power supply, while another pole of the switch supplies the normal power to the load to keep it operating at full power level. The dc supply is adjusted so that the semiconductor case temperature remains approximately constant when the switch is thrown to each position for about 10 seconds. The dc voltage and current values are multiplied together to obtain average power. It is generally necessary that a Kelvin connection be used for the device voltage measurement.

APPENDIX A
THERMAL RESISTANCE CONCEPTS

The basic equation for heat transfer under steady–state conditions is generally written as:

$$q = hA\Delta T \quad (1)$$

where q = rate of heat transfer or power dissipation (P_D)
h = heat transfer coefficient,
A = area involved in heat transfer,
ΔT = temperature difference between regions of heat transfer.

However, electrical engineers generally find it easier to work in terms of thermal resistance, defined as the ratio of temperature to power. From Equation 1, thermal resistance, R_θ, is

$$R_\theta = \Delta T/q = 1/hA \quad (2)$$

The coefficient (h) depends upon the heat transfer mechanism used and various factors involved in that particular mechanism.

An analogy between Equation (2) and Ohm's Law is often made to form models of heat flow. Note that T could be thought of as a voltage thermal resistance corresponds to electrical resistance (R); and, power (q) is analogous to current (I). This gives rise to a basic thermal resistance model for a semiconductor as indicated by Figure 35.

The equivalent electrical circuit may be analyzed by using Kirchoff's Law and the following equation results:

$$T_J = P_D (R_{\theta JC} + P_{\theta CS} + R_{\theta SA}) + T_A \quad (3)$$

where T_J = junction temperature,
PD = power dissipation
$R_{\theta JC}$ = semiconductor thermal resistance (junction to case),
$R_{\theta CS}$ = interface thermal resistance (case to heat–sink),
$R_{\theta SA}$ = heat sink thermal resistance (heatsink to ambient),
T_A = ambient temperature.

The thermal resistance junction to ambient is the sum of the individual components. Each component must be minimized if the lowest junction temperature is to result.

The value for the interface thermal resistance, $R_{\theta CS}$, may be significant compared to the other thermal resistance terms. A proper mounting procedure can minimize $R_{\theta CS}$.

The thermal resistance of the heatsink is not absolutely constant; its thermal efficiency increases as ambient temperature increases and it is also affected by orientation of the sink. The thermal resistance of the semiconductor is also variable; it is a function of biasing and temperature. Semiconductor thermal resistance specifications are normally at conditions where current density is fairly uniform. In some applications, such as in RF power amplifiers and short–pulse applications, current density is not uniform and localized heating in the semiconductor chip will be the controlling factor in determining power handling ability.

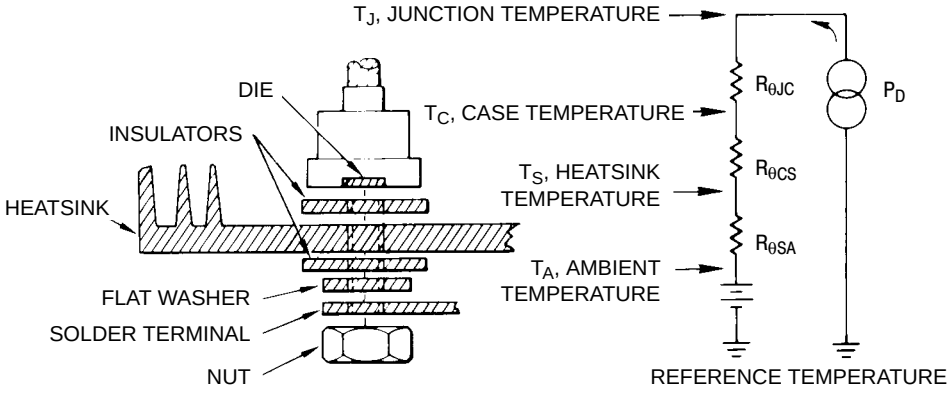


Figure 35. Basic Thermal Resistance Model Showing Thermal to Electrical Analogy for a Semiconductor

APPENDIX B
MEASUREMENT OF INTERFACE THERMAL RESISTANCE

Measuring the interface thermal resistance $R_{\theta CS}$ appears deceptively simple. All that's apparently needed is a thermocouple on the semiconductor case, a thermocouple on the heatsink, and a means of applying and measuring DC power. However, $R_{\theta CS}$ is proportional to the amount of contact area between the surfaces and consequently is affected by surface flatness and finish and the amount of pressure on the surfaces. The fastening method may also be a factor. In addition, placement of the thermocouples can have a significant influence upon the results. Consequently, values for interface thermal resistance presented by different manufacturers are not in good agreement. Fastening methods and thermocouple locations are considered in this Appendix.

When fastening the test package in place with screws, thermal conduction may take place through the screws, for example, from the flange ear on a TO–3 package directly to the heatsink. This shunt path yields values which are artificially low for the insulation material and dependent upon screw head contact area and screw material. MIL–I–49456 allows screws to be used in tests for interface thermal resistance probably because it can be argued that this is "application oriented."

Thermalloy takes pains to insulate all possible shunt conduction paths in order to more accurately evaluate insulation materials. The ON Semiconductor fixture uses an insulated clamp arrangement to secure the package which also does not provide a conduction path.

As described previously, some packages, such as a TO–220, may be mounted with either a screw through the tab or a clip bearing on the plastic body. These two methods often yield different values for interface thermal resistance. Another discrepancy can occur if the top of the package is exposed to the ambient air where radiation and convection can take place. To avoid this, the package should be covered with insulating foam. It has been estimated that a 15 to 20% error in $R_{\theta CS}$ can be incurred from this source.

Another significant cause for measurement discrepancies is the placement of the thermocouple to measure the semiconductor case temperature. Consider the TO–220 package shown in 36. The mounting pressure at one end causes the other end – where the die is located – to lift off the mounting surface slightly. To improve contact, ON Semiconductor TO–220 Packages are slightly concave. Use of a spreader bar under the screw lessens the lifting, but some is inevitable with a package of this structure. Three thermocouple locations are shown:

1. The ON Semiconductor location is directly under the die reached through a hole in the heatsink. The thermocouple is held in place by a spring which forces the thermocouple into intimate contact with the bottom of the semi's case.
2. The JEDEC location is close to the die on the top surface of the package base reached through a blind hole drilled through the molded body. The thermocouple is swaged in place.
3. The Thermalloy location is on the top portion of the tab between the molded body and the mounting screw. The thermocouple is soldered into position.

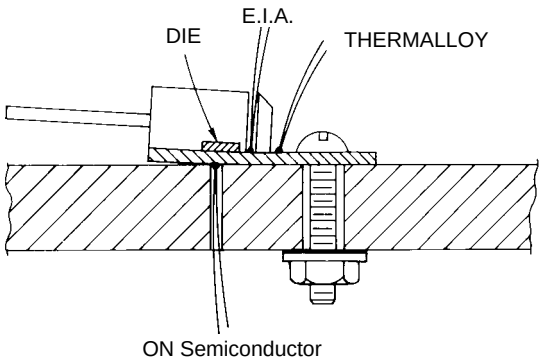


Figure 36. JEDEC TO–220 Package Mounted to Heatsink Showing Various Thermocouple Locations and Lifting Caused by Pressure at One End

Temperatures at the three locations are generally not the same. Consider the situation depicted in the figure. Because the only area of direct contact is around the mounting screw, nearly all the heat travels horizontally along the tab from the die to the contact area. Consequently, the temperature at the JEDEC location is hotter than at the Thermalloy location and the ON Semiconductor location is even hotter. Since junction–to–sink thermal resistance must be constant for a given test setup, the calculated junction–to–case thermal resistance values decrease and case–to–sink values increase as the "case" temperature thermocouple readings become warmer. Thus, the choice of reference point for the "case" temperature is quite important.

There are examples where the relationship between the thermocouple temperatures are different from the previous situation. If a mica washer with grease is installed between the semiconductor package and the heatsink, tightening the screw will not bow the package; instead, the mica will be deformed. The primary heat conduction path is from the die through the mica to the heatsink. In this case, a small temperature drop will exist across the vertical dimension of the package mounting base so that the thermocouple at the EIA location will be the hottest. The thermocouple temperature at the Thermalloy location will be lower, but close to the temperature at the EIA location as the lateral heat flow is generally small. The ON Semiconductor location will be coolest.

The EIA location is chosen to obtain the highest temperature on the case. It is of significance because power ratings are supposed to be based on this reference point. Unfortunately, the placement of the thermocouple is tedious and leaves the semiconductor in a condition unfit for sale.

The ON Semiconductor location is chosen to obtain the highest temperature of the case at a point where, hopefully, the case is making contact to the heatsink. Once the special heatsink, to accommodate the thermocouple has been fabricated, this method lends itself to production testing and does not mark the device. However, this location is not easily accessible to the user.

The Thermalloy location is convenient and is often chosen by equipment manufacturers. However, it also blemishes the case and may yield results differing up to 1°C/W for a TO–220 package mounted to a heatsink without thermal grease and no insulator. This error is small when compared to the thermal resistance of heat dissipaters often used with this package, since power dissipation is usually a few watts. When compared to the specified junction–to–case values of some of the higher power semiconductors becoming available, however, the difference becomes significant and it is important that the semiconductor manufacturer and equipment manufacturer use the same reference point.

Another EIA method of establishing reference temperatures utilizes a soft copper washer (thermal grease is used) between the semiconductor package and the heatsink. The washer is flat to within 1 mil/inch, has a finish better than 63 μ–inch, and has an imbedded thermocouple near its center. This reference includes the interface resistance under nearly ideal conditions and is, therefore, application–oriented. It is also easy to use, but has not become widely accepted.

A good way to improve confidence in the choice of case reference point is to also test for junction–to–case thermal resistance while testing for interface thermal resistance. If the junction–to–case values remain relatively constant as insulators are changed, torque varied, etc., then the case reference point is satisfactory.

APPENDIX C
Sources of Accessories

Manufacturer	Joint Compound	Adhesives	Insulators						Heatsinks	Clips
			BeO	AlO_2	Anodize	Mica	Plastic Film	Silicone Rubber		
Aavid	–	–	–	–	–	–	X	X	X	X
AHAM–TOR	–	–	–	–	–	–	–	–	X	–
Asheville–Schoonmaker	–	–	–	–	–	X	–	–	–	–
Astrodynamis	X	–	–	–	–	–	–	–	X	–
Delbert Blinn	–	–	X	–	X	X	X	X	X	–
IERC	X	–	–	–	–	–	–	–	X	–
Staver	–	–	–	–	–	–	–	–	X	–
Thermalloy	X	X	X	X	X	X	X	X	X	X
Tran–tec	X	–	X	X	X	X	–	X	X	–
Wakefield	X	X	X	–	X	–	–	X	X	X

Other Sources for silicone rubber pads: Chomerics, Bergquist

Suppliers Addresses

Aavid Engineering, Inc., P.O. Box 400, Laconia, New Hampshire 03247 (603) 524–1478
AHAM–TOR Heatsinks, 27901 Front Street, Rancho, California 92390 (714) 676–4151
Asheville–Schoonmaker, 900 Jefferson Ave., Newport News, VA 23607 (804) 244–7311
Astro Dynamics, Inc., 2 Gill St., Woburn, Massachusetts 01801 (617) 935–4944
Bergquist, 5300 Edina Industrial Blvd., Minneapolis, Minnesota 55435 (612) 835–2322
Chomerics, Inc.,16 Flagstone Drive, Hudson, New Hampshire 03051 1–800–633–8800
Delbert Blinn Company, P.O. Box 2007, Pomona, California 91769 (714) 623–1257
International Electronic Research Corporation, 135 West Magnolia Boulevard, Burbank, California 91502 (213) 849–2481
The Staver Company, Inc., 41–51 Saxon Avenue, Bay Shore, Long Island, New York 11706 (516) 666–8000
Thermalloy, Inc., P.O. Box 34829, 2021 West Valley View Lane, Dallas, Texas 75234 (214) 243–4321
Tran–tec Corporation, P.O. Box 1044, Columbus, Nebraska 68601 (402) 564–2748
Wakefield Engineering, Inc., Wakefield, Massachusetts 01880 (617) 245–5900

PACKAGE INDEX

PREFACE

When the JEDEC registration system for package outlines started in 1957, numbers were assigned sequentially whenever manufacturers wished to establish a package as an industry standard. As minor variations developed from these industry standards, either a new, non–related number was issued by JEDEC or manufacturers would attempt to relate the part to an industry standard via some appended description.

In an attempt to ease confusion, JEDEC established the present system in late 1968 in which new packages are assigned into a category, based on their general physical appearance. Differences between specific packages in a category are denoted by suffix letters. The older package designations were re–registered to the new system as time permitted.

For example the venerable TO–3 has many variations. Can heights differ and it is available with 30, 40, 50, and 60 mil pins, with and without lugs. It is now classified in the TO–204 family. The TO–204AA conforms to the original outline for the TO–3 having 40 mil pins while the TO–204AE has 60 mil pins, for example.

The new numbers for the old parts really haven't caught on very well. It seems that the DO–4, DO–5 and TO–3 still convey sufficient meaning for general verbal communication.

ON Case Number	JEDEC Outline		Notes	Mounting Class
	Original System	Revised System		
001	TO-3	TO-204AA		Flange
003	TO-3		2	Flange
009	TO-61	TO-210AC		Stud
011	TO-3	TO-204AA	–	Flange
011A	TO-3	–	2	Flange
012	TO-3	–	2	Flange
036	TO-60	TO-210AB	–	Stud
042A	DO-5	DO-203AB	–	Stud
044	DO-4	DO-203AA	–	Stud
054	TO-3	–	2	Flange
056	DO-4	–	–	Stud
058	DO-5	–	2	Stud
61-04				Flange
63-02	TO-64	TO-208AB		Stud
63-03	TO-64	TO-2088AB		Stud
077	TO-126	TO-225AA	–	Plastic
080	TO-66	TO-213AA	–	Flange
086	–	TO-208	1	Stud
086L	–	TO-298	1	Stud
144B-05				Stud
145A-09				Stud
145A-10				Stud
145C	TO-232		1	Stud
157	–	DO-203	1	Stud
160-03	TO-59	TO-210AA	–	Stud
167	–	DO-203	1	Stud
174-04				Pressfit

ON Case Number	JEDEC Outline		Notes	Mounting Class
	Original System	Revised System		
175-03				Stud
197	–	TO-204AE	–	Flange
211-07				Flange
211-11				Flange
215-02				Flange
221	–	TO-220AB	–	Tab
221C-02				Plastic
221D-02	–	–	Isolated TO-220	Plastic
235	–	TO-208	1	Stud
235-03				Stud
238	–	TO-208	1	Stud
239	–	TO-208	–	Stud
244-04				Stud
245	DO-4	–	–	Stud
257-01	DO-5	–	–	Stud
263	–	TO-208	–	Stud
263-04				Stud
283	DO-4	–	–	Stud
289	–	TO-209	1	Stud
305-01				Stud
310-02				Pressfit
311-02			Isolated	Stud
311-02				Pressfit
311-02				Stud
314B-03				Tab

ON Case Number	JEDEC Outline		Notes	Mounting Class
	Original System	Revised System		
314D-03				Tab
316-01				Flange
319-06				Flange
328A-03				Flange
332-04				Stud
333-04				Flange
333A-02				Flange
336-03				Flange
337-02				Flange
340		TO-218AC		Tab
340A-02				Plastic
340B-03			Isolated TO-218	Plastic
342-01				Flange
357B-01				Flange
361-01				Flange
368-02				Flange
369-06		TO-251		Insertion
369A-12		TO-252		Surface
373-01			Isolated	Flange
383-01			Isolated	Flange
387-01		TO-254AA	Isolated 2	Tab
388A-01		TO-258AA	Isolated 2	Tab
744-02				Flange
744A-01				Flange
043-07	DO-21	DO-208AA		Pressfit

Notes: 1. Would fit within this family outline if registered with JEDEC.
2. Not within all JEDEC dimensions.

(1) MIL–HANDBOOK – 2178, SECTION 2.2.
(2) "Navy Power Supply Reliability – Design and Manufacturing Guidelines" NAVMAT P4855–1, Dec. 1982 NAVPUBFORCEN, 5801 Tabor Ave., Philadelphia, PA 19120.
(3) Catalog #87–HS–9, (1987), page 8, Thermalloy, Inc., P.O. Box 810839, Dallas, Texas 75381–0839.
(4) ITW Shakeproof, St. Charles Road, Elgin, IL 60120.
(5) Robert Batson, Elliot Fraunglass and James P Moran, "Heat Dissipation Through Thermalloy Conductive Adhesives," EMTAS '83. Conference, February 1 – 3, Phoenix, AZ; Society of Manufacturing Engineers, One SME Drive, P.O. Box 930, Dearborn, MI 48128.
(6) Catalog, Edition 18, Richco Plastic Company, 5825 N. Tripp Ave., Chicago, IL 60546.
(7) Herb Fick, "Thermal Management of Surface Mount Power Devices," Powerconversion and Intelligent Motion, August 1987.

Grafoil is a registered Union Carbide.
Kon–Dux and Rubber–Duc are trademarks of Aavid Thermal Technologies, Inc.
Thermasil is a registered trademark and Thermafilm is a trademark of Thermalloy, Inc.
Kapton is a registered trademark of du Pont de Nemours & Co., Inc.
Sil–Pad is a registered trademark of the Bergquist Company.
CHO–THERM is a registered trademark of Chomerics, Inc.
FULLPAK, ICePAK and Thermopad are trademarks of Semiconductor Componenets Industries, LLC.

ON Semiconductor and are trademarks of Semiconductor Components Industries, LLC (SCILLC). SCILLC reserves the right to make changes without further notice to any products herein. SCILLC makes no warranty, representation or guarantee regarding the suitability of its products for any particular purpose, nor does SCILLC assume any liability arising out of the application or use of any product or circuit, and specifically disclaims any and all liability, including without limitation special, consequential or incidental damages. “Typical” parameters which may be provided in SCILLC data sheets and/or specifications can and do vary in different applications and actual performance may vary over time. All operating parameters, including “Typicals” must be validated for each customer application by customer’s technical experts. SCILLC does not convey any license under its patent rights nor the rights of others. SCILLC products are not designed, intended, or authorized for use as components in systems intended for surgical implant into the body, or other applications intended to support or sustain life, or for any other application in which the failure of the SCILLC product could create a situation where personal injury or death may occur. Should Buyer purchase or use SCILLC products for any such unintended or unauthorized application, Buyer shall indemnify and hold SCILLC and its officers, employees, subsidiaries, affiliates, and distributors harmless against all claims, costs, damages, and expenses, and reasonable attorney fees arising out of, directly or indirectly, any claim of personal injury or death associated with such unintended or unauthorized use, even if such claim alleges that SCILLC was negligent regarding the design or manufacture of the part. SCILLC is an Equal Opportunity/Affirmative Action Employer.

PUBLICATION ORDERING INFORMATION

NORTH AMERICA Literature Fulfillment:
Literature Distribution Center for ON Semiconductor
P.O. Box 5163, Denver, Colorado 80217 USA
Phone: 303–675–2175 or 800–344–3860 Toll Free USA/Canada
Fax: 303–675–2176 or 800–344–3867 Toll Free USA/Canada
Email: ONlit@hibbertco.com
Fax Response Line: 303–675–2167 or 800–344–3810 Toll Free USA/Canada

N. American Technical Support: 800–282–9855 Toll Free USA/Canada

EUROPE: LDC for ON Semiconductor – European Support
German Phone: (+1) 303–308–7140 (Mon–Fri 2:30pm to 7:00pm CET)
Email: ONlit–german@hibbertco.com
French Phone: (+1) 303–308–7141 (Mon–Fri 2:00pm to 7:00pm CET)
Email: ONlit–french@hibbertco.com
English Phone: (+1) 303–308–7142 (Mon–Fri 12:00pm to 5:00pm GMT)
Email: ONlit@hibbertco.com

EUROPEAN TOLL–FREE ACCESS*: 00–800–4422–3781
*Available from Germany, France, Italy, UK, Ireland

CENTRAL/SOUTH AMERICA:
Spanish Phone: 303–308–7143 (Mon–Fri 8:00am to 5:00pm MST)
Email: ONlit–spanish@hibbertco.com
Toll–Free from Mexico: Dial 01–800–288–2872 for Access –
then Dial 866–297–9322

ASIA/PACIFIC: LDC for ON Semiconductor – Asia Support
Phone: 1–303–675–2121 (Tue–Fri 9:00am to 1:00pm, Hong Kong Time)
Toll Free from Hong Kong & Singapore:
001–800–4422–3781
Email: ONlit–asia@hibbertco.com

JAPAN: ON Semiconductor, Japan Customer Focus Center
4–32–1 Nishi–Gotanda, Shinagawa–ku, Tokyo, Japan 141–0031
Phone: 81–3–5740–2700
Email: r14525@onsemi.com

ON Semiconductor Website: http://onsemi.com

For additional information, please contact your local
Sales Representative.

Elliott Sound Products **Project 25**

Phono Preamps For All

Diagrams by Richard Crowley, Text by Rod Elliott

Introduction

This is a collection of phono (black vinyl for the youngsters) preamps and equalisation circuits, one of which is sure to meet your requirements. These are not my circuits, but were contributed by a reader (see above), so I am not really in a position to make any specific recommendations. They are provided as a service to the experimenters out there, and may be found useful for other applications as well.

Moving Coil Preamps

The first set of offerings are moving coil preamps, specifically intended for the very low output voltage and impedance of the majority of moving coil pickups. These were much favoured in their day for superior quality over the entire audio range, and for any serious listening I still use my moving coil pickup.

MC Preamp #1

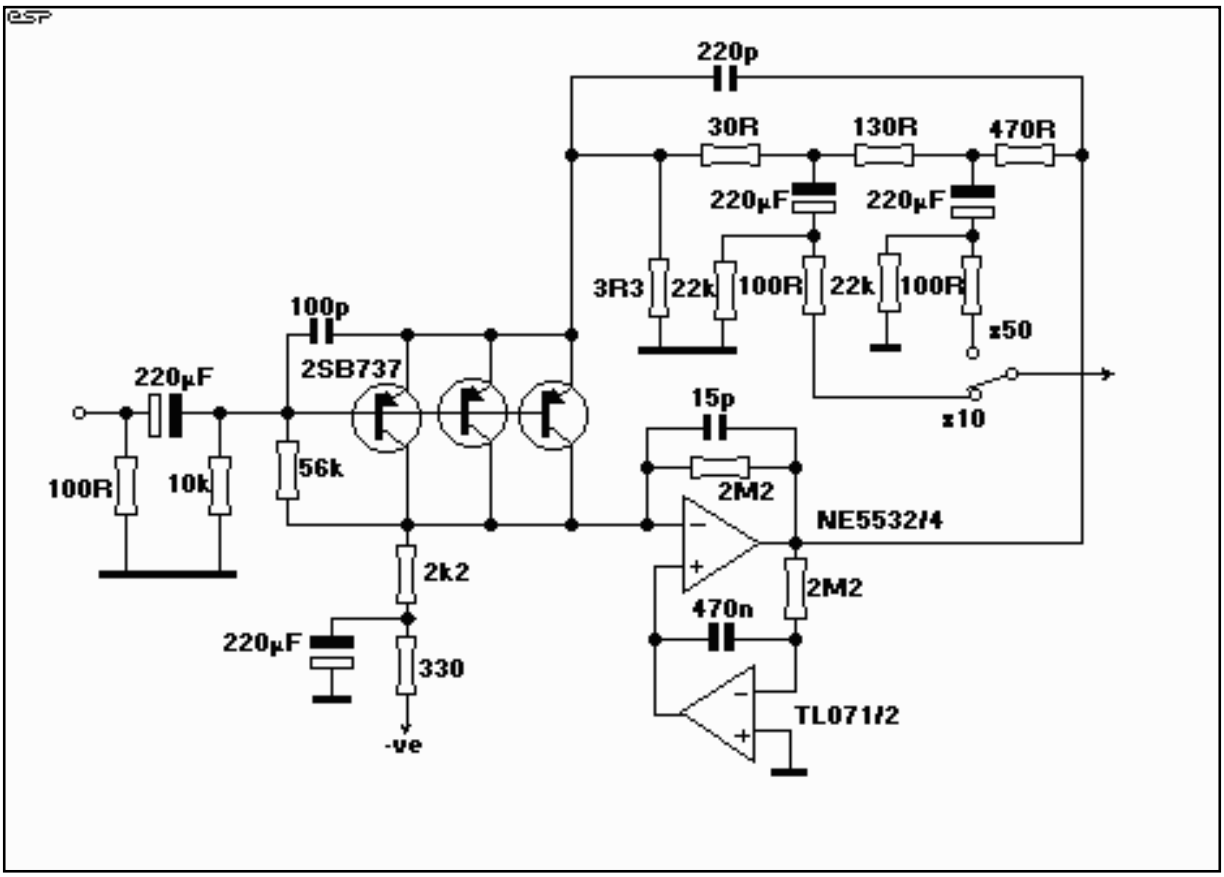


Figure 1 - Moving Coil Preamplifier (After Douglas Self)

This design uses multiple transistors as the initial amplifying stage. The transistors chosen have very low noise, and this is reduced even further by the parallel technique.

As can be seen from the diagram, the circuit gain can be changed to suit high and low output cartridges with a single switch. The gains as shown are x10 and x50 (20 dB and 34 dB respectively), but could be modified if desired. I will leave this up to the reader to experiment with.

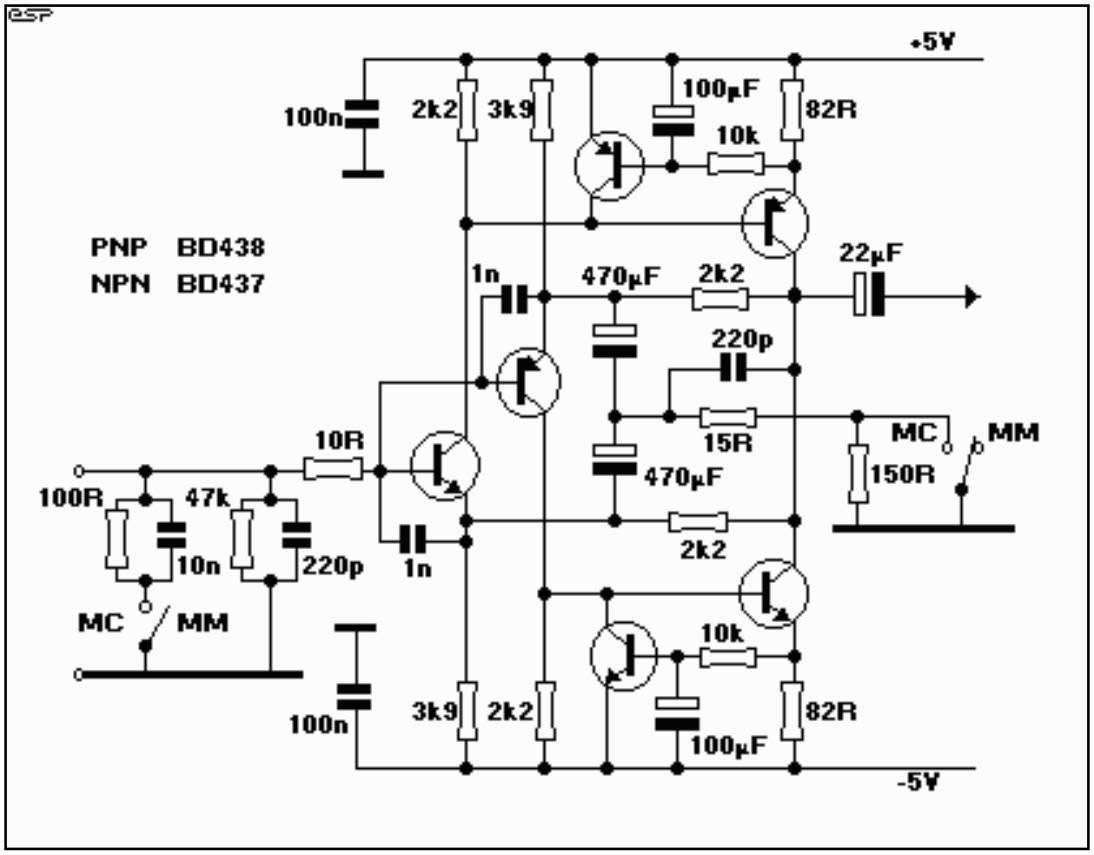
As I look at the circuit, I'm not sure why the gain resistors were set up as they are. With the values shown, the circuit actually has a gain of 192 (set the feedback resistors in series and the 3.3 Ohm emitter resistor), and this is then attenuated to provide the gains of 50 and 10 as shown. I'm sure there was an excellent reason for this arrangement, but I cannot see it.

The second opamp (TLO71/72) acts as a bias servo device, and ensures that the inverting input of the NE5532 is at the same voltage as its non-inverting input. With this arrangement, output offset voltage can be expected to be very low - typically no more that a couple of millivolts.

MC Preamp #2

This next design is based on the work of John Linsley-Hood, and again can be expected to give

good results. As can be seen, there are no opamps used, and the circuit is completely symmetrical. This offers a reduction in noise over a single circuit, since the two complete mirror image halves are in parallel. Distortion is probably reduced as well, but at the signal levels produced by a moving coil pickup, this is not likely to be significant.



Moving Coil Preamp (After John Linsley-Hood)

The circuit is (more or less) conventional, using a common emitter stage modulating a constant-current source. As the circuit is mirrored, these are reproduced for both positive and negative supplies. The net output voltage will again have a low DC offset, but probably somewhat higher than the previous design due to the lack of a bias "servo" to maintain an exact 0V DC output.

Gain is set by the 2.2k resistors from the collectors of the output devices (in parallel, so 1.1k) and the 15 Ohm resistor (MC or Moving Coil position) optionally in series with the 150 Ohm (MM or Moving Magnet position).

These give a theoretical gain of 74 (37 dB) in the MC position, or 7.6 (17.6 dB) for the MM setting. In practice, one can expect the gain to be very slightly less than these figures, especially for the MC setting.

I cannot comment on the relative noise performance of these two preamps, but as they are both based on designs by well respected audio designers, I would expect that noise would not be an issue with either circuit.

Phono Equalisers

With two different circuits and several additional EQ networks to choose from (plus my own version, published as Project 06), there has to be one for you!

I do not know the origin of these circuits (other than from Richard), so cannot be too specific about them. They are both reasonably conventional, as can be seen.

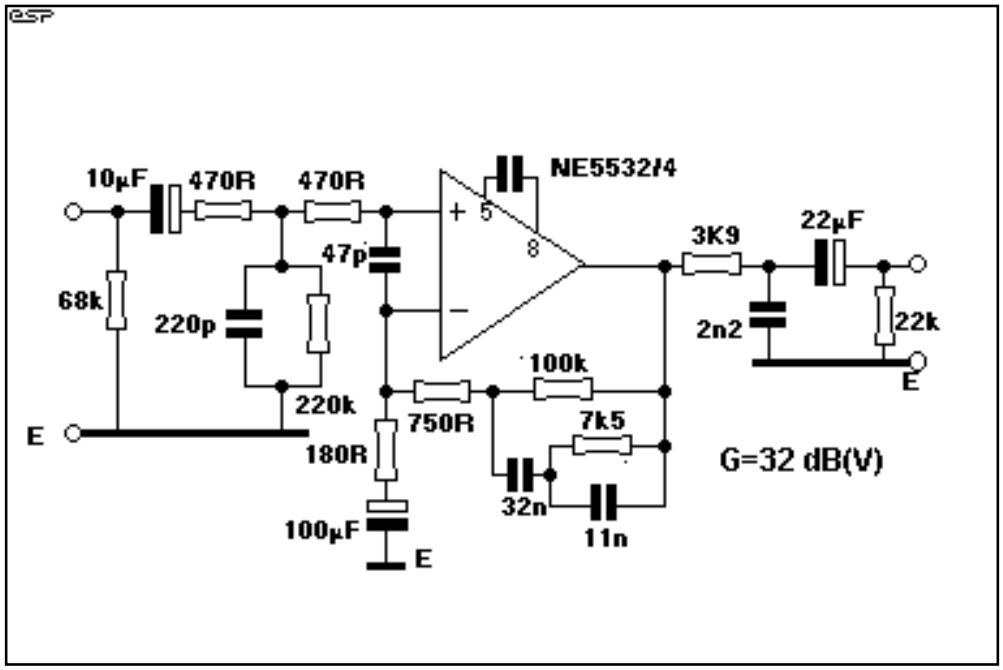


Figure 3 - Phono Preamp #1

This is a simple, one opamp phono preamp, but has a few added features. These are mainly for radio frequency (RF) suppression, and with the input circuit shown can be expected to be highly resistant to even high levels of RF interference.

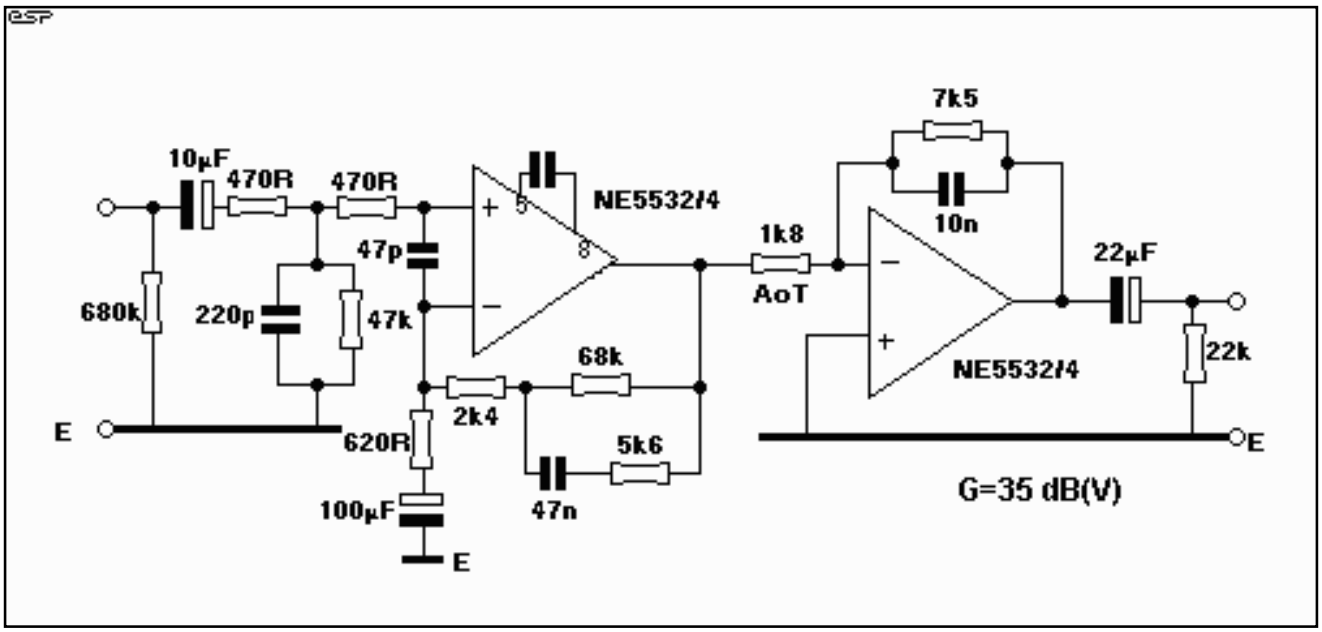


Figure 4 - Phono Preamp #2

This preamp splits the equalisation into two stages - the first stage provides the low frequency boost required by the RIAA specification, and the second reduces the high frequency component (again, to the RIAA spec.).

The advantage of this approach is that the two filter sections have less interaction, and much of the circuit noise produced by the first (and second) stage is attenuated by the top-cut circuit.

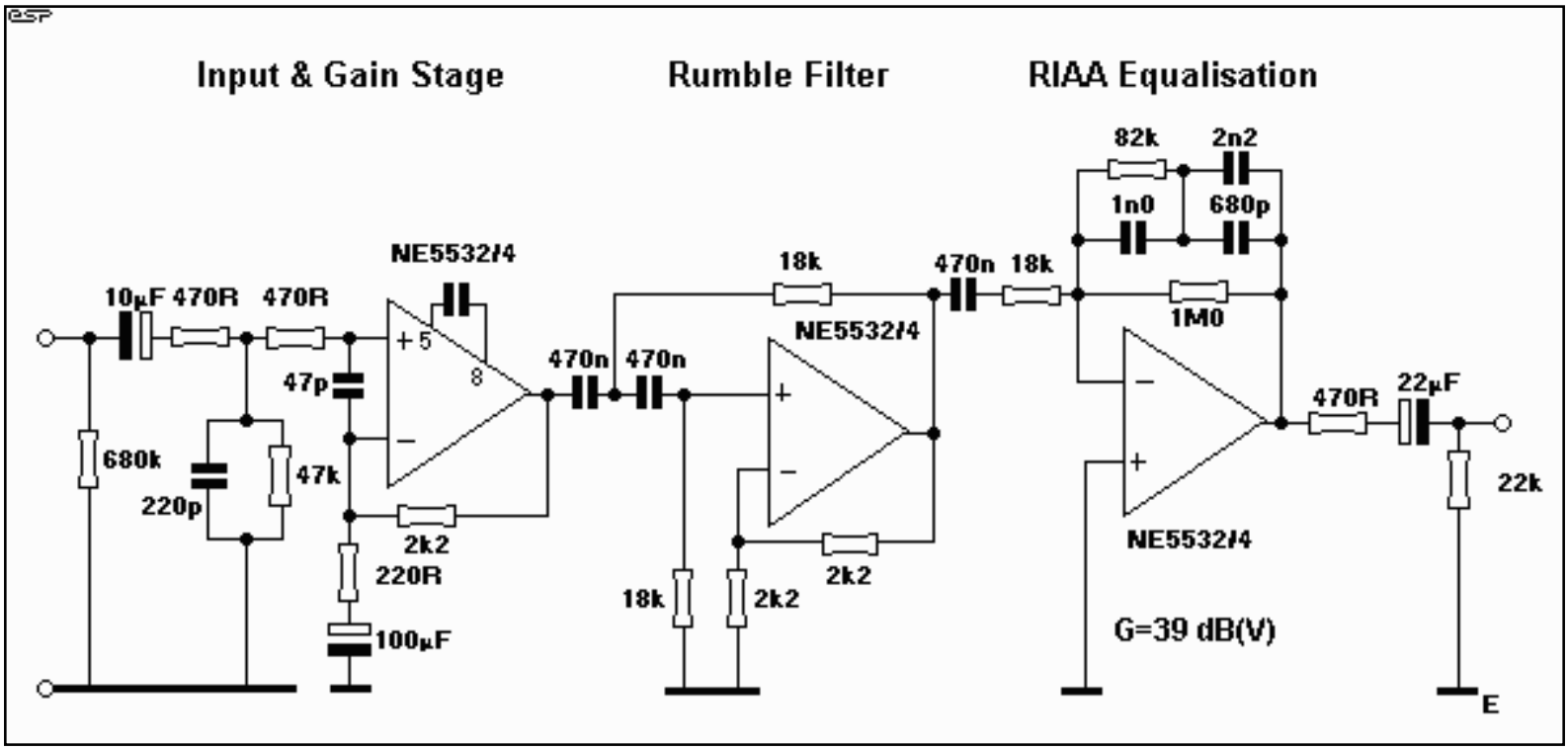


Figure 5 - Phono Preamp #3

As can be seen, this preamp is considerably more complicated than the first two, but includes a

3rd order rumble filter with a cutoff frequency of about 18 Hz.

The first stage is a simple amplifier, again with complete RF interference protection. This is followed by the rumble (infra-sonic) filter, and finally the equalisation stage. Note that in this circuit arrangement, the opamp is operating as an inverting amplifier, which has no bearing on the final result, but is (very) slightly noisier than the non-inverting configuration. This is unlikely to be audible in practice, since the gain contributed by the final stage is much lower than normal.

Total gain is 39 dB, of which 17.5 dB is contributed by the input stage, 6 dB by the filter, leaving only 15.5 dB gain in the final stage. All preceeding high frequency opamp noise is attenuated by the RIAA equalisation, leading to a design which should have an excellent overall noise figure.

Equalisation Networks

These EQ networks can be used around any opamp, and will provide RIAA equalisation to an accuracy as shown next to each diagram. Some care must be used to ensure that the feedback resistor (to ground) is selected to give the required gain.

This is specified at 1 kHz for all phono preamps, because of the frequency characteristics of the filter network.

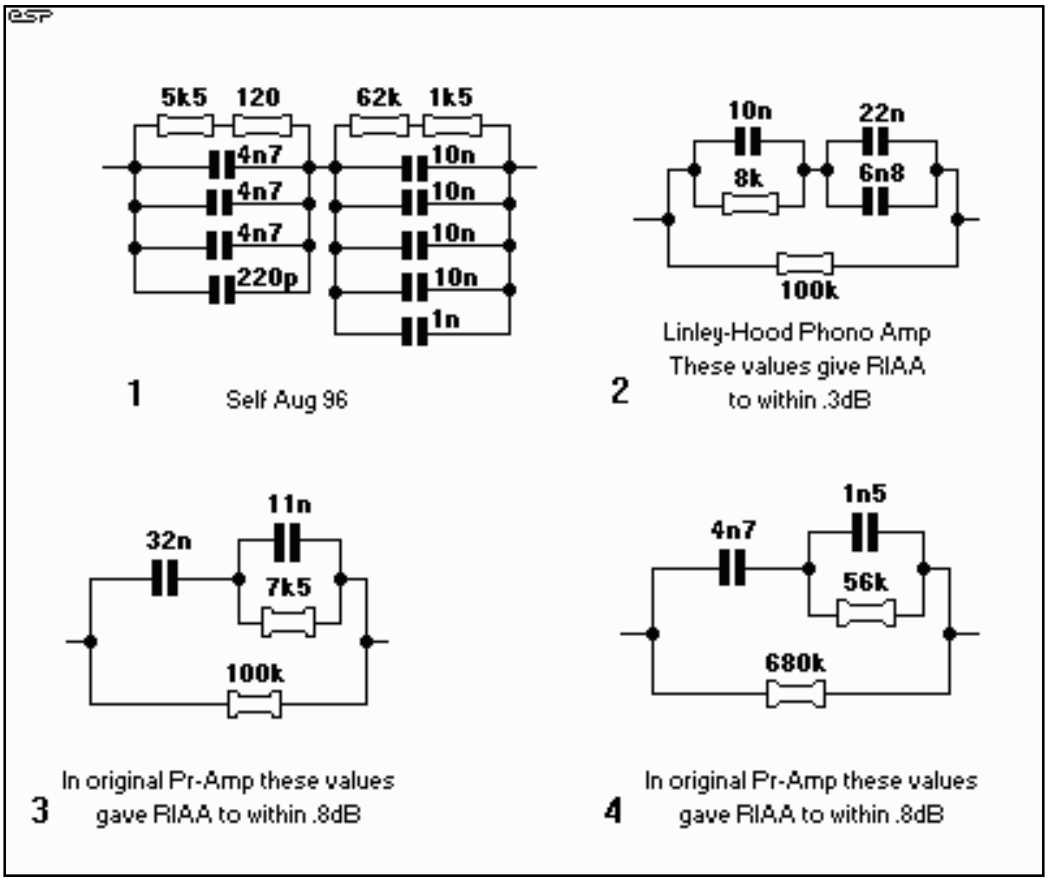


Figure 6 - Various RIAA Phono EQ Networks

Note - I have been advised by a reader that the 1nF capacitor in Figure 6-1 should be 10nF. I do not have the original circuit diagram for this preamp, so cannot be sure, but it is worth bearing in mind if you use the EQ network. A calculation reveals that there should be 50nF in parallel with 63,500 ohms to obtain 50Hz, so I expect that the 1nF cap is indeed incorrect.

The 1 kHz impedance of each network is quite different, so the required feedback resistor has been calculated for a nominal gain of 35 dB (x 56) . These work out to (approximately)

Circuit Nbr.	Network Z	Feedback Res.	Actual Gain
1	7.75 k Ohms	138 Ohms	56.6
2	10 k Ohms	180 Ohms	57.0
3	9.43 k Ohms	168 Ohms	56.6
4	67 k Ohms	1.2 k Ohms	56.8

With this in mind, the 1 kHz gain resistor can be calculated. For example, using (1) above, for a gain of 35 dB (a gain of 56 for ease of calculation), this will need a resistor to ground from the -ve input of the opamp of 138 Ohms.

RIAA Equalisation Response

For the sake of reference, the table below shows the response curves for both RIAA equalisation, and the IEC modified version. The latter was not very popular, because it represented a loss of bass (-3 dB ref RIAA at 20 Hz). For anyone wanting to know a little more about how and why the equalisation was done in this way, read on.....

Hz	*RIAA*	*IEC*	*Hz*	*RIAA*	*IEC*	*Hz*	*RIAA*	*IEC*
20	19.36	16.35	**240**	7.04	7.01	**2400**	-3.39	-3.39
22	19.24	16.62	**270**	6.25	6.23	**2700**	-4.04	-4.04
25	19.04	16.89	**300**	5.57	5.55	**3000**	-4.65	-4.65
28	18.83	17.04	**340**	4.80	4.79	**3400**	-5.43	-5.43
31	18.61	17.09	**380**	4.16	4.15	**3800**	-6.17	-6.17
35	18.29	17.06	**430**	3.49	3.48	**4300**	-7.02	-7.02
39	17.96	16.95	**480**	2.93	2.92	**4800**	-7.82	-7.82
44	17.54	16.73	**540**	2.38	2.38	**5400**	-8.70	-8.70
49	17.12	16.45	**610**	1.86	1.86	**6100**	-9.64	-9.64

55	16.61	16.07	**680**	1.43	1.43	**6800**	-10.50	-10.50
62	16.02	15.59	**760**	1.02	1.02	**7600**	-11.39	-11.39
70	15.37	15.03	**850**	0.63	0.63	**8500**	-12.30	-12.30
79	14.67	14.40	**950**	0.26	0.26	**9500**	-13.22	-13.22
89	13.93	13.72	**1100**	-0.23	-0.23	**11000**	-14.44	-14.44
100	13.18	13.01	**1200**	-0.52	-0.52	**12000**	-15.17	-15.17
110	12.54	12.39	**1300**	-0.79	-0.79	**13000**	-15.85	-15.85
120	11.94	11.82	**1500**	-1.31	-1.31	**15000**	-17.07	-17.07
130	11.38	11.27	**1700**	-1.80	-1.80	**17000**	-18.14	-18.14
150	10.36	10.28	**1900**	-2.27	-2.27	**19000**	-19.09	-19.09
170	9.46	9.40	**2100**	-2.73	-2.73	**21000**	-19.95	-19.95
190	8.67	8.62						
210	7.97	7.93						

RIAA Time Constants

The time constants (**2 π f**) that provide RIAA equalisation are

- High Frequency = 75us (2122 Hz)
- Mid Frequency = 318us (500.5 Hz)
- Low Frequency = 3180us (50.05 Hz)

The basic principle behind the equalisation curve was quite simple, and was designed to reduce the grove modulation to a manageable level (both for the cutter and the reproducer), and provide some basic high frequency noise reduction.

With this in mind, frequencies below 500 Hz were "cut" using constant amplitude, which means that the signal from the cartridge will increase at 6 dB / Octave, since if the amplitude remains constant, the velocity must increase with the frequency. Because the output of a magnetic cartridge is dependent upon the velocity, bass boost must be applied to bring the levels back to normal.

It is this rather large amount of bass boost that accentuates the mechanical noise of a turntable, producing what is commonly known as "rumble". There is also the risk of low frequency feedback if the turntable is not capable of isolating the platter and tone-arm from the listening room environment. Timber floors and rigid suspension increase the risk of feedback, and many manufacturers went to extreme lengths to provide isolation and very low levels of low frequency noise.

Above 500Hz, the cutter mode changed to constant velocity, so the output of the cartridge will now be independent of frequency. Vinyl (or any other material for that matter) will collect dust, and will also have minor surface imperfections, so all signals above 2100 Hz are boosted (again at 6 dB / Octave), so the equalisation curve now applies treble "cut". This brings the signal level back where it should be, and reduces disc surface noise as well.

Return to Projects Index Page

Back to Home Page

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Richard Crowley and Rod Elliott, and is Copyright (c) 1999. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The authors (Richard Crowley and Rod Elliott) grant the reader the right to use this information for personal use only, and further allow that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott and Richard Crowley.

Minor re-format 11 Aug 2000

Elliott Sound Products **Project 17**

'A' Weighting Filter For Audio Measurements

Rod Elliott - ESP (Design based on an old Ampex circuit)

(Updated with improved filter - 29 Aug 2003)

Introduction

I have to say from the outset that I do not agree with the use of weighting filters, since they are not (despite the claims to the contrary) an accurate representation of human hearing.

In fact, the standard "A" weighting curve is accurate at only one SPL (Sound Pressure Level), assuming that the listener has British Standard Ears. I have no idea what SPL this filter is meant to be accurate at, and I doubt that anyone else does either (although at a rough guess I would suggest somewhere around 50dB and below).

When the police measure the noise from a car exhaust or a party, they happily use A-weighting (it's probably in the legislation - that has to be scary - politicians thinking that they know about SPL? Next thing they will tell us that they understand fiscal policy. But I digress .....

The purpose is supposedly to account for the fact that human hearing is less sensitive at low and high frequencies than in the upper midrange, and that this variation is dependent upon the sound intensity (SPL). The Fletcher-Munson curve (as it is commonly known, and reproduced below) shows the variation, and it is clear that any loss of sensitivity is highly dependent upon the actual SPL. The idea that a single filter can represent this at all levels is clearly wrong, but it is a standard nonetheless. (Interestingly, being wrong has never stopped a standard from being imposed, and this is probably truer in the audio industry than almost any other I can think of.)

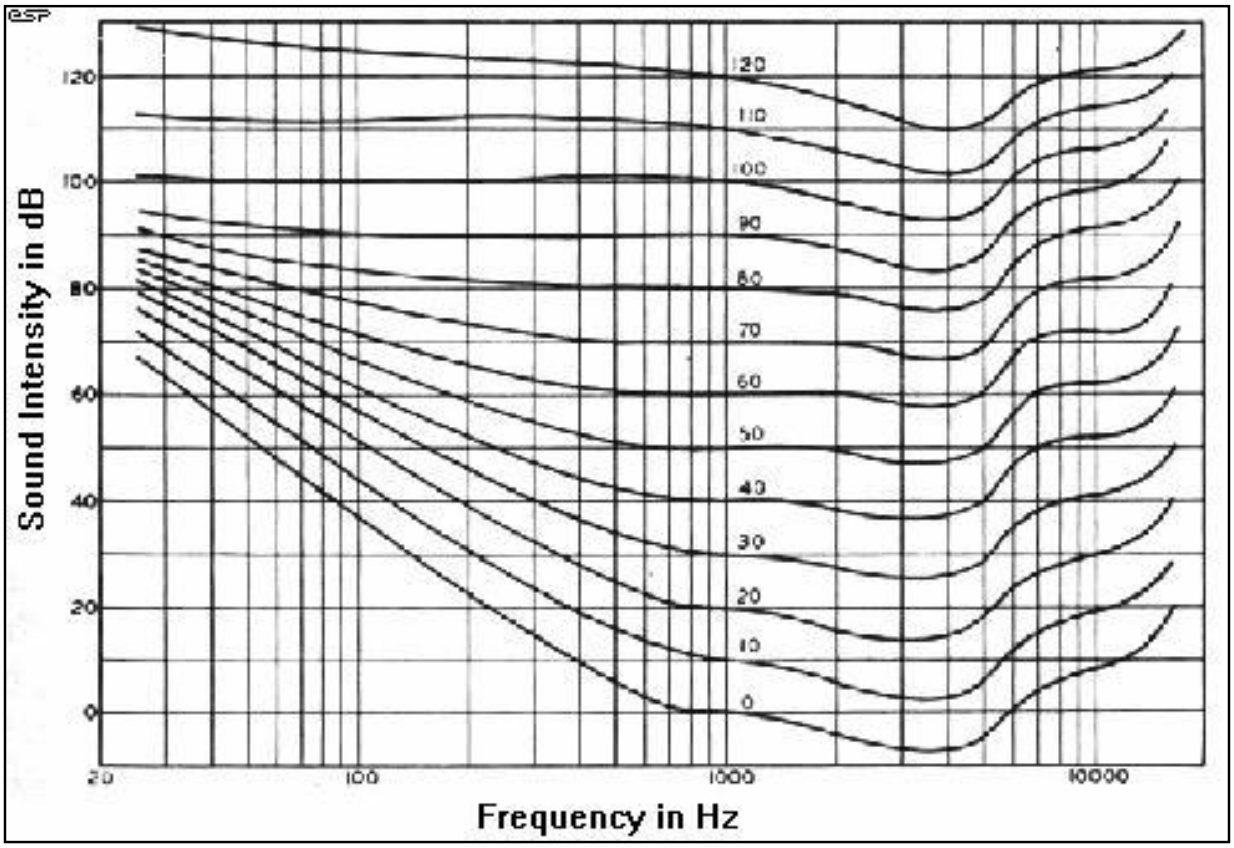


Figure 1 - Equal Loudness Curves (After Fletcher and Munson)

The premise behind all this is that as the SPL is reduced, our ability to detect low or high frequency noise is reduced, so measurements should reflect this phenomenon. While it is undeniable that the chart above represents reality in terms of human hearing [1], I remain unconvinced that A-weighting is a valid test methodology unless the absolute sound intensity is specified.

Ok, I agree that there just might be some validity hiding in there somewhere for noise measurements of amplifiers and the like, but just because the meter tells me that I should not be able to hear the harmonics of the 50/60Hz mains, does not mean that I cannot. There are some sounds that seem (at a casual glance) to defy all measurement standards, and remain audible (albeit at very low level) despite all the "evidence" that this should not be so. As with all such things, experience and practical application are more important than the absolute indication on a meter.

A piece of equipment that is essentially "noise-free" for all intents and purposes is in reality a waste of time, since the ambient noise level in most urban or suburban areas is likely to be far higher than the residual noise of most audio equipment. How useless is 100dB signal to noise ratio for a car hi-fi system, for example? Even the most expensive luxury cars generate far more noise than any tuner/cassette/CD system (and this is apart from all the other external noise generated by other vehicles on the road).

Description

Since it is unlikely that I shall be able to convince the entire industry that it is using flawed reasoning, I shall describe an A-weighting filter so that we can at least make some meaningful comparisons with other systems where this has been used. Note that A-weighting is generally applied only to noise measurements, so *might* have some validity in this respect (as long as the noise we are measuring is of very low amplitude - the neighbour's party is unlikely to fit this mould, but will be measured with A-weighting anyway - oh dear - so much for getting some sleep! (And yet again I digress ....))

The curve of the described filter is shown in Figure 2, and it can be seen that it is essentially a tailored bandpass filter, having a defined rolloff above and below the centre frequency. The reference point is at 1kHz, where the gain is 0dB. The filter response is supposed to be the inverse of one of the curves of the equal loudness graph shown in Figure 1 - it is a little hard to tell which one, but this is a standard, so we shall leave it at that.

As can be seen from Figure 3, the circuit is very simple, but even with this frequency response it is not particularly hard to calibrate accurately so that it really does account for our perception of noise level - especially if an accurate sinewave tone is used as the reference.

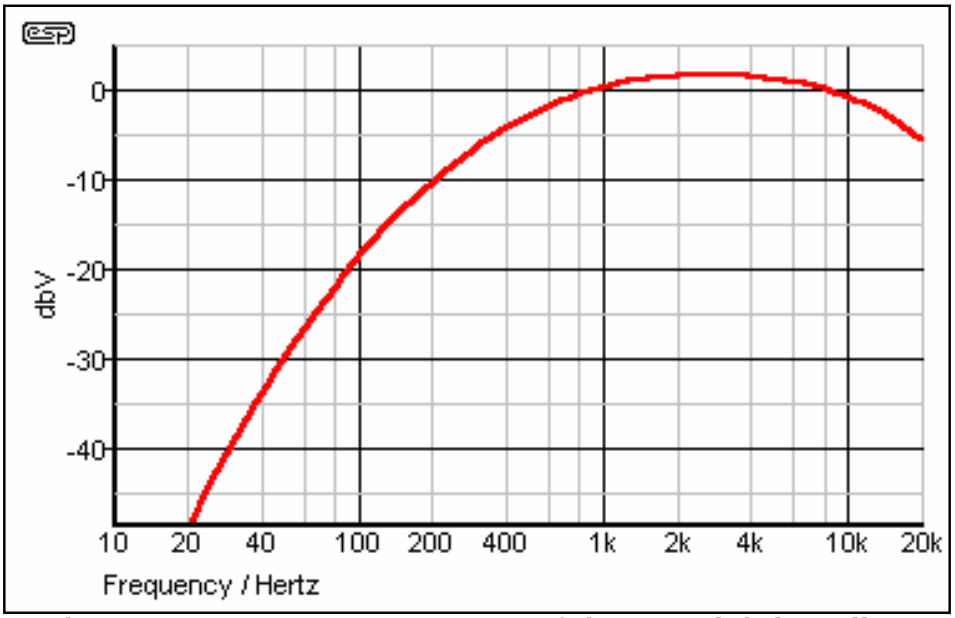


Figure 2 - Frequency Response of the A-Weighting Filter

The filter itself is passive, and the opamps are there only to buffer the input and output, and to adjust the gain so there is some correlation with reality (however slight). Note that the input impedance is quite low, and the output impedance is high, so the unit should be well shielded to prevent noise pickup from the outside world.

As always, I suggest the use of 1% metal film resistors, and the capacitors should be measured and selected, or close tolerance types used. If "ordinary" capacitors are used, their

tolerance will adversely affect the accuracy, but for normal use (i.e. non-certified laboratory), it should be close enough even if 10% caps are used. After all, the noise level of any semiconductor amp is likely to be somewhat variable anyway, so extreme precision is not normally warranted.

The circuit can be operated from a pair of 9 Volt batteries, or a regulated supply of up to +/- 15V. There is no need to use premium opamps unless extremely low noise levels are to be measured, and even then are not needed if there is a gain stage at the front end.

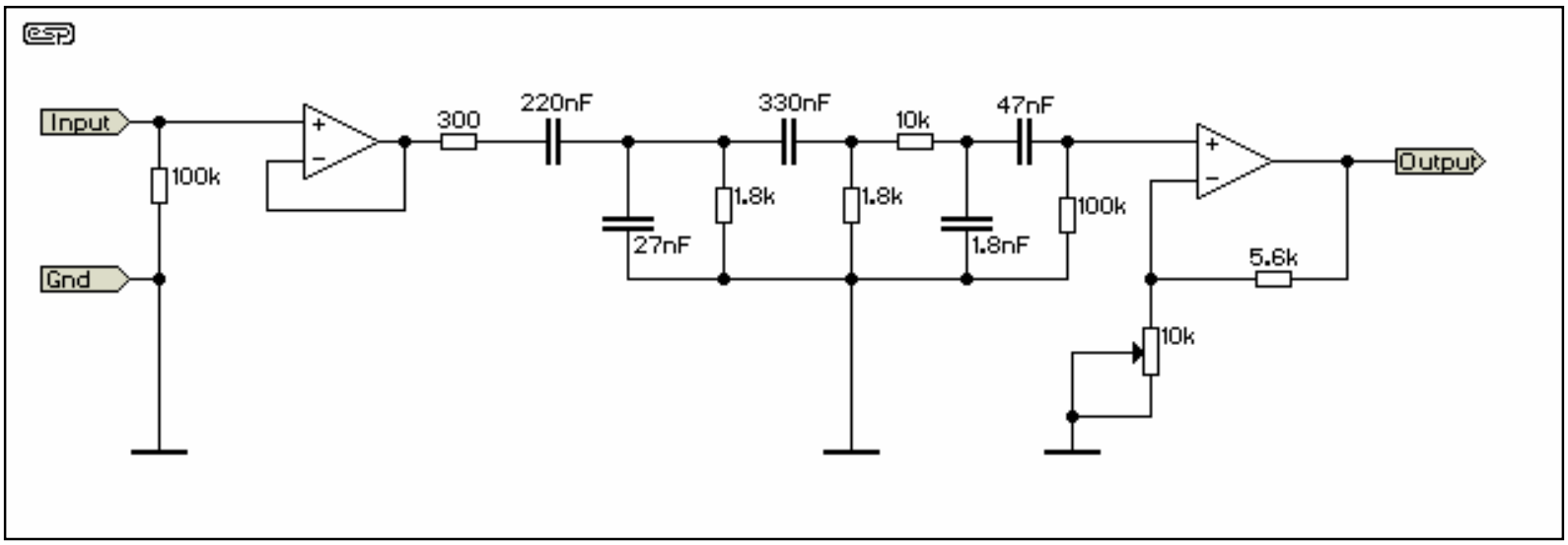


Figure 3 - The A-Weighting Filter Schematic

I will leave it up to the reader to decide on the opamps - I suggest a TL072 dual FET device or similar, which should be ok for most applications (they are not too bad for general purpose work). No opamp pinouts have been included, these are available on any manufacturers' data sheet if you don't know them.

Please note that the filter circuit has a very low impedance at high frequencies. This is liable to stress many opamps, and if a voltage in excess of a few hundred millivolts is expected at high frequencies, I suggest that the driving opamp either be cabable of reasonable current (for example the NE5534) or use two opamps in parallel. At 10kHz and 1V, the opamp will be expected to supply about 8mA.

Basic calibration is not hard - the overall gain at 2,700Hz is supposed to be about 1.3dB, so if the input is set to 1V RMS, the output at 2.7kHz should be 1.162V. Alternatively, at 1kHz, the gain (or insertion loss) should be 0dB - I would suggest that it is checked at both frequencies if possible, and if necessary, average the error between the two frequencies.

Use the 10k trimpot to adjust the level (you need to be accurate with your measurements if true A-weighting is to be obtained). Note that the trimpot should be a quality multi-turn "Cermet" (Ceramic-Metal Film) type to enable accurate setting and long-term stability. Alternatively the trimpot may be replaced with a 5.6k resistor, and accuracy will be quite

acceptable for most applications (the error is less than 0.2dB).

So, there you have it. This project will enable you to make "industry standard" measurements of amplifier noise levels, it is up to you to decide which particular standard you want to make comparisons against. Life would be so much easier if all noise measurements were made "flat" - with no filters of any kind, but this is not to be.

Update - 29 Aug 2002

The filter as originally shown was a little off at 2.7kHz relative to 1kHz (it should be 1.3dB higher at the higher frequency), and this has been corrected. The version shown should be accurate to within about 0.1dB.

It was pointed out (May 2000) that the curve of the original filter shown is not a very good fit to more modern measurement sets, and a small modification will cure this. The low frequency response of the original was not quite what it should be, and at high frequencies the rolloff was too slow. The circuit now shows the final version which is more accurate than the original.

Original modification contributed by

Jürgen Fehringer
Elektronikentwicklung
AKG Acoustics GmbH

My thanks for this useful update.

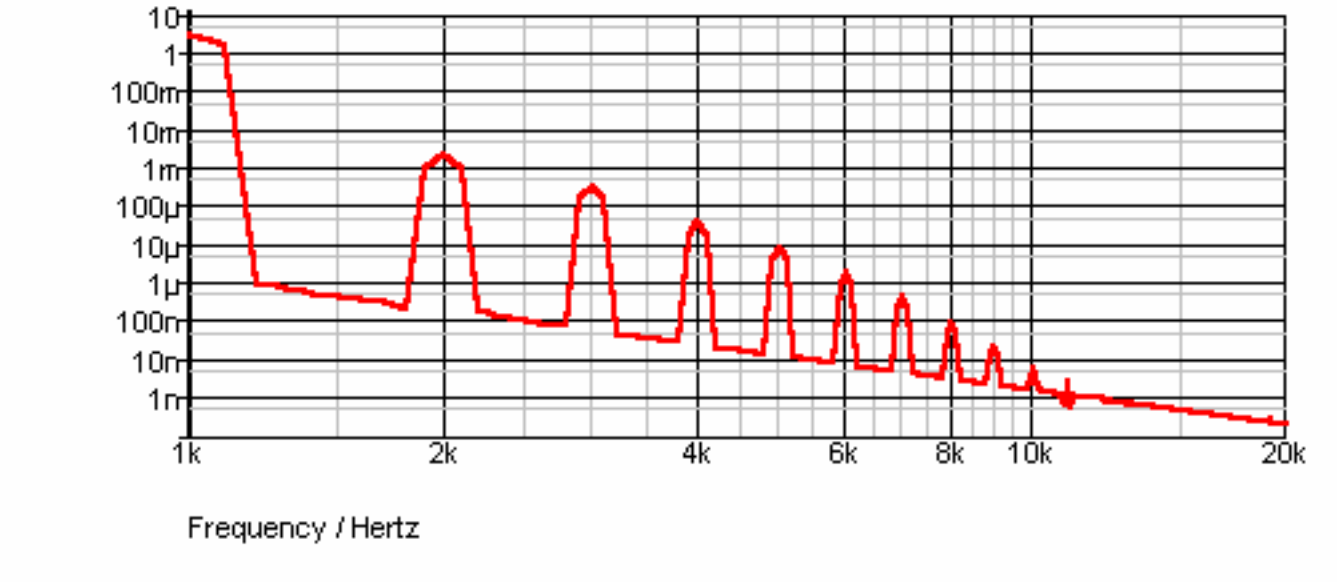
Reference:

1 Fletcher and Munson, Journal of the Acoustic Society of America - Vol.4, No. 2, 1933

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 1999-2003. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.


10
1
100m
10m
1m
100µ
10µ
1µ
100n
10n
1n
1k
2k
4k
6k
8k
10k
20k
Frequency / Hertz

TRUE AUDIO®
Updated: Wednesday, 18 January 2006
Welcome to the Loudspeaker Design Center!

Home
Catalog
Tech Topics
Audio Links

NEW v 3! TrueRTA™ Real Time Analyzers from $39.95! LOUDSPEAKER Design Software only $79!

Tech Topics
Technical Discussions of Audio and Loudspeaker Issues

Visit the Audio Engineering Book Store

Topic No. 2

Loudspeaker Diffraction Loss and Compensation

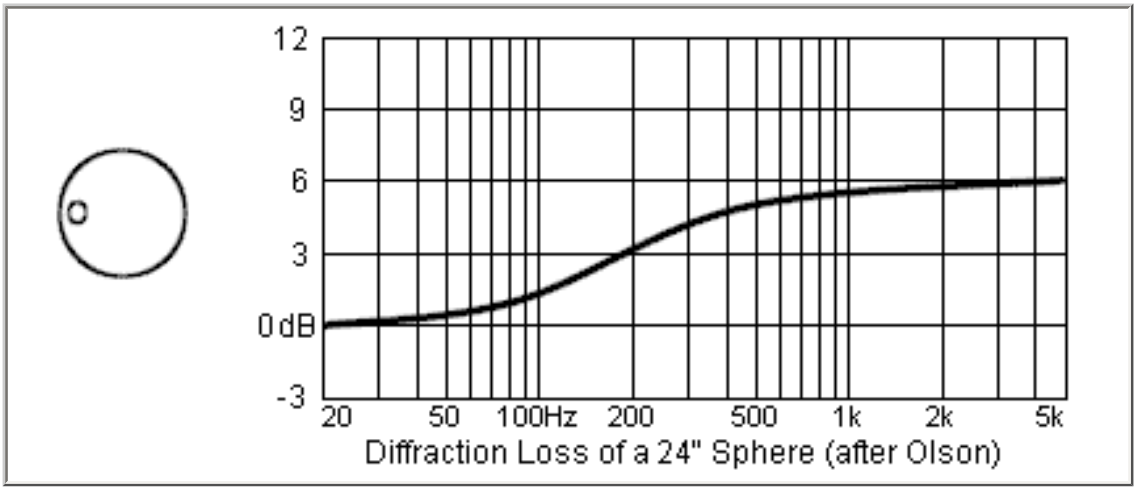
by
John L. Murphy
Physicist/Audio Engineer

Loudspeaker enclosure "diffraction loss" occurs in the low frequency range of loudspeakers in enclosures that are located in the open, away from walls or other surfaces. The essence of it is this: At high frequencies the speaker is radiating into "half space" i.e. it is only radiating into the forward hemisphere. No significant energy is radiated to the rear of the speaker. At low frequencies the speaker is radiating into both the forward hemisphere and the rear hemisphere. That is, at low frequencies the speaker radiates into "full space". Because the "energy density" at low frequencies is reduced there is a loss of bass. In short, speaker systems designed for radiation into half space (mounted flush on an infinite plane) exhibit a loss of bass when implemented in typical speaker enclosures. Fortunately, this bass loss can be accurately modeled and subsequently compensated.

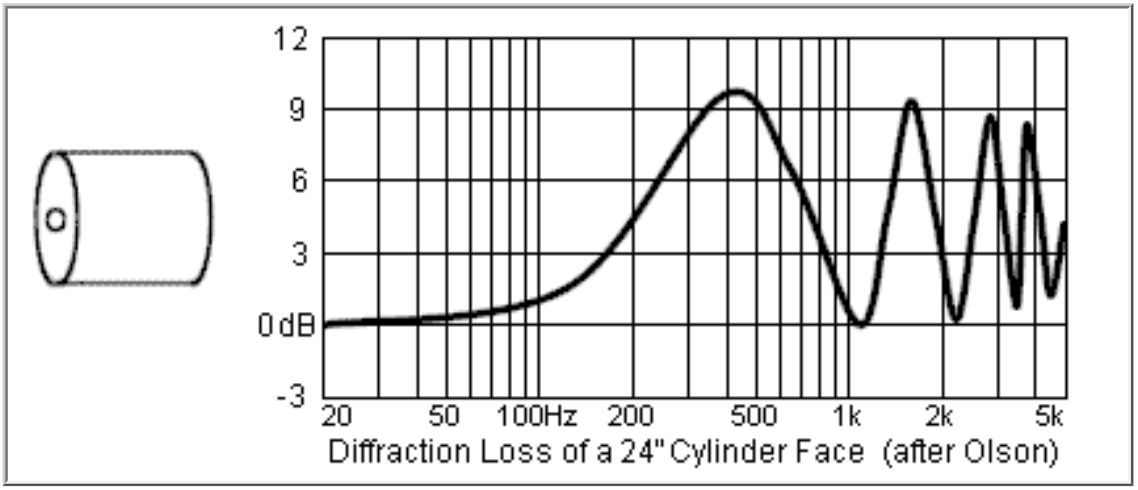
Most loudspeaker modeling is performed based on the assumption of radiation into half space. A speaker radiating into half space plays 6 dB louder than the same speaker radiating into full space. This is the crux of the diffraction loss. A full range speaker finds itself radiating into half space at the upper frequencies but radiating into full space at lower frequencies. As a result, there is a gradual shift of -6dB from the highs to the lows. This is what is called the "6 dB baffle step" or the enclosure's "diffraction loss". The center frequency of the transition is dependent on the dimensions of the baffle. The smaller the baffle the higher the transition frequency.

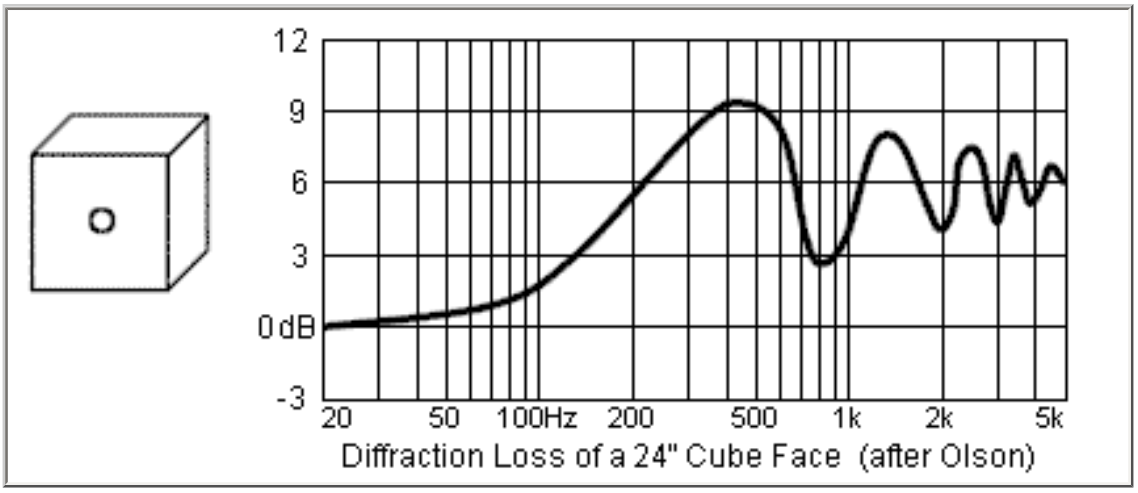
The shape of the diffraction loss frequency response curve depends on the size and shape of the enclosure. Olson has carefully documented the diffraction loss of enclosures of various shapes (see

references below). All enclosure shapes exhibit a basic 6 dB transition (or "step") in the response with the bass ending up 6 dB below the treble. A spherical enclosure exhibits this transition clearly with a very smooth diffraction loss curve. In the curves below I have taken the liberty of extending the frequency range of Olson's original graphs from 100 Hz to 20 Hz at the low end and from 4kHz to 5 kHz at the high end. The low frequency response was extended to more clearly reveal the "stepped" nature of the response. I wanted it to be clear that the response levels off at the low end. Olson's own reproductions of the measured diffraction loss of a sphere by Muller, Black, and Davis tend to confirm that my extensions to the responses are correct.

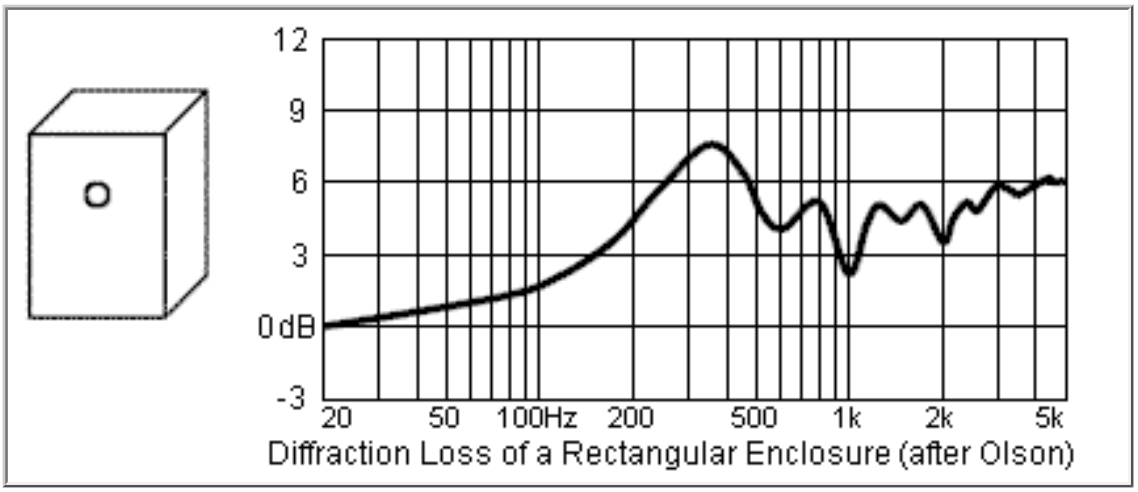


More "angular" enclosures exhibit the underlying 6 dB step along with a series of response ripples that are dependent on the placement of the speaker with respect to the baffle edges. The worst case appears to be placing the driver at the center of a circular baffle so that it is the same distance from all diffracting edges.



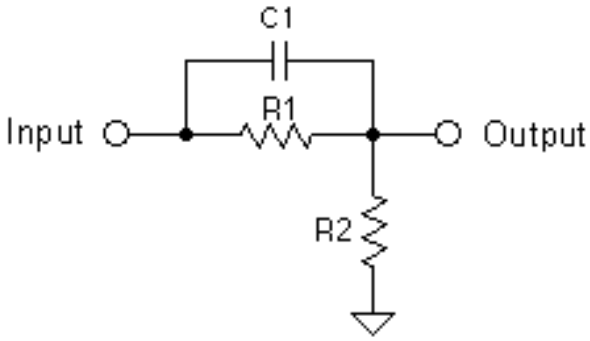


Placing the driver on the baffle so that it is a different distance from each edge tends to minimize the response ripples and make the diffraction loss look more like the smooth loss of the sphere. Olson's rectangular enclosure is an improvement over the cube and the cylinder face but the driver is still equidistant from three edges. Other authors report further reduction in the ripples with careful driver placement and edge rounding.



Because the spherical diffraction loss is a common element for the diffraction of all enclosures and the response ripples are much more difficult to predict (and can be minimized anyway) it makes sense to approximate the diffraction loss of a loudspeaker as the diffraction loss of the equivalent sphere.

One simple electrical circuit which produces a 6 dB step reduction in the bass response is shown below.

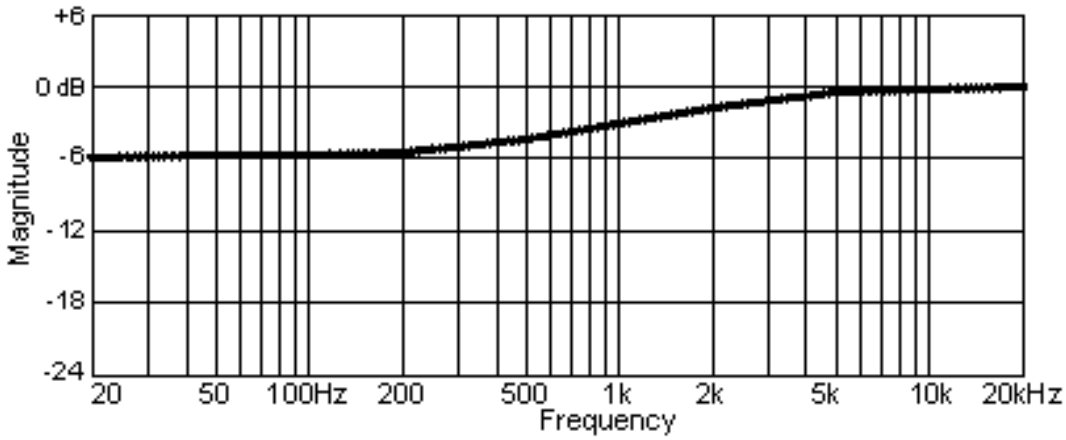


If we let R1 = R2 = R then a 6 dB attenuation results at low frequencies. At higher frequencies (where C1 becomes a low impedance) the attenuator is effectively bypassed and the signal is passed without attenuation.

It can be shown that the 3 dB "center" frequency for the above network is given by:

$$f_3 = \frac{\sqrt{2}}{2\pi RC}$$

The frequency response of diffraction modeling network typically looks like this:



Careful inspection of Olson's spherical diffraction loss curve reveals a -3dB frequency of about 190 Hz for the 24" sphere. Assuming that the 3 dB frequency is inversely proportional to the baffle diameter I have arrived at the following approximation for calculating the -3dB frequency as a function of baffle diameter.

f(3) = 115/W(B)
(where W(B) is the baffle width in **meters**)

or

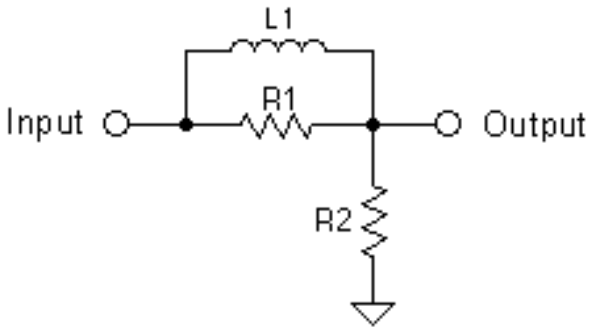
f(3) = 380/W(B)

(where W(B) is the baffle width in **feet**)

Sanity Check: for Olson's 24" (2 feet) baffle we calculate f(3) = 380/2 = 190 Hz . . .**OK!**

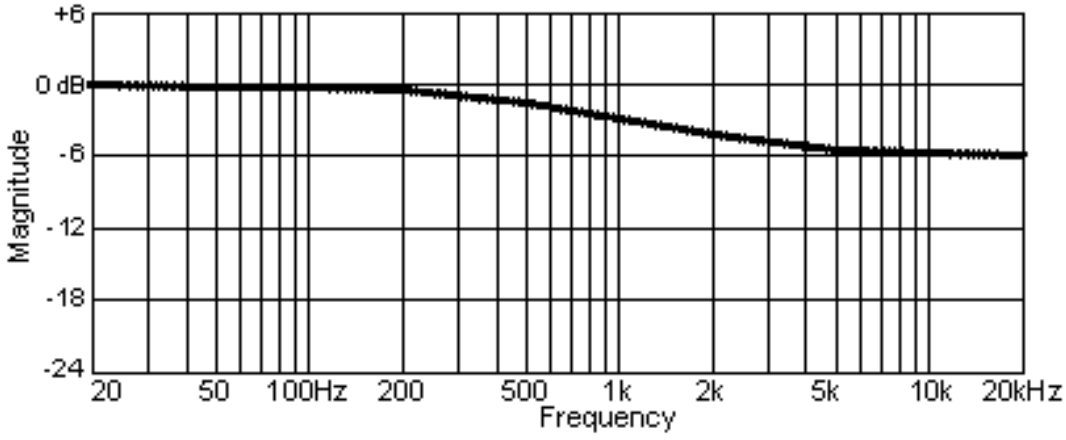
Once the diffraction loss is known it is possible to design a simple electrical network that will exactly mirror the spherical diffraction loss and restore the lost bass to a speaker system. Loudspeaker designers have traditionally compensated for the diffraction loss by reducing the level of the tweeter and making other adjustments in the crossover. The method I propose is to design for half space but then do a precise mirror image compensation for the diffraction loss by way of an R-L network wired in series with the (impedance compensated) speaker. Alternately, the diffraction loss can be compensated at line level with a simple R-R-C network. Line level correction would reduce the requirement for the large inductors typically needed for a speaker level compensation network.

A simple electrical network which produces a 6 dB step reduction in the treble response is shown below.



Here R2 represents the loudspeaker load impedance. If we let R1 = R2 = R then a 6 dB attenuation results at high frequencies. At lower frequencies (where L1 becomes a low impedance) the attenuator is effectively bypassed and the signal is passed to the driver (R2 here) without attenuation.

The frequency response of 6 dB diffraction compensation network looks like this:



To design an RL network which will compensate for diffraction loss of a particular system we start by setting:

R1 = R2 = R =Nominal System Impedance (8 Ohms for example)

Next, you can use my empirically derived equation to calculate the value of the inductor L1:

$$L1 = \frac{W_B R}{1.021}$$

Where W(B) is the baffle width in meters, R is in Ohms, and L1 in millihenrys.

I arrived at this equation for L1 by forcing the 3 dB frequency of the compensating network to match the 3 dB frequency for the diffraction loss of the baffle.

The resulting RL network should be wired in series with the speaker system it is compensating. The correction will be most accurate if the loudspeaker itself approximates a resistive load.

Example:

Find the network required to compensate the spherical diffraction loss of a 4 Ohm speaker system with a 0.25 meter wide baffle.

R = 4 (the nominal impedance of the system)
(this resistor should have a power rating something like a quarter of the **system** power rating)

L1 = .25 x 4 / 1.021 = 1 / 1.021

L1 = .979 mH
(1 mH will be close enough)

To **build** this network, start by connecting a 4 Ohm resistor in parallel with a 1 mH inductor. Then connect this RL network in series with the speaker. You should hear reduced treble when the diffraction loss compensator is used.

WinSpeakerz models the diffraction loss of the enclosure as a simple spherical diffraction loss. Provided the driver is located "irregularly" on the baffle this gives very good approximation to the actual diffraction loss of the enclosure. The frequency of the transition is controlled by the "Baffle Width" parameter at the System Editor page 1. The response of the speaker can be viewed with or without the diffraction loss and the diffraction loss can also be viewed separately.

jlm

25Jun99 A follow-up post on this topic:

Question:

> I've been researching the idea of adding a baffle step compensation circuit
> to my Marchand XM9 active crossover. The only thing that seems hard to do
> is calculate the amount of boost actually needed as this is room dependent.
> The theory, see John Murphy's article at www.trueaudio.com, suggests 6db
> boost. An article at the TL web page suggests 3db is more likely with room
> reinforcement. Has anyone looked at this.

The 6 dB loss is correct for a speaker enclosure in free space. When the enclosure is placed in a room it will encounter various effects due to the room (reverb, standing waves, boundary effect, cavity effect . . .)

Diffraction loss and room effects are independent and completely different effects. The diffraction loss is nicely predictable whereas the effects of the room are highly variable, not only from room to room but also with speaker placement and room furnishing. This typically means that each listening environment will be unique and will require unique compensation.

I suggest the 6 dB diffraction loss correction as a correction for the diffraction loss alone. I don't suggest that it will neutralize all the effects due to a unique listening environment. Others suggest you "deal with diffraction" in the crossover, usually by just lowering the tweeter level a bit.

In some situations 3 or 4 dB of diffraction loss correction may result in an overall response that is closer to neutral (flat). But the most correct way to compensate the room would be to do it separately from any diffraction loss correction. Room compensation might take the form of several notch filters tuned to the worst peaks resulting from room modes. Next you might want to tilt the treble up a smidge to compensate for reverberation that has significant treble loss. Dark room reverb will make the playback sound a little darker. Bright reverb . . . bright. Next, depending on the size of the room and speaker response, you might need to compensate for the cavity effect. In larger rooms cavity effect can be ignored but in vehicle cabins it is a major effect.

Diffraction loss compensation is only part of the job of precisely compensating for the difference between a theoretical half-space acoustic load and what happens when we place an enclosure in a real world listening room. Reducing the degree of diffraction loss compensation MAY reduce the coloration from the room effects as these effects largely tend to "boost the bass" but such an adjustment is imprecise at best.

If we can systematically identify each source of color between our half space model and our particular listening room then we can then take steps to precisely neutralize the response in our own listening room. Spherical diffraction compensation is one effect we can correct with a high degree of precision. As we move toward a better understanding and modeling of our listening rooms I'm sure we will work out more practical and precise ways to compensate our rooms.

Comments and critique are welcome. :-)

Regards,

John

////////////////////////////////////
John L. Murphy
Physicist/Audio Engineer
True Audio
http://www.trueaudio.com
Check out my new book "Introduction to Loudspeaker Design" at Amazon.com

Loudspeaker Diffraction Technical References

Revised: 20Jun00

Wright, J. R.
"Fundamentals of Diffraction"
Journal of the Audio Engineering Society
Vol. 45, No. 5, 1997 May, pp.347-356

> **NOTE:** The above reference was published after this tech topic was completed and thus was not included in the original list of references. I would suggest that this be the **first** reference studied by anyone beginning an investigation into loudspeaker diffraction phenomena. In particular, note that Wright questions the validity of certain details in some of these previous papers, especially regarding the phase change of the diffracted wave. Some of the previous papers incorrectly indicated that there was no phase difference between incident and diffracted waves in the illuminated zone. jlm 20Jun00

Rasmussen, Soren and Rasmussen, Karsten Bo
"On Loudspeaker Cabinet Diffraction"

Journal of the Audio Engineering Society
Vol. 42, No. 3, 1994 March, pp.147-150
(for comments on the above article see: Verification of Vanderkooy's "Simple Theory".
JAES, Vol. 42, No. 10, 1994 October, p.826)

Vanderkooy, John
"A Simple Theory of Cabinet Edge Diffraction"
Journal of the Audio Engineering Society A modification of GTD.
Vol. 39, No. 12, 1991 December, pp.923-933

Gonzalez, Ralph E.
"A Dual-Baffle Loudspeaker Enclosure for Balanced Reverberant Response"
Audio Engineering Society Preprint No. 3203
Presented at the 91st AES Convention, Oct 1991

Porter, James, and Geddes, Earl
"Loudspeaker Cabinet Edge Diffraction"
Journal of the Audio Engineering Society
Vol. 37, No. 11, 1989 November, pp.908-918

Backman, Juha
"Computation of Diffraction for Loudspeaker Enclosures"
Journal of the Audio Engineering Society
Vol. 37, No. 5, 1989 May, pp.353-362

Bews, R. M., and Hawksford, M. J.
"Application of Geometric Theory of Diffraction (GTD) to
Diffraction at the Edges of Loudspeaker Baffles"
Journal of the Audio Engineering Society
Vol. 34, No. 10, 1986 October, pp.771-779

Kral, Robert C.
"Diffraction - the True Story"
Speaker Builder Magazine, 1/80, pp.28-33

Olson, H. F.
"Direct Radiator Loudspeaker Enclosures"
Journal of the Audio Engineering Society
Vol. 17, No. 1, 1969 October, pp.22-29

Muller, G.G., Black, R., and Davis, T. E.
"The Diffraction Produced by Cylindrical and Cubical Obstacles
and by Circular and Square Plates"

Journal of the Acoustical Society of America
Vol. 10, No. 1 (July 1938), pp. 6-13

New v 3! TrueRTA™ Real Time Analyzers from $39.95! LOUDSPEAKER Design Software only $79!

TRUE AUDIO®

387 Duncan Lane
Andersonville TN
37705

Phone/Fax: 865-494-3388

True Audio Home Page | Catalog | Tech Topics | Audio Links | Book Store
WinSpeakerz Info | WinSpk Support | WinSpk Drivers | Try the WinSpeakerz Demo
MacSpeakerz Info | MacSpk Support | MacSpk Drivers | Try the MacSpeakerz Demo
TrueRTA Info | TrueRTA Support | TrueRTA Downloads
Introduction to Loudspeaker Design | About the Author | Box Chart | Privacy Page | Security Policy

Your comments are welcome at webmaster@trueaudio.com
Copyright©1990-2006 True Audio, Andersonville, TN USA, All rights reserved.

LINKWITZ LAB

Design & Consultation in Electronics & Acoustics

| **Page 1** | Page 2 | Page 3 | Page 4 |

Issues in loudspeaker design - 1

When it comes to the practical requirements for accurate sound reproduction, then there are a surprising number of issues that are not fully understood. Sound reproduction is about creating illusions in our mind, in that amazing information processor between our two ears. Sound reproduction involves physical processes to generate air vibrations that move our eardrums and in turn stimulate mental processes, that have evolved over millions of years for the purpose of survival and for communication, processes that lead to perception of the sound source and its environment. While we can measure many different physical parameters involved in the generation and propagation of a sound field, it becomes exceedingly difficult to assess their relative importance to what we hear and how it helps or detracts from the illusion. The published scientific material on the psychology of hearing (psychoacoustics) is extensive and sometimes helpful in explaining phenomena in two-channel and surround sound reproduction.

In the following I describe what I see as frontiers, if not in yours, then at least in my understanding of what matters for accurate sound reproduction. While I strive as an engineer to find answers and explanations based on the physics of the situation, I try to be honest to my listening observations as ultimate arbiter. I use my memory of un-amplified sound as reference for judging accuracy of the illusion. What I see as frontiers may be settled areas to others, or clear separations between opposing camps, that hold on to cherished convictions. I would hope that the exposé stimulates a few readers to research a subject further and share their insights.

HOME

Products

ORION - US

ORION - EU

ORION by the sea

Do-It-Yourself

A - Acoustic absorption and acoustic resistors

B - Frequency response of open baffles

C - Dipoles below the Schroeder frequency of a room

D - Room and baffle reflections

Concepts

Design of loudspeakers

Sound reproduction

Loudspeakers

Room acoustics

Crossovers

Issues in speaker design

Projects

ORION challenge

PLUTO omni speaker

THOR subwoofer

PHOENIX dipole speaker

Active filters

Microphone

FAQ

Reference earphones

E - Line sources

F - Group delay and transient response

G - Multitone distortion

H - Suggestions and hearing

I - Digital crossovers

J - Doppler distortion

K - Room response equalization

L - H-frame and U-frame open-baffle woofers

M - Investigation of energy storage

N - Mounting a driver to a baffle

O - Unreasonably high distortion in midrange drivers?

P - Estimate of $L_e(x)$ induced distortion at high frequencies

Q - Estimate of $L_e(i)$ induced distortion at high frequencies

R - Recommended study material about distortion in loudspeaker drivers

S - Dipole and monopole woofer response

A - Acoustic absorption and acoustic resistors

When a speaker driver is mounted in a box it radiates as much energy into the space in front of the cone as it does into the much smaller space behind the cone. What happens to the air borne energy inside? At long wavelengths it is common practice to store it in resonant

Surround sound

Loudspeaker evaluation

Resources

Website CD

Toneburst CD

Demo CD

Publications

Music

Links

Site map

About me

HOME

ORION Forum

structures to extend the steady-state low frequency response of the speaker. In general, the energy leads to very high sound pressures inside the box. A small amount of the energy is lost as heat in the stuffing material, some in the process of flexing the cabinet walls. Much of it reappears outside the box, because the thin cone presents a weak sound barrier. Just how much is difficult to measure, but it is a contributor to the frequency response. I am of the opinion that the effect is most notable in the low hundreds of Hz region, where stuffing materials are ineffective and the internal dimensions not small enough for the internal air volume to act as a pure compliance. Consequently, enclosures should be either very small (less than 1/16th of a wavelength) or extremely large, both of which are not very practical for different reasons.

To make progress with box speakers an acoustic resistor is needed that can more effectively dissipate energy in the 80 Hz to 800 Hz frequency range at high volume velocities. Such device would not only be useful for closed box speakers, but also for speakers that use the rear radiation from the driver to form a specific polar radiation pattern, such as a cardioid.
A cardioid speaker can be made with two opposite polarity monopole sources separated by a distance D, and with the signal to one of the sources delayed by a time T = D/c. An implementation of this concept could be a driver in a box of depth D where the rear wall is an acoustic resistor R. At long wavelengths the box internal air volume behaves as a compliance or acoustic capacitor C. The acoustic output from the rear of the box is low-passed by the RC filter and delayed relative to the front output by T = RC.

The acoustic resistor should be purely dissipative, with vanishing reactive component, and be independent of frequency. It also should be linear over the range of volume velocities encountered for high SPL. Traditionally cloth type materials have been used for cardioid speakers. Long fiber wool, synthetic fibers or fiber glass matting have been used to attenuate sound inside enclosures. The properties of these materials are neither frequency independent nor linear.

It may not be widely known that filter media for the filtration of liquids and gases in the chemical and other industries can have applications in acoustics. Such filters may be thin sheets (<1 mm thick) of a non-woven, sintered, stainless steel fibre matrix for filtration levels from 5 to 50 micron. Airflow at a constant velocity v through the filter material causes a pressure drop Δp between input and output sides corresponding to a flow resistance $R_f = \Delta p/v$ [Ns/m^3].

It is common in this industry to specify an inverse quantity which is Permeability P [l/dm^3/min] at 200 Pa pressure drop. Flow resistance and permeability are related by $R_f = 1200/P$ in this case.

Resistance values between 150 and 3500 Ns/m^3, or 15 to 350 rayl in the older cgs system of units (1 rayl = 10 Ns/m^3), are obtainable from a single filter sheet. For comparison the free-space acoustic field impedance $p/v = \rho c$ is resistive and has a value of 414 Ns/m^3 = 41.4 rayl. Materials are available with greater structural rigidity such as Feltmetal with thickness up to 6 mm and resistance between 6 and 50 rayl. The impedance is resistive and constant over the 20 Hz to 2 kHz range that I tested. Linearity should also be quite good, but I have not measured it.

The challenge remains to build an acoustic termination for the inside of a box. Feltmetal and filters should be readily useable for a cardioid speaker, but for a woofer application their linearity at high volume velocities needs investigation.
Top

B - Frequency response of open baffles

At this point in time the design of an open baffle speaker using cone type drivers is highly empirical. The baffle shape which determines the on-axis and off-axis frequency response must be optimized by experiment. This is how the dimensions of the PHOENIX main panel were determined (Ref.2), because no modeling software was available. Some guidance for baffle size can be derived from the response of a flat,

thin circular disk baffle of radius D with the driver in its center. This shape is less than optimum for use at frequencies above f = 0.5 c/D, where comb filtering occurs. The peaks and dips in response are smoothed by a flat rectangular baffle. As the driver itself becomes more directional the outline of the baffle has less effect.

For size, appearance and structural reasons the baffle may be folded back into a shallow U-shape. At low frequencies the H-shape and W-shape baffles lead to space savings. All of theses structures are considerably more difficult to analyze, because lumped and distributed acoustic elements have been added in the process of folding.

The radiation of the driver itself to front and rear is not symmetrical. The basket structure adds a complicated acoustic lowpass filter. This affects not only the output to the rear of an open baffle, but also the response in front due to additional phase shift and change in amplitude.

While it is not too difficult to estimate the response of H and W baffle structures at long wavelengths, it would be highly desirable to have modeling software for flat and U-shaped baffles for the whole frequency range. To be of usefulness the software would have to account for the measured frequency response of the driver in an 'infinite' size baffle, for radiation in front and rear and at different angles.
Top

C - Dipoles below the Schroeder frequency of a room

The Schroeder frequency describes the approximate border between the frequency range of discrete, sparse room mode distribution below it and high mode density above it.

$$f_s = 2000 \left(T_{60} / V \right)^{1/2} \quad [\text{Hz, s, m}^3]$$

For typical domestic listening rooms f_s is in the order of 100 - 200 Hz.

Above f_s a dipole source is in the reverberant field and has the following behavior:

- 1/3rd or -4.8 dB of the radiated acoustic power compared to a monopole for the same direct sound pressure level, due to the cos (angle) free-space polar pattern.
- A 4.8 dB higher level of direct to reverberant sound compared to a monopole.
- Rotation of the source affects the direct sound level, but this is strongly masked, because of the typically higher reverberant sound level.
- Room placement and rotation have little practical effect upon the reverberant sound field level. The number of possible modes is so high, that reducing the excitation of some increases it for others.
- Reflections are affected by placement and rotation.

Below f_s in the sparse mode region:

- Fewer modes are excited compared to a monopole, because of the cos(angle) directivity.
- A monopole and a dipole source in the same location will excite a given mode to a different degree, because their source impedances are very different. Consequently, the match to the acoustic field impedance of the mode is different.
- Monopole and dipole source in the same location will excite a given mode to a different degree for an additional reason. For maximum

excitation by a dipole source the direction of mode propagation has to coincide with the free-space direction of maximum source radiation.

The observed audible difference between monopole and dipole for bass below f_s is:

- Less boom and droning of specific notes with a dipole,
- A higher degree of articulation and resolution of complex musical bass lines,
- A sense of air, spaciousness, and very natural reproduction of acoustic bass,
- Less physical impact,
- Less absolute bottom foundation below the lowest room mode frequency.

To my knowledge these and similar subjective observations have not been explained satisfactorily by theory and only partially by measurement. In Ref.1 I have investigated dipole behavior on the basis of:

- Steady-state frequency response of the transfer function between source and listener for a simple 2-dimensional model and measurements in a very small office room and a living room,
- Cumulative spectral decay measurements,
- Energy-time curve of the low frequency impulse response,
- Modulation transfer function for low frequencies,
- Frequency response to shaped tonebursts,
- Frequency and time response to a multi-burst signal and its correlation to listening impression.

The best physical evidence for dipole benefits, that I have found, comes from multi-burst tests. A CD with such test signals is available.
The steady-state frequency response has been used for arguments pro and con dipole low frequency sources. While the transfer functions for dipole and monopole are clearly different I would not dare to make sound quality judgments based either on the data that I have seen or the ones I have taken myself. The exception to this is the very low frequency region, below the first mode, where the monopole response gains consistently from pressure buildup compared to a dipole.

I suspect that the perceived low frequency benefits of dipoles are related to the psychology of hearing and involve much more complex processes than a simple steady-state response analysis. Our mind is geared towards analyzing time, frequency and amplitude variant sound pressure between two ears and selectively draws information from it.

My hypothesis is that three effects contribute to the subjectively more accurate reproduction of bass:
1 - The open baffle, dipole woofer has a figure-of-eight radiation pattern and therefore excites fewer room modes.
2 - Its total radiated power is 4.8 dB less than that of a monopole woofer for the same on-axis SPL. Thus the strength of the excited modes is less.
3 - A 4.8 dB difference in SPL at low frequencies is quite significant, due to the bunching of the equal loudness contours at low frequencies, and corresponds to a 10 dB difference in loudness at 1 kHz.
Thus, bass reproduced by a dipole would be less masked by the room, since a dipole excites fewer modes, and to a lesser degree, and since the perceived difference between direct sound and room contribution is magnified by a psychoacoustic effect.
Top

D - Room and baffle reflections

Reflections are part of everyday life. If you have been inside an anechoic chamber and experienced the eerie silence and feeling of disorientation, then you know what I mean. In more common surroundings, every sound we make generates a set of reflections. If the sound is sustained, then its amplitude and phase at a given location are affected by the surroundings. This is what happens in a listening room. To quote:
"The acoustics of a room resemble either a linear filter or a hall of mirrors, depending upon the time scale of interest. As a filter, a room causes linear distortion in the amplitudes and phases of steady-state tones received by the listener's ears. As a hall of mirrors, a room appears to be a collection of spatially distributed reflecting surfaces that create multiple versions of each transient sound. The reflected transients ultimately reach the listener from diverse directions, almost none of which agree with the direction of the source. Thus, both the steady-state and the transient response of a room result in the listener receiving misleading cues about the location of a source of sound in the room." [1]

Modeling and understanding of how we manage to cope with this situation, and actually localize real and virtual sources surprisingly well in most instances, is subject of ongoing scientific investigation [2]. Inter-aural time delay (ITD), and intensity differences (IID), spectral envelope cues, and the "precedence effect", all play their part in recognizing the location of sources when room reflections are present. In addition to affecting localization: "The lagging sounds can contribute markedly to perceived timbre and to the sense of a room's acoustics. If the acoustical features of the lagging sounds are consistent with them being echoes of the leading sound, then the lagging sounds are perceptually suppressed, and are not heard as separate auditory events. On the other hand, if the lagging sounds cannot be plausibly interpreted as echoes, or if insufficient information is available to decide whether they are echoes or not, then they will tend to be heard as separate auditory events." [2]

Most of the research into audibility of reflections has been performed under anechoic conditions. Detection thresholds and phantom image shifts for a single reflection versus delay time and amplitude have been studied extensively. But, the results cannot necessarily be translated to binaural listening in reverberant spaces. Nevertheless, the prevailing school of thought amongst audiophiles holds all reflections to be detrimental. The prescription is to either absorb or diffuse reflections as much as possible. A room with only hard reflective surfaces and with no furnishings, having echoes and long reverberation time, is clearly not an ideal playback environment, but nor is an overstuffed room, full of absorbing materials and resembling an anechoic chamber. The ideal room may be a function of the type of speaker used.
The typical, conventional box speaker has a power response that decreases over 10 dB from low to high frequencies, assuming a flat on-axis response. It is omni-directional up to a few hundred Hz and becomes ever more forward radiating with increasing frequency. An open baffle, dipole speaker has at least 5 dB less power response variation between low and high frequencies. While box speakers seem to perform best in relatively dead rooms, I have found that dipole speakers show their qualities better in a rather live room.

In general, the more a room resembles an anechoic chamber, the less the off-axis radiation contributes to the sound at the listening place. The type of dipole speakers I have worked on, all try to duplicate the on-axis response for off-axis horizontal angles up to +/-60 degrees, albeit at decreasing levels and up to about 5 kHz. This has two effects. The level of the reflected sound is angle dependent and can be reduced by speaker orientation. The acoustic features of the reflected sound are consistent with the direct sound. Both of these help in suppressing a perception of timbre change and decrease a sense of the room's acoustics. Consequently, you hear more of the speaker and less of the room.

Much has been made about the wall behind an open baffle speaker, the front wall from the listeners perspective. The wall reflects the rearwards radiated sound which is 180 degrees out-of-phase with the forward radiated sound. The interference between the two leads to progressive cancellation at low frequencies and comb filtering at higher frequencies for steady-state sounds. A corresponding effect occurs with a box speaker, at least up to several hundred Hz, only the comb filter frequencies are different and the low frequencies are boosted. All of this is controlled by the distance between speaker and wall. If the speaker is at least 1 m (3 ft) in front of the wall, then the reflected sound

is delayed by about 6 ms, which is sufficient not to be perceptually summed with the direct sound. The cognitive process disassociates it from the speaker and attributes it to the room with a sense of space.
Panel type open baffle speakers can have specific problems with the back wall. The radiation impedance seen by the panel goes to extreme values at frequencies where the distance is a multiple of a quarter wavelength. The motion amplitude of the very low mass panel may change due to the change in air load, causing a change in frequency response. Also, a large panel area creates a lobing polar response at higher frequencies and the reflection can become colored in timbre, shift angle and cause unstable imaging.

I have observed that a certain amount of lateral reflection is necessary for creating a believable illusion of the space in which the recording took place or to give a satisfyingly rich musical experience [3,4,5]. Flush (soffit) mounted speakers strike me as producing a sound that is two-dimensional, analytical, but ultimately lifeless. Highly directional horn speakers create a similar negative impression. In the tweeter range, wide horizontal dispersion removes harshness and adds airy-ness in my experience. This only makes sense if the reflected sound intensity has been increased. It might also explain why ribbon tweeters, which can be very narrow sources leading to wide dispersion, are so highly regarded, despite their poor vertical polar response.

RT60 for my listening room is around 550 ms and pleasantly lively for conversation and socializing. I do not like a muffled cocktail lounge ambience. The two side walls are mostly double pane glass. The room is very long. A large brick fire place is 1.5 m (5 ft) behind the speakers. Imaging is very precise, given a good recording, with a smooth spread of the venue acoustics between the speakers and slightly beyond. The images are placed behind the plane of the speakers, never in front. The speakers disappear completely.

Sound diffraction at the edge of a box speaker baffle is claimed to degrade stereo imaging. I have not experienced this myself for speakers with typical flat baffles for drivers up to 8" (20 cm), nor have I seen scientific evidence. The low amplitude diffracted sound wave occurs so early after the direct sound (<500 us), that the precedence effect does not come into play and summing localization occurs [6]. What is significant, is that the "baffle step", the transition from 4π to 2π-space radiation, occurs at higher frequency for a narrower baffle. This keeps the power response more uniform over a wider frequency range and the spectral signature of any reflection more like the on-axis response. Large radius or wide chamfered edges smooth the high frequency response on-axis and off-axis.

The question remains: When is room treatment necessary and for what purpose?
Specular reflections [7] can be diffused with quadratic residue diffusor panels, but this will not change the reverberant sound energy in the room. Diffusors become very large for frequencies below 500 Hz. Reflections can also be absorbed. This too works best above 500 Hz. Treatment for lower frequencies becomes exceedingly more space consuming and below 100 Hz requires custom solutions.

[1] William M. Hartmann, "Auditory localization in rooms", Proc. AES 12th International Conf., Copenhagen, 1993, "The perception of reproduced sound", pp. 34-39
[2] Brian C.J. Moore, "Controversies and mysteries in spatial hearing", Proc. AES 16th International Conf., Rovaniemi, 1999, "Spatial sound reproduction", pp. 249-258
[3] Arthur H. Benade, "From instrument to ear in a room: Direct or via recording", JAES, vol. 33, no. 4, 1985 April
[4] David Moulton, "The creation of musical sounds for playback through loudspeakers", Proc. AES 8th International Conf., Washington, 1990, "The sound of audio", p. 167
[5] http://www.sawonline.com/wp_mix_article1.html
[6] Jens Blauert, "Spatial hearing. The psychophysics of human sound localization", MIT Press, 1997
[7] James A. S. Angus, "The effects of specular versus diffuse reflections on the frequency response at the listener", JAES, vol. 39, no. 3, 2001 March
Top

E - Line sources

As a line source I define a speaker with a very narrow and long radiating surface, which is used at frequencies above f = c/2L, where its length L is greater than 1/2 wavelengths. Often such line source consists of a vertical array of a large number of identical drivers on a narrow baffle. In other cases it might be a ribbon like, narrow and long conductor in a magnetic field, or a narrow and tall electrostatic source. Specifically, I am referring to lines that are at least 1.5 m (5ft) long.

Under free-field conditions such lines will have a wide and fairly uniform horizontal radiation pattern The vertical polar response will be very irregular and highly distance dependent, because of the finite length of the source [1]. In a room, though, the images of the source in floor and ceiling will make it appear as infinitely long, the more so, the less the gaps between the ends of the source and the reflecting floor and ceiling surfaces. Consequently, the line launches a cylindrical wave and illuminates the room very uniformly. The acoustics of the room, therefore, are of similar importance as for an omni-directional point source. At sufficient distance the direct SPL falls off with the square root of distance, which improves the direct-to-reverberant sound ratio over that of a point source.

It surprises me that in a reverberant environment, such as a typical living room, these radiators give little indication of poor vertical polar behavior, even when the line does not fully extend to floor and ceiling. Clearly, the reverberant field in the room fills in any perceptual notches in the polar response. Standing close to the line and moving with one's ears up and down, the acoustic image does not change and follows the ear over the active length of the line, almost to its ends. Imaging is very precise, when standing or seated in the near-field of the line, which is closer than the typical listening distance given by the apex of an equilateral triangle. At greater distances the images grow in height. This can be impressive on big orchestral pieces, as if looking through a large window onto the sound stage. Small acoustic sources, like a vocalist or a cello, though, become overblown in size. Line source speakers are at their best when listened to close-up like giant headphones. This preserves proper size imaging and gives the illusion of a large sound stage.

Vertical height of the sound stage adds to its realism. I find small 2-way speakers very unsatisfactory in this respect, even when they are correctly positioned at ear level height. Line sources provide this height, as do single tweeter symmetric driver layouts like LM-UM-T-UM-LM. In the latter case separate frequency ranges are assigned to each symmetrical set of drivers LM-LM and UM-UM, such that any free-space vertical lobing is minimized. This type of source suffers considerably less from the image size distortion of a uniform line.

An added advantage of tall sources is their relatively large radiating surface area, which means small excursions and, therefore, progressively lower non-linear distortion for a given maximum sound level. The low distortion aspect is even more important, than sound stage height, wide angle horizontal dispersion and uniform power response. Low distortion means increased dynamics and clarity, and the ability to play back at near live levels without listener strain. It brings the naturally occurring distortion mechanisms of the ear into play, which are necessary elements for creating an illusion of reality.
Sensitivity can be high. A line with 25 identical drivers connected in series-parallel to obtain the same combined impedance as that of a single driver, will have 14 dB higher SPL for the same voltage applied.
Stored energy is not necessarily low because of the many cabinet and driver chassis cavities.

I would encourage more investigation of line sources for use in very large venues, as for public address or entertainment speakers. For typical domestic environments they are unnecessarily large. Today, the same low distortion sound pressure levels can be obtained from far fewer drivers and without the image distortion.

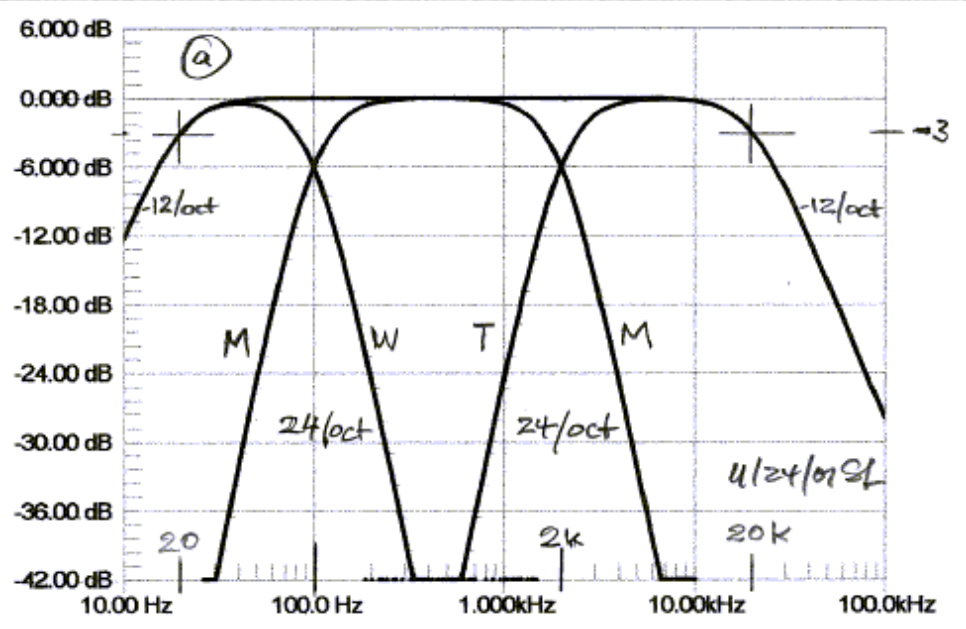
[1] Stanley P. Lipshitz, John Vanderkooy, "The acoustic radiation of line sources of finite length", 81st AES Convention, Los Angeles, 1986,

Preprint 2417
Top

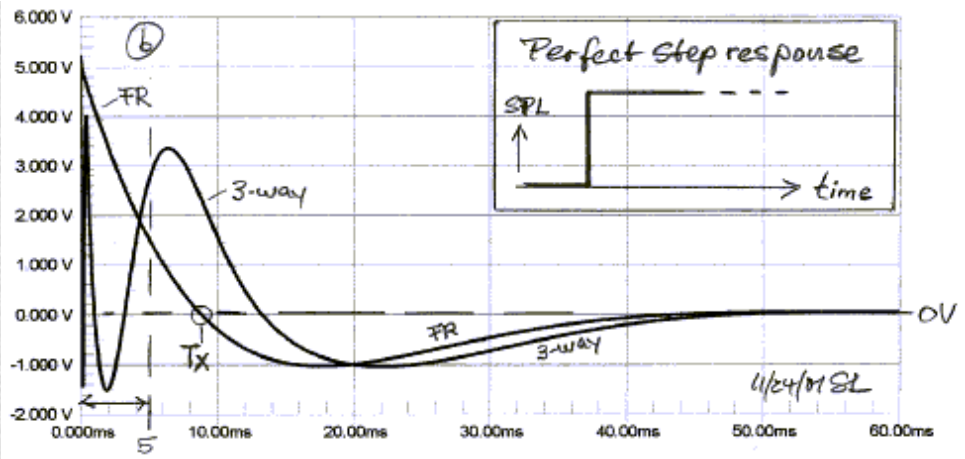
F - Group delay and transient response

Few studies have been made about the audibility of linear distortion of the time waveform for an acoustic signal. Some people have great expectations about the improvements that DSP technology can bring to loudspeakers and crossovers through linearizing their phase response, which should result in a perfect transient response. The step response fidelity is often taken as a measure of performance. The step response is the integral of the impulse response over time, and belongs to a frequency response, that has been rolled off at 6 dB/oct towards high frequencies, starting from zero frequency.

It is instructive to review the step response of a practical 3-way speaker and to compare it to a full-range speaker. Both speakers are assumed to have a flat frequency response that is 3 dB down at 20 Hz and 20 kHz, rolling off at 12 dB/oct like a filter with Q = 0.7. The 3-way speaker has LR4 crossovers at 100 Hz and 2 kHz as in graph (a).

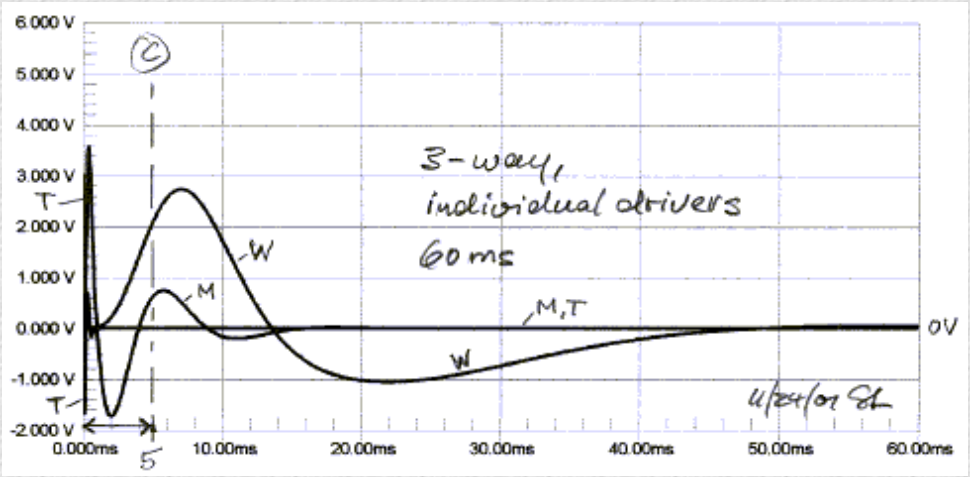


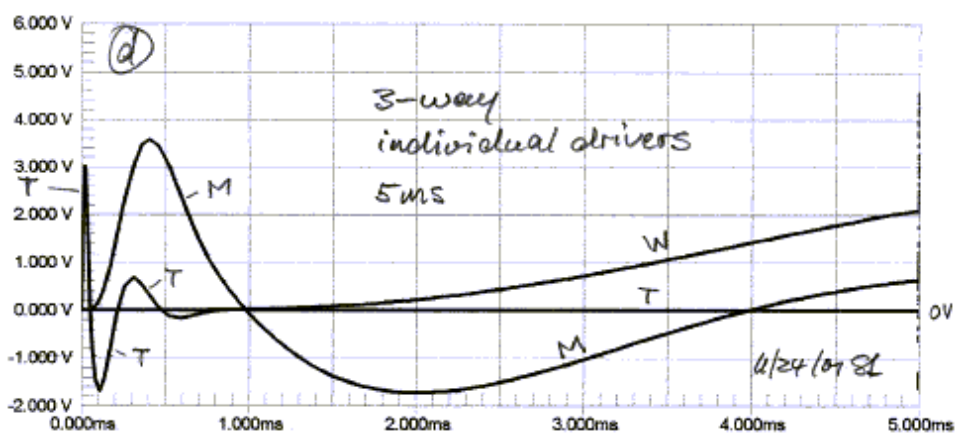
The step response of the full-range speaker FR is far from perfect, because the loudspeaker is a bandpass and in this view (b), over a 60 ms time window, the low frequency highpass dominates the shape of the curve. At time Tx the response crosses the zero axis. The positive area under the curve between zero and Tx equals the negative area between Tx and very large time, because a bandpass system cannot pass dc. Equivalently, a speaker cannot set up a permanent pressure increase, as would be required for perfect step response.



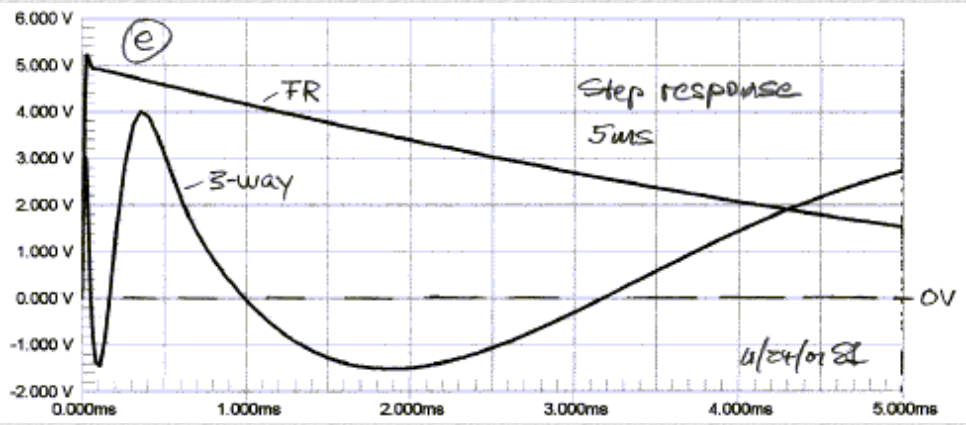
DSP based equalization can move Tx out in time, but how far is enough needs to be established. The aural integration time window is about 50 ms, before echoes are noted.

The step response of the 3-way looks even further removed from the ideal and it is easy to understand why people like to think this should be accompanied by audible defects. The response is the sum of woofer, midrange and tweeter outputs (c). The tweeter response, barely visible on the 60 ms time scale, really shows up in a display of the first 5 ms (d).



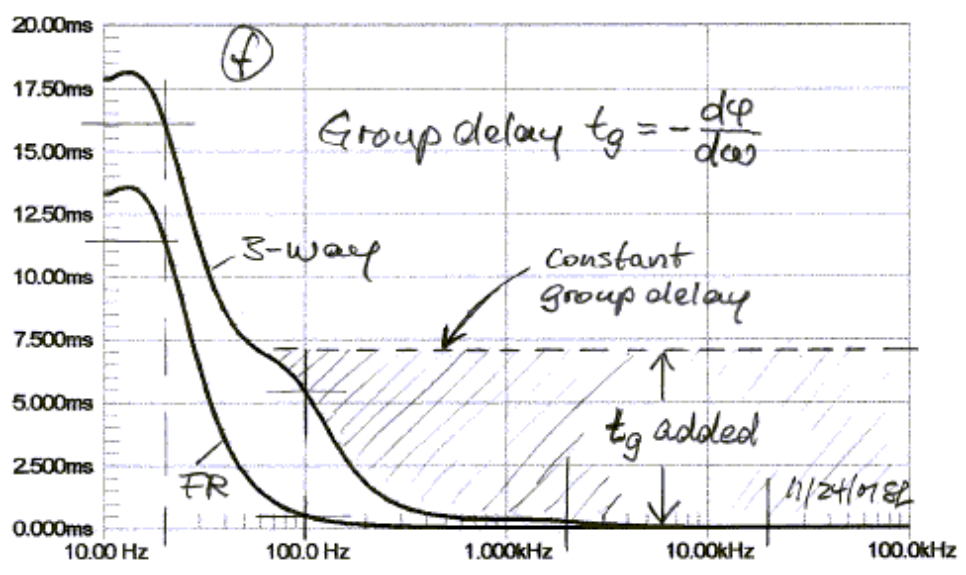


Note that the tweeter response has settled in 1 ms. The midrange has its second zero crossing at 4 ms. The woofer has barely started to move after 2 ms. A time window of 5 ms or less is frequently used in speaker reviews to comment about driver polarities and waveform preservation (e). This covers less than 10% of the transient response time for many speakers.



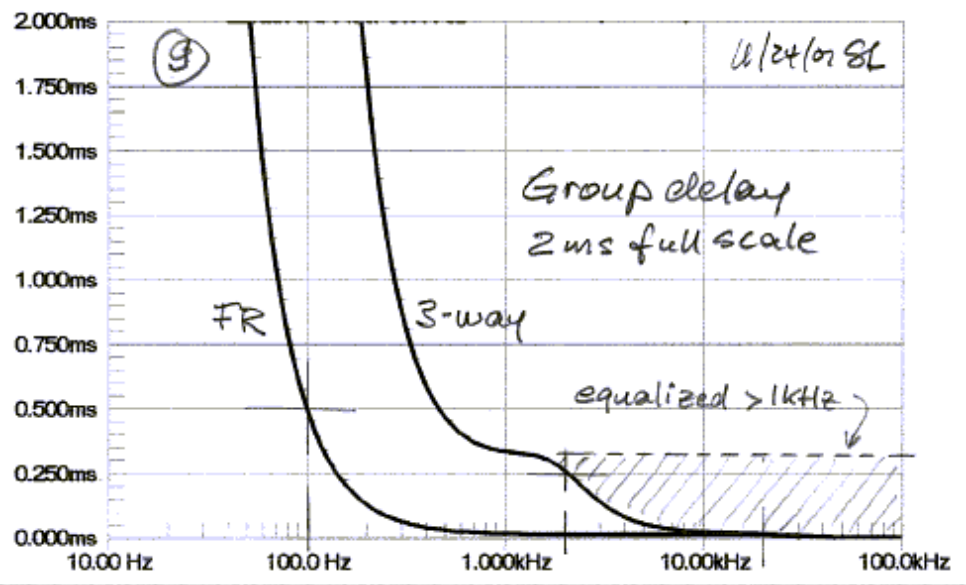
A recent audiophile magazine review of a 3-way speaker, with response similar to (e) above claimed: "... The step response indicates that the tweeter and midrange unit are connected with positive acoustic polarity (the sharp up/down spike at 0 ms and the lazier triangle 0.4 ms later), while the woofer is connected with negative polarity (the broad, negative-going hump centered between 1.5 and 2.5 ms). The actual drive-unit polarity doesn't matter except when it comes to waveform preservation -- and the jury is out on that subject."
More understanding of the subject matter is obviously needed.

The phase of a speaker's frequency response has to change linearly with frequency in order to preserve waveform fidelity. This is equivalent to saying, the group delay has to be constant (f).



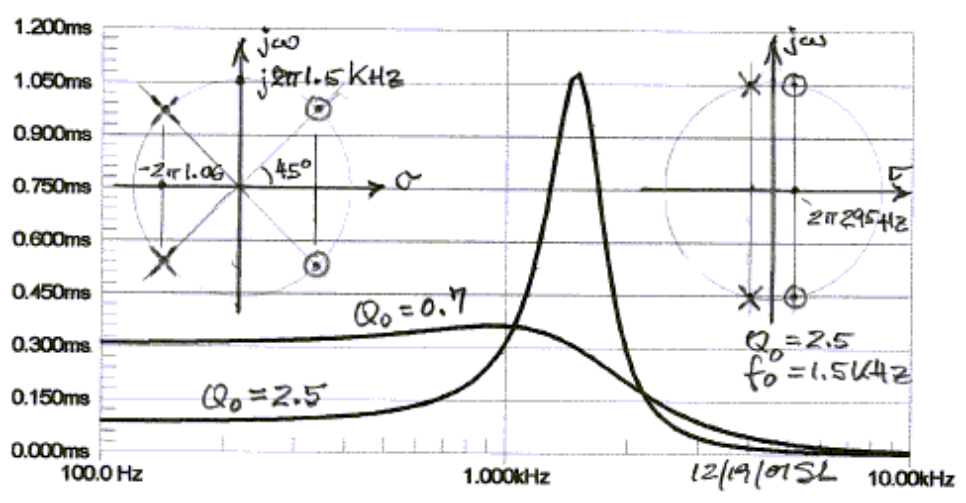
Group delay increases dramatically at low frequencies, even for the full-range speaker. To equalize the 3-way speaker down to 50 Hz a great amount of delay would have to be added to the higher frequency portion of the spectrum. This is practically impossible to accomplish with analog means and not completely trivial with DSP either.

If delay equalization is restricted to higher frequencies, where digital processing becomes easier because of shorter time record lengths, then large portions of the spectrum remain distorted (g). Such solution may improve the first few milliseconds of the step response, but it remains to be shown unambiguously that if this is audibly noticeable. It will enable fairly accurate reproduction of square-waves with frequencies above 1 kHz though.

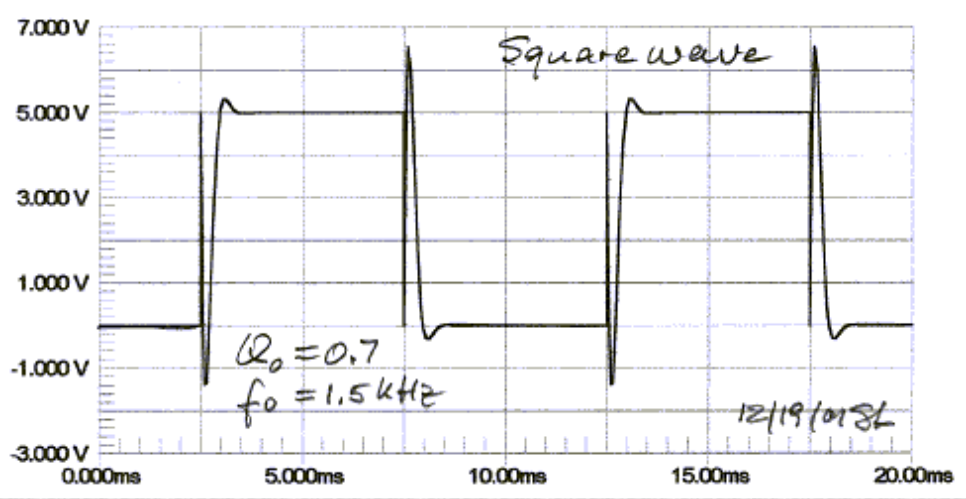


With similar effort a 2-way speaker could be delay equalized over a wider frequency range. A 2-way loudspeaker is either marginal in frequency coverage or in dynamic range and improvements in its transient response might be easily masked by other shortcomings. A full-range or single driver speaker is even more limited and not a contender, except in the form of headphones.

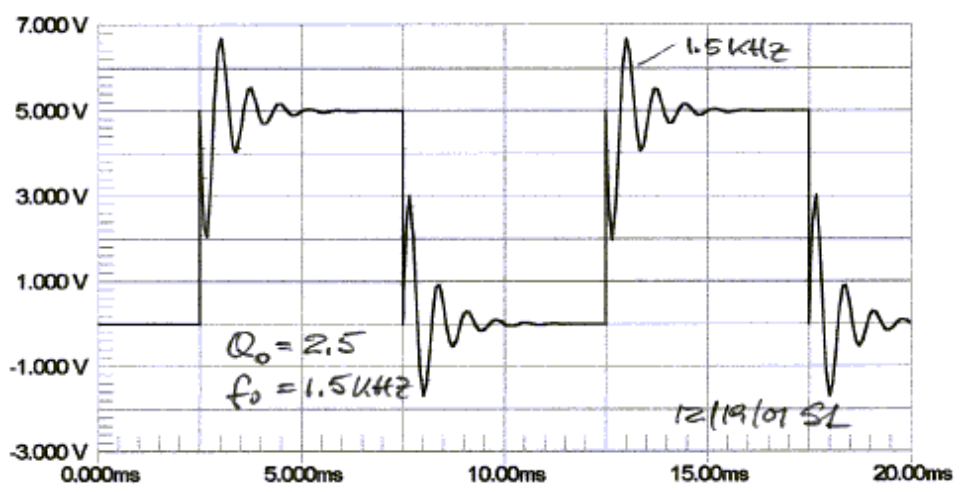
I have used headphones to convince myself that the phase distortion introduced by LR4 crossovers around 2 kHz and at 100 Hz are not audible. For example, a 1.5 kHz LR4 crossover with 24 dB/oct slopes forms a second order allpass filter with Q0 = 0.7 for its pole and zero locations.



Its non-linear phase shift and non-constant group delay in the crossover region cause waveform distortion which is not audible, though.



If the allpass Q0 is significantly higher, like Q0 = 2.5, then the phase distortion does become audible. In the case of the 100 Hz square-wave a high pitched sound is added. The frequency response, though, is completely flat.

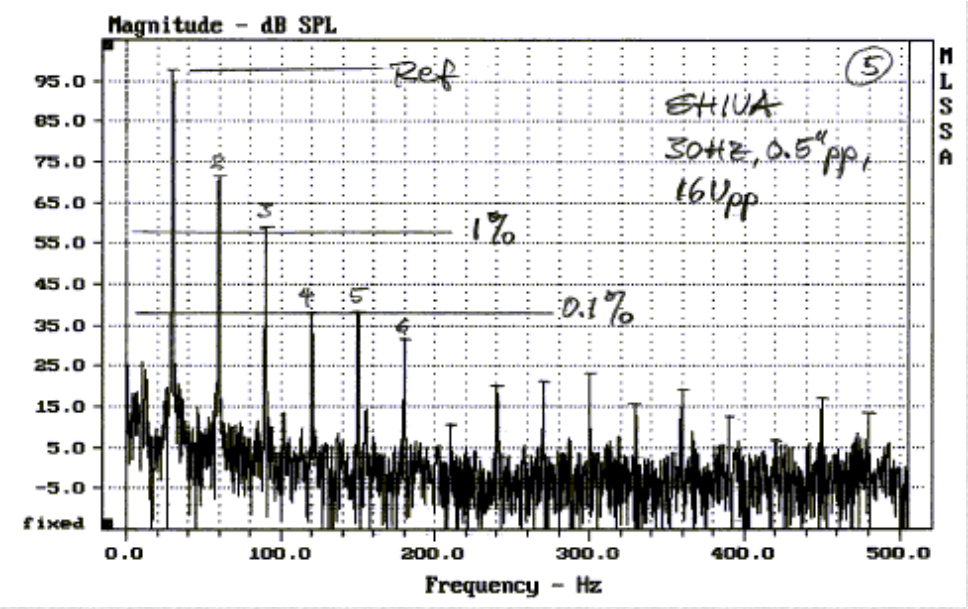


I am not certain what happens in the range below 100 Hz and I have strong suspicions that this is the region where delay distortion is audible. It is also the region where delay really accumulates via vented and bandpass woofers, and the great numbers of dc blocking capacitors in the signal chain from microphone to speaker terminal.

Further experiments might be performed with a high quality 3-way speaker system, whose phase response or group delay have been carefully measured or modeled. Knowing the phase distortion, which the speaker will introduce, use digital signal processing with the appropriate software to pre-distort various music files on a PC, such that the overall playback phase response will be linearized. Burn the newly created files to CD-R. For comparison take the same music files and add a constant delay to them, so that pre-distorted and un-distorted material have undergone the same process and any artifacts are common to both. In this way one could switch directly between nearly identical tracks of a CD-R and listen for differences that can only be the result of phase response differences.
Top

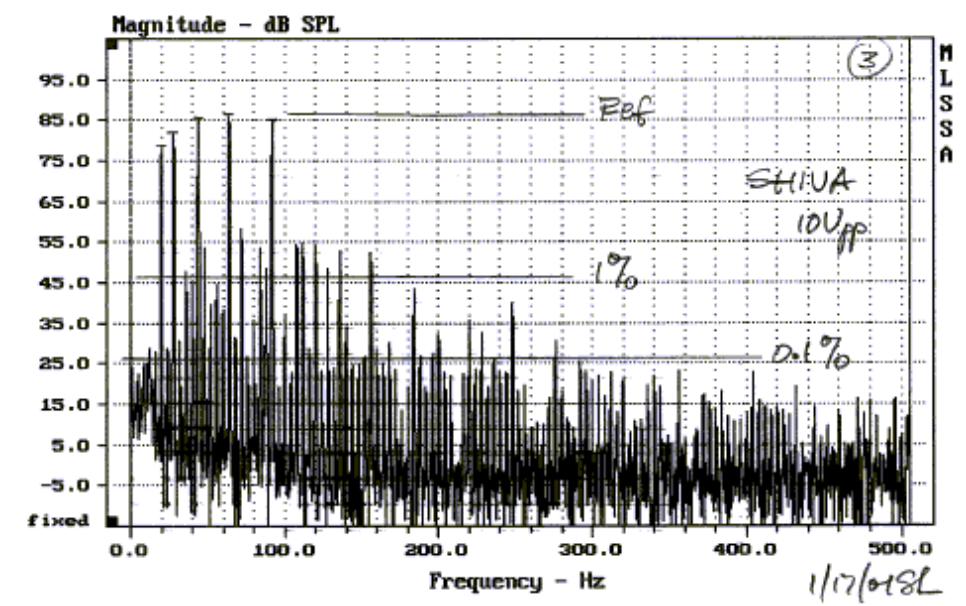
G - Multitone distortion

The measurement of harmonic distortion gives an incomplete description of the non-linear behavior of a loudspeaker, even when the exciting sinewave is swept to generate a continuous display of distortion products. Still, it is very useful data for comparing devices.

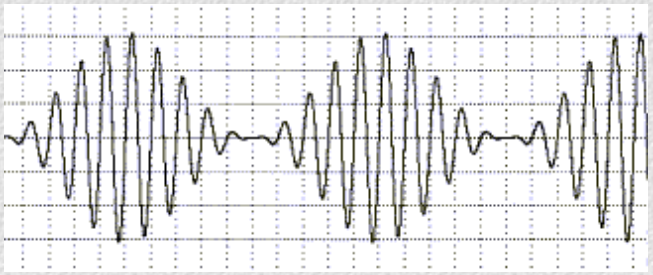


More information is obtained when two tones are used, because now also the various sum and difference products of the two stimuli and their harmonics can be measured with a spectrum analyzer. These intermodulation products are usually higher in amplitude than the related harmonic distortion products. I cannot emphasize enough, that anytime when a device introduces harmonic distortion, it will generate intermodulation distortion, when more than one tone is involved. Some seem to think that a little bit of 2nd harmonic distortion, as often generated by tube equipment, has a pleasing effect, because it enriches the natural even harmonics of acoustic instruments. It will also generate non-harmonic intermodulation, and has little to do with accurate sound reproduction or even generating the illusion of a real sound. Instead it imparts a euphonic sameness to all sounds.

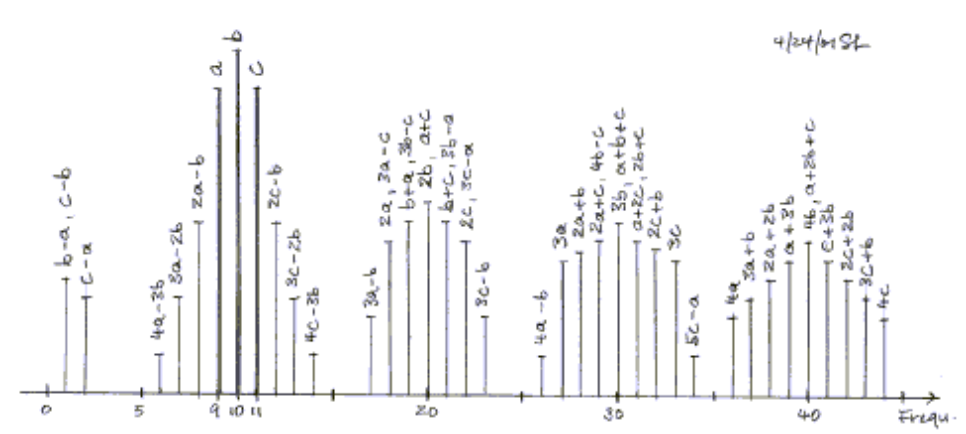
Real sounds contain many tones. A woofer might be tested with a set of 5 tones that are not harmonically related, such as 20, 28, 44, 64, and 92 Hz, but which cover the woofer's intended operating range. Each of the tones is a multiple of 4 Hz, with factors 5, 7, 11, 16, 23, and the resulting distortion products will occur at 4 Hz intervals.



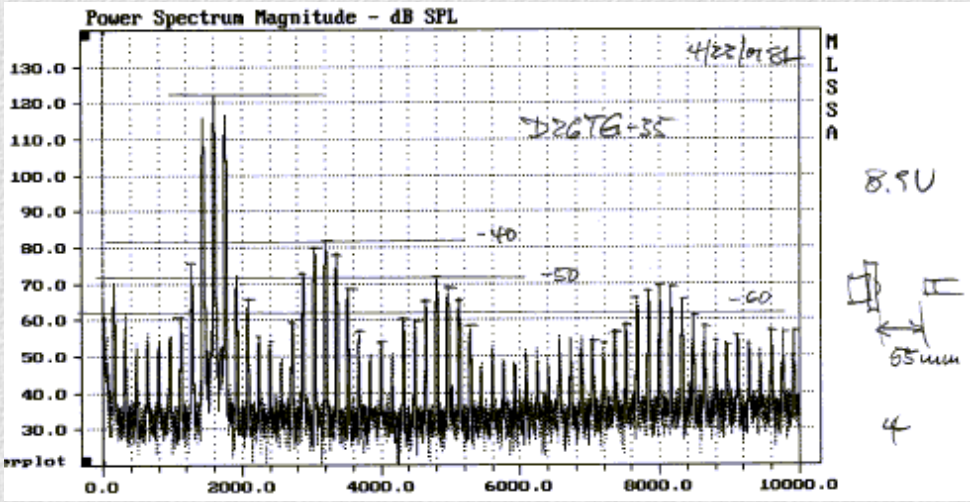
The richer distortion spectrum makes comparison between devices easier than relying on the few data points that harmonic distortion provides. The time waveform for the sum of the 5 tones looks very erratic but is, of course, periodic. A very simple looking, but highly effective test signal is a sinewave that is 100% amplitude modulated at 1/10th its frequency. This type of stimulus is especially useful for stressing vibrating systems and audibly detecting the onset of distortion as its amplitude is increased. To avoid permanent damage only bursts of this signal should be used.



The undistorted spectrum consists of the carrier b and two sidebands a, c, which are 6 dB down and offset by b/10. The potential distortion spectrum contains harmonic and intermodulation products. A few of those are shown below. They occur at intervals b/10.



The close-up measurement of a tweeter reveals the dense sound spectrum generated by the cluster of 3 tones at 1.6 kHz.



I have found these and other types of multitone test signals useful for sorting and relative ranking of drivers. Their subjective evaluation when part of a complete speaker correlates closely with the measurements. It should be understood that distortion has audible effects long before it is perceived as distortion. For example, it might enhance sonic detail, analogous to contrast enhancement in a photo.

I think a recent experience while attending a San Francisco Symphony performance gave me further insight into why we are able to tell a loudspeaker from a live instrument. Loudspeakers reproduced pre-recorded material, interspersed with the musicians' playing, during the performance of several modern pieces of music. It seemed, that coming out of the loudspeaker a sound carried a comet like trail of newly created sounds with it, that formed a continuum of background sounds, which was completely absent from the live instruments and made the speaker immediately recognizable as such. Live instruments had a space between tones, like a black background, and even en mass always remained articulate. I believe the observation had nothing to do with recorded ambience, which was present occasionally, or with hall reverberance. It reminded me of other incidences, where I had easily recognized live instruments, though there was no reflection free path between the source and my ears, and where I then struggled to describe to myself what would have been different, if a loudspeaker was playing instead. It seems to me that the ongoing-ness of sound is one of the major problems with speakers. It shows less with recordings of a

single voice with small accompaniment, so often favored by audiophiles, but in complex passages of classical orchestral music and choral works.

Much work remains to be done before firm predictions of sound character can be made from distortion measurements. Wolfgang Klippel has developed a practical methodology for characterizing and auralizing the effects of various distortion parameters in woofers at large excursions. Researchers at Cerwin Vega have done extensive theoretical work [1]. They concluded amongst other things: "... With an increase in the number of tones, the multitone provides increasingly detailed information about the nonlinear behavior of a DUT. An ultimate increase in the number of tones leads to a noise-like signal that may hypothetically provide complete information about a nonlinearity-impaired DUT, but demands an enormous amount of calculation, if a high-order nonlinearity is to be assessed."

[1] Gene Czerwinski, Alexander Voishvillo, Sergei Alexandrov, Alexander Terekhov, "Multitone testing of sound system components - Some results and conclusions, Part 1: History and theory", JAES, Vol. 49, No. 11, 2001 November, pp. 1011-1048.
"Part 2: Modeling and Application", JAES, Vol. 49, No. 12, 2001 December, pp. 1181-1192.
Top

H - Suggestions and hearing

It is common practice in promoting audio products to tell a customer what to pay attention to in what he is going to hear, then to follow that with a demonstration, and afterwards to tell him again what he heard. It is the same process, whether it is a manufacturer's representative speaking with a store sales person, or the sales person talking to a prospective buyer. It is almost impossible for the target person not to hear, what has been suggested they will hear.

The power of suggestion has been used to perform major surgery without anesthesia, where the patient experiences no pain. It has been shown with CAT scans of the brain that in such cases an information disconnect occurs between the region that responds to pain, and the region of the brain that gives the patient the perception of pain. Thus, the normal perception process has been altered through the verbal, non-physical preparation that the patient received beforehand.

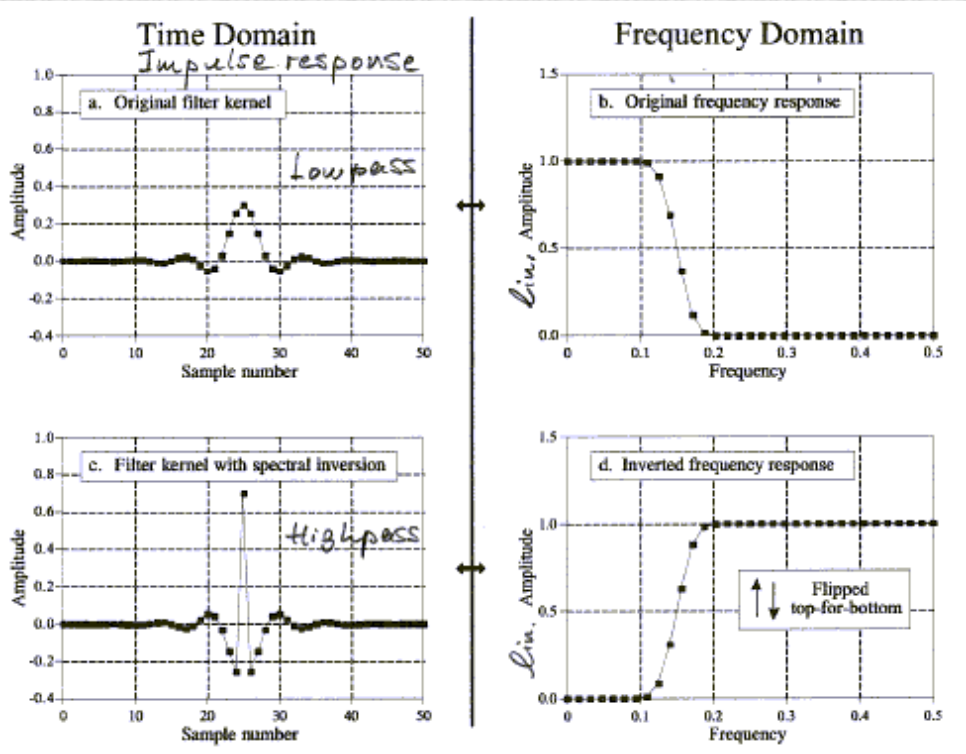
The influence of suggestion, especially when frequently repeated, should not be underestimated in the field of hearing. Many listeners have not acquired an auditory reference that is based on un-amplified sounds and are thus easily misled in matters of accuracy. Double-blind and ABX product comparison tests are frowned upon in some audio circles, though they remove any influence of suggestion, self or other generated.

I have no problem, if people enjoy their music more, because it has passed through a certain interconnect, has been transformer balanced, amplified without feedback, processed through a reissued WW2 transmitter tube and sent to their speakers on litz wires of specific arrangement and purity, except that the money for these suggestions could have been spent on better speakers, for real improvements in accuracy of dynamic range and resolution. After all, the loudspeakers are by far the weakest link in the chain of components that are needed for sound reproduction.
Top

I - Digital crossovers

Some people think that digital crossovers will replace analog ones, because digital filters can be designed with desirable characteristics that are impossible to realize with analog circuitry. In particular, lowpass and highpass filters with extremely steep slopes and linear phase shift are possible. Steep slopes reduce the overlap region between drivers. Linear phase shift eliminates waveform distortion and merely causes a delay of the signal. Such characteristics can be obtained from the digital equivalent of tapped delay line filters, which have a finite impulse response (FIR) duration that depends upon the number of taps used. Digital FIR filters can have almost any desired frequency response, if the number of weighted taps is made sufficiently high. [1]

The linear phase shift comes at a price. The impulse response rings. The more so, the steeper the filter slopes. Both lowpass and highpass sections of the crossover ring, but when the outputs are combined, as for a crossover, then the two impulse responses add to a non-ringing, delayed pulse.



All would be fine, if we listened only in anechoic spaces or to speakers with coincident drivers. In reality we use speakers in rooms with reflections and reverberation and the the drivers are separated from each other due to their sizes. As a consequence the off-axis response of the speaker matters and contributes to what we hear. With the drivers non-coincident, the lowpass and highpass outputs are delayed different amounts at points off-axis, and the ringing is no longer canceled in the addition. In the best case the drivers might be coaxial, but this has another set of problems. Very steep crossovers can also cause a very abrupt change in the polar pattern of the speaker, when transitioning from a large diameter driver to a small one. Under reverberant conditions and/or listening off-axis this may have audible consequences.

A different type of digital filter has an infinite impulse response (IIR), from using recursive or feedback topologies. IIR filters can closely simulate all the known analog filter functions, but also have their associated non-linear phase shifts. The frequency response of a given IIR

filter, such as a biquad, can be easily changed by reprogramming its coefficients. They are readily adapted to different needs.

Digital filters have high accuracy and are very repeatable. They require high performance ADC, DAC and DSP. Analog anti-aliasing and reconstruction filters are part of their design. Dynamic range, clipping and noise must be dealt with in the digital domain of numbers. In particular, the effects of numeric round-off upon filter shapes and distortion must be analyzed. For reasons of limited linear digital range the volume control must be located after the DSP to maximize bit resolution. Non-linear distortion in all its forms between input and analog output has to be considered and minimized. Loudspeaker drivers have their own inherent acoustic highpass amplitude and phase response, which must be accounted for in the digital crossover filter design. Electromagnetic compatibility (EMC) should be designed in and tested for.

[1] Steven W. Smith, "The Scientist and Engineer's Guide to Digital Signal Processing"
Top

J - Doppler distortion

Drivers are now available that generate very low intermodulation distortion, because their motor behavior is very linear. This may bring Doppler distortion to the foreground under certain conditions. Doppler distortion occurs when two tones of largely different frequency are radiated from the same membrane. It is a linear form of distortion, at least to the first order, which means its magnitude is independent of the amplitude of the higher frequency tone and distortion increases linearly with the amplitude of the lower frequency tone..

For example, a 18 cm (6.5 inch) driver with 126 cm^2 cone area like the Seas W18E001 may radiate a 50 Hz low frequency tone at 5 mm peak excursion, which produces 86 dB SPL at 1 m. The sinusoidal cone displacement has a peak velocity of:

$$V_p = 2 \pi f_L X_p = 1.57 \text{ m/s}$$

A 2000 Hz tone which is radiated simultaneously will have its wavelength shortened and lengthened at a 50 Hz rate as the cone velocity is added to and subtracted from the speed of the sound's propagation (c = 343 m/s). Based on a relatively simple analysis this results in a frequency modulation of the 2000 Hz tone with a peak deviation of:

$$\Delta f_H = 2 \pi f_L f_H X_p / c = 9.2 \text{ Hz}$$

The modulation distortion can be reduced to 5.5 Hz for the same 50 Hz SPL by using a W22EX001 driver with 220 cm^2 cone area.

The audibility of Doppler distortion depends upon three criteria:

1. the FM threshold of hearing,
2. masking by the low frequency tone,
3. FM to AM conversion due to room resonance.

The FM threshold is typically determined at a low modulation rate of 4 Hz and is about 0.35% for frequencies above 500 Hz, corresponding to 7 Hz deviation at 2000 Hz. Higher modulation rates around 70 Hz may be perceived as roughness of the tone.
Masking is dependent upon the relative levels and frequency separation of the two tones.

A room with a reverberation time T_{60} = 500 ms has resonances with a 3 dB bandwidth of BW = 2.2 / T_{60} = 4.4 Hz. The slope of the filter curves is about 3 dB per BW/2 or 1.4 dB/Hz. Thus, a 9.2 Hz deviation of the 2000 Hz tone could generate up to 12.9 dB change in its amplitude, if the rate was slow enough so that the sidebands of the FM signal are not significantly attenuated by the filter.

Doppler distortion of a driver can be measured with a spectrum analyzer by observing the amplitude of the frequency or phase modulation sidebands A_s relative to the carrier A_H. The relative amplitude for the first pair of sidebands is approximately

$$A_s / A_H = 0.5 \Delta f_H / f_L = \pi f_H X_p / c = 0.09$$

or 20.7 dB below the 2000 Hz carrier A_H, regardless of f_L as long as X_p = 5 mm. This corresponds to a phase modulation with a maximum phase deviation of $2 A_s / A_H = 2 \times 0.09$ rad = 0.18 x 180 / π = 10.3 degree.

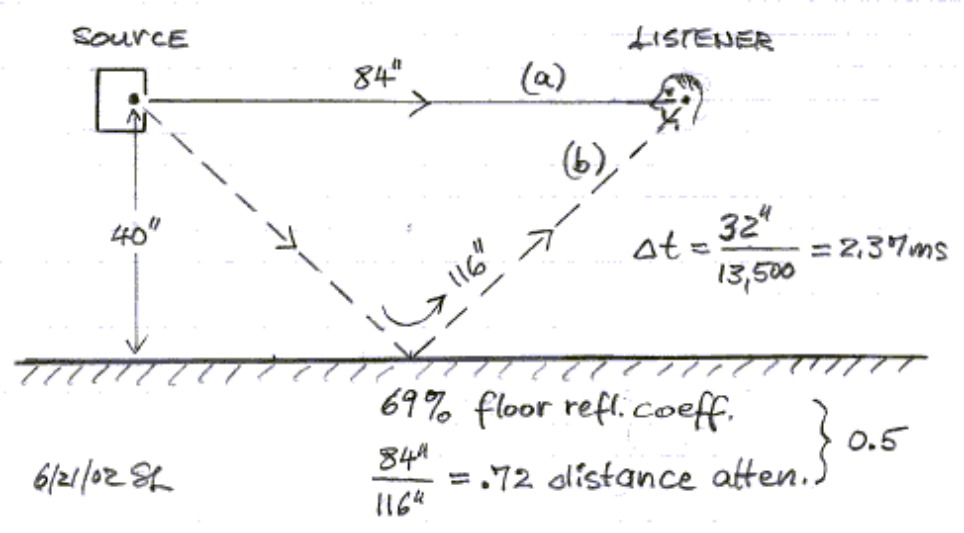
The sidebands occur at the same frequencies $f_H - f_L$ and $f_H + f_L$ as the 2nd order intermodulation products of the two tones. They can be differentiated by changing the applied signal levels. Doppler sidebands decrease 3 dB for a 3 dB reduction in signal levels, whereas 2nd order intermodulation sidebands decrease 6 dB. The two sidebands may be of different amplitude when both forms of distortion are present, because the intermodulation products are in-phase with each other, the FM sidebands are out-of-phase. This fact allows a decomposition of unequal sidebands into their two constituent odd and even parts.

I have experimentally verified the results of the above analysis by measuring the relative amplitude of the sidebands in the frequency spectrum. The Doppler effect is very difficult to measure in the time domain, though, because of the small amount of phase shift (e.g. 10.3 degree) that it causes for the higher frequency tone. Rod Elliott claims to have been able to measure the shift in the time waveform. An analysis of his observation confirms that the maximum phase shift occurs when the cone is at its maximum excursion, where its velocity is zero. The Doppler analysis usually mentions only the maximum frequency shift, which occurs when the cone has its highest velocity, or when going through zero excursion for a sinusoidal low frequency tone. Both results are correct and are a consequence of the fact that the radian frequency ω = d(phase) / dt . This leads to a 90 degree phase shift between time varying frequency and phase for a sinusoidal signal. Either way of looking at the situation results in identical magnitudes of the spectral components. It all is the Doppler effect, though I have not seen the associated phase modulation pointed out in text books.

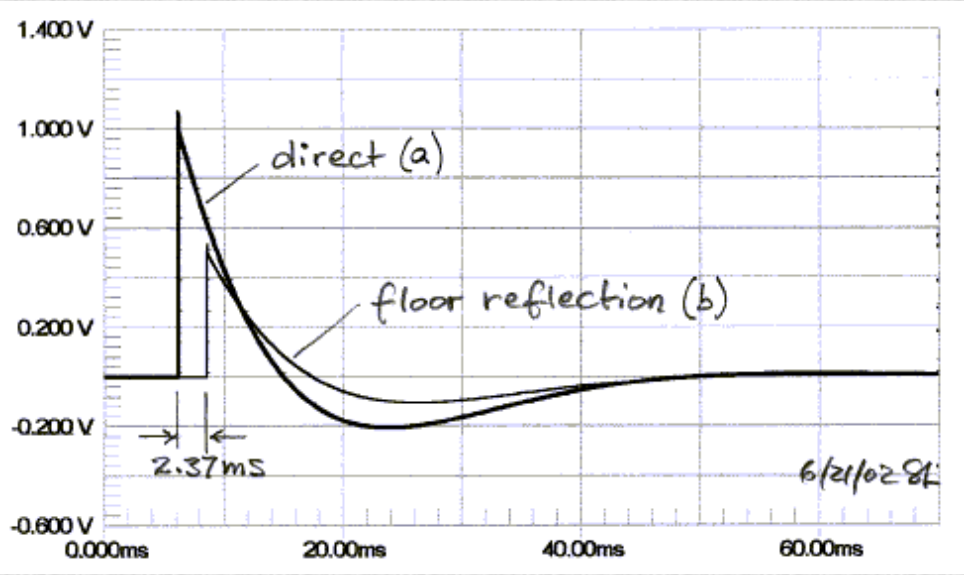
Top

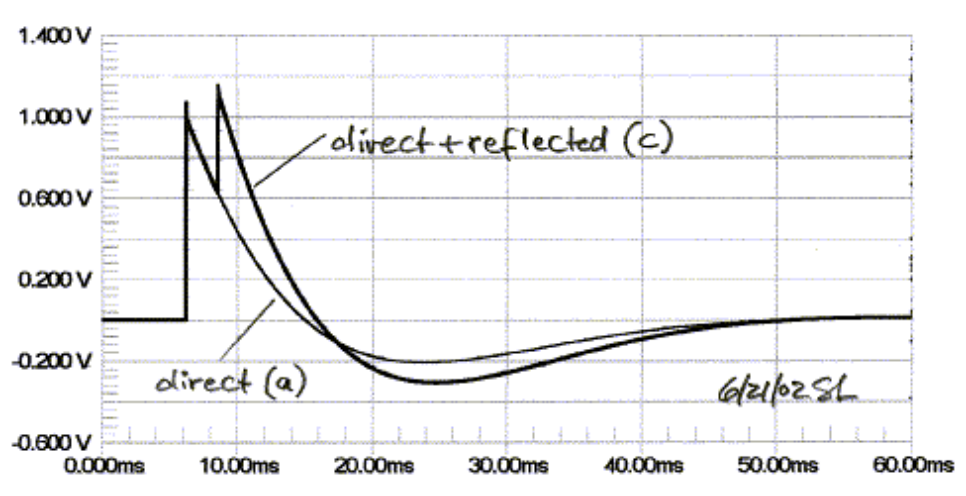
K - Room response equalization

In theory, digital signal processing allows to remove completely the effect of the room upon the frequency response of a loudspeaker at the listening position, but only over a very small region in space. The principle is illustrated by correcting the output signal from a loudspeaker in order to cancel the effect of the inevitable floor reflection upon the response at the listener's place.

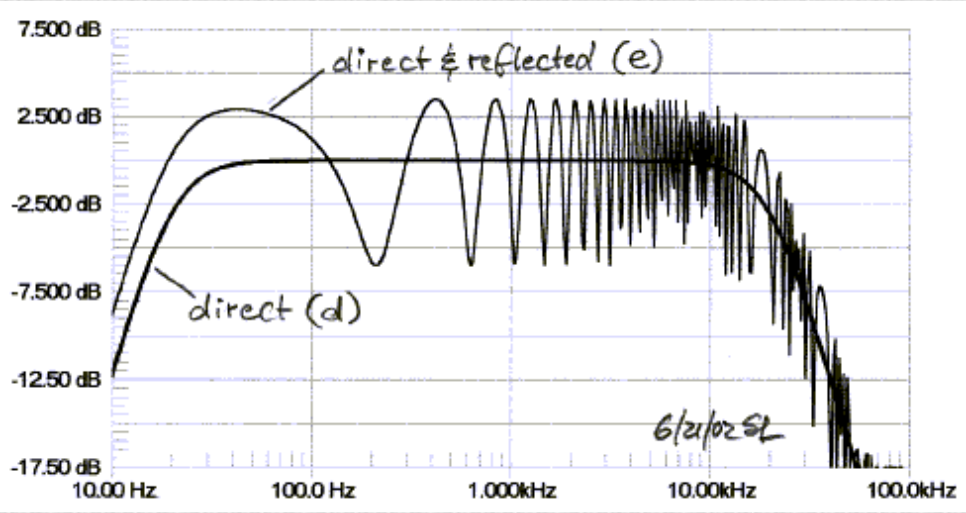


The model assumes a speaker with 20 Hz to 20 kHz response. The floor is not perfectly reflecting and together with the attenuation, due to a 32" (0.81 m) longer path length for the reflected signal (b) than the direct signal (a), the amplitude of the reflected signal at the listener will be half that of the direct signal. The floor reflection is delayed by 2.37 ms. The step response at the listener position is the sum of direct and reflected signals (c).

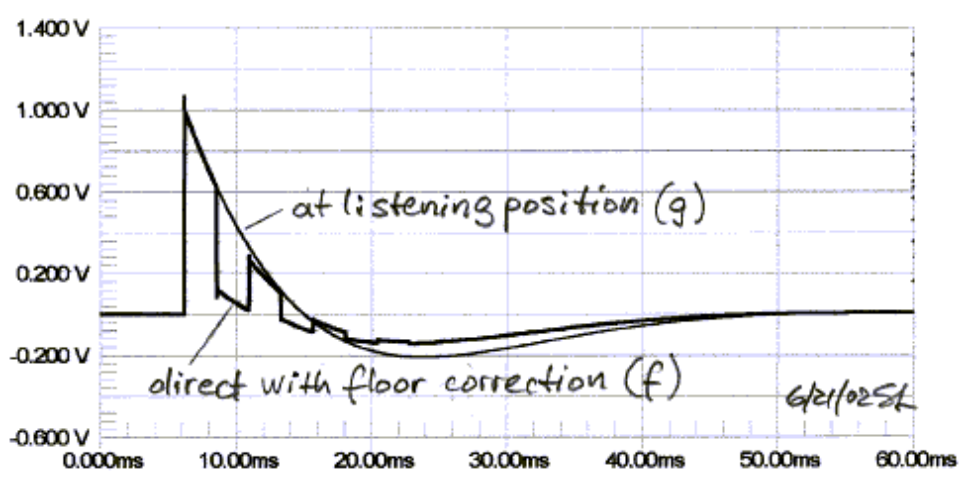




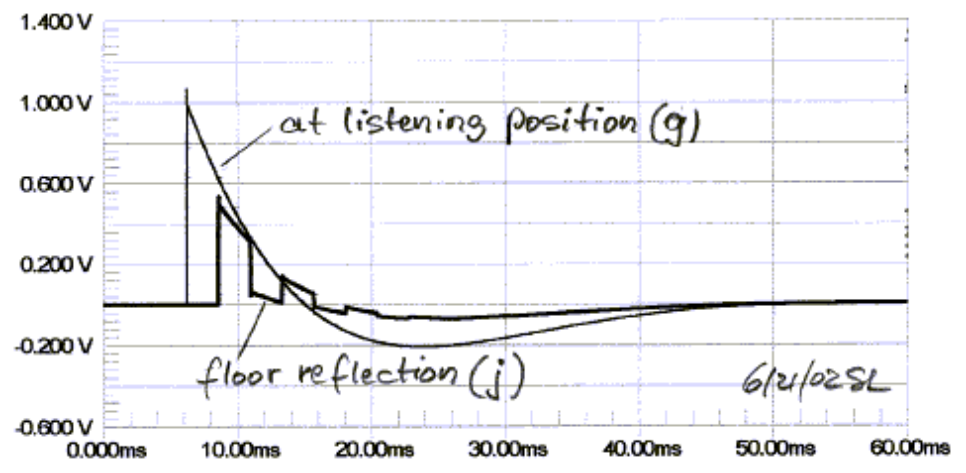
The frequency response at the listener position shows the typical comb filter behavior (e). At low frequencies the floor reflection adds to the direct sound (d). The first minimum in the response occurs when the 32" path length difference corresponds to 1/2 wavelength and causes 180 degrees of phase shift, which is at 211 Hz. The next peak is at 422 Hz, followed by a minimum at 633 Hz, and so on. The response is non-minimum phase and can be corrected with a minimum phase parametric equalizer to some degree, but only at the very low frequencies.



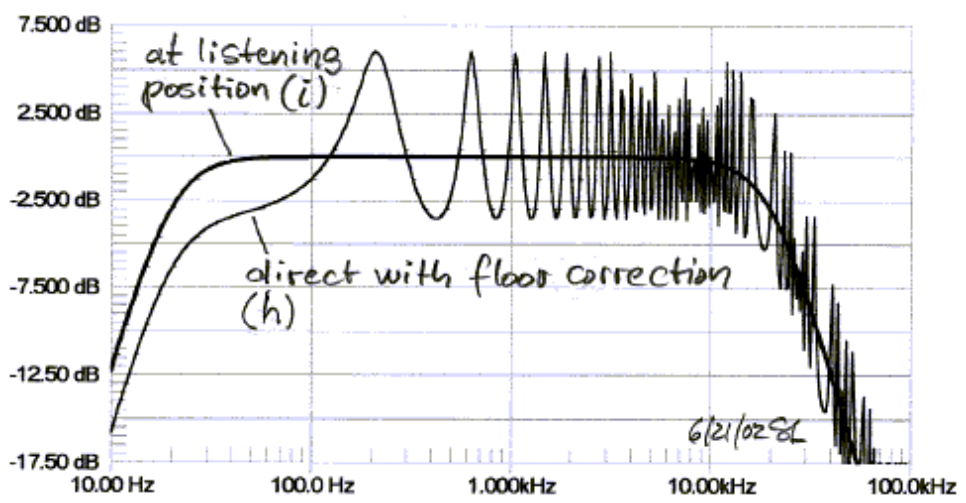
The effect of the floor reflection can be cancelled by pre-distorting the direct signal from the speaker. After 2.37 ms from the start the direct signal has to have a negative step to cancel the first reflected signal which arrives at the listener. The negative step in turn causes a floor reflection which must be cancelled by a positive step 4.74 ms later, and so on. Since the floor reflection is attenuated, consecutive correction steps will decrease in amplitude (f).



Direct (f) and reflected (j) signals combine at the listener position for a perfect step response (g), identical to the anechoic response (a).



The frequency response of the pre-distorted direct signal (h) may have large peaks to correct for the notches in the room response (e).



It should be obvious that this form of room correction is a delicate balancing act, particularly to maintain equalization at higher frequencies. Small deviations in listener position from the point for which the speaker was equalized, cause large changes in room response. In real rooms there is not only the floor reflection, but all the many other reflections from walls, ceiling, furniture, etc. that enter the correction of the speaker. Since the correction can at best be valid only for a single sweet spot, it will necessarily send a distorted signal to every place else. Even at the sweet spot the system performance does not degrade gracefully, when the correction is less than perfect. Gross spectral distortion is introduced at higher frequencies and may become aurally distractive.

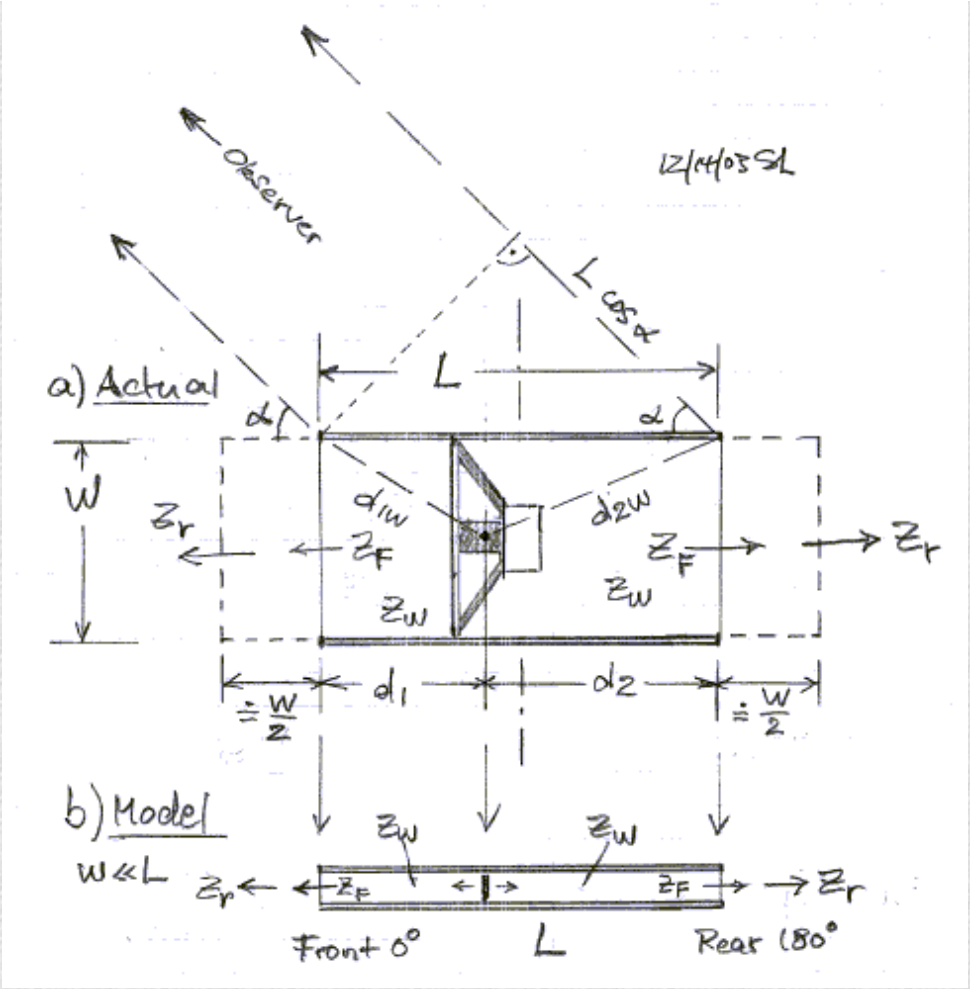
Room equalization with DSP will require a new understanding of what should be corrected, and what should be left alone, to meet a specific overall objective for sound reproduction. For example, the comb filter response due to floor reflection is intimately familiar to us, since most sounds we hear, have at least a floor reflection. As we grew up and became taller the comb spectrum shifted towards lower frequencies. How do we use the information conveyed by it? How important is it to correct for it in a speaker setup? The issue with room equalization is not about sufficient computing horse power, but how and where to apply it.

The following two references are highly recommended reading before embarking on any digital room equalization project.

[1] Louis D. Fielder, Practical Limits for Room Equalization, 111th AES Convention, 2001, Preprint 5481
[2] Scott D. Snyder, Active Noise Control Primer, Springer/AIP, 2000
Top

L - H-frame and U-frame open-baffle woofers

The analysis of a H-frame dipole woofer is rather simple due to the symmetry of the mechanical structure combined with the fact that front and rear acoustic outputs are of reversed polarity. The influence of a number of important parameters upon the radiated field can be ignored, since their effect is equal in magnitude on both sides of the baffle and cancels because of the 180 degree phase difference. The analysis becomes more complicated when the frame is no longer symmetrical and the driver is moved closer to one of the openings of the frame.

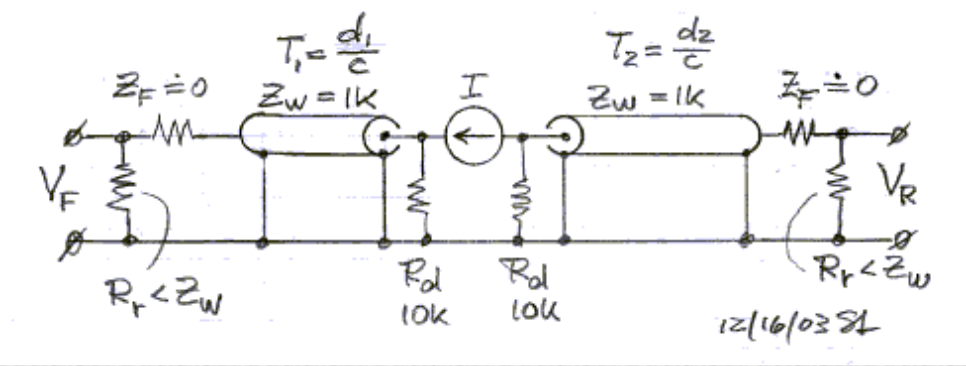


The baffle structure has a physical length L = d1 + d2 and width and height W. The acoustic length is larger by about W/2 on each end, similar to the end correction of the vent of a Helmholtz resonator. For determining the polar pattern I will assume the case where W << L, and work only with the physical length L. Any calculated 1/4 wavelength resonance frequency will be higher than in actuality, but that does not invalidate the relative performance at different driver positions d1. The observation point is far enough away so that the outputs from front and rear are seen at the same angle α. The two propagation paths then differ in length by Lcos(α).
I assume that the driver is operated in its mass controlled region and acts as a volume source or like an electrical current generator. Using an actual driver model would complicate the analysis and only be beneficial if we needed to know absolute numbers. I will also assume that the source is fully symmetrical, as would be the case with two drivers having their magnets pointed in opposite directions. A single driver would introduce additional phase shift on its backside due to the acoustic elements associated with the cavities around the basket and magnet structure.

Looking from the driver cone towards the rear of the baffle the displaced volume propagates through a waveguide of acoustic impedance Z_W.

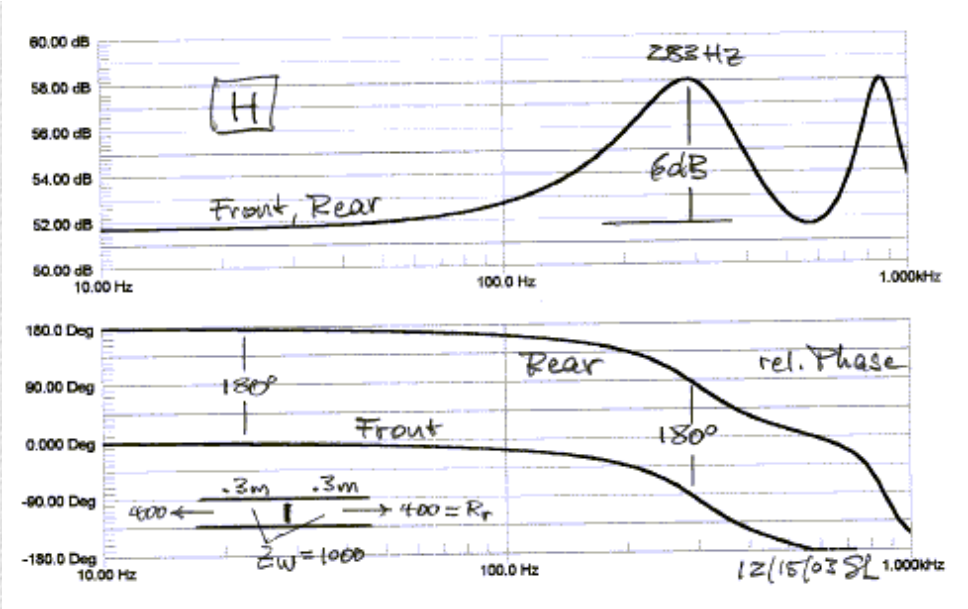
If the waveguide were closed off at its end by a wall, then the volume flow would come to a stop as it reaches the wall and be reflected back to the source. The wall can be described by an infinite flow impedance Z_F. When the the waveguide is open ended the corresponding flow

impedance Z_F is nearly zero. In reality, Z_F is an acoustic mass or, electrically, an inductive element. A small percentage of the flow is reflected back from the opening towards the source. When the wave hits the cone it is reflected back towards the opening, where again a small fraction of it bounces back to the cone. This back and forth travel sets up a resonance at the frequency where d2 equals a quarter wavelength. The major percentage of the volume flow exits the opening and expands into the larger free space. Consequently, its pressure drops relative to what it was inside the waveguide. This can be described by an impedance Z_r which is less than Z_w. The impedance mismatch exists at all frequencies. The pressure developed across the real part R_r of this radiation impedance Z_r is the quantity of interest.

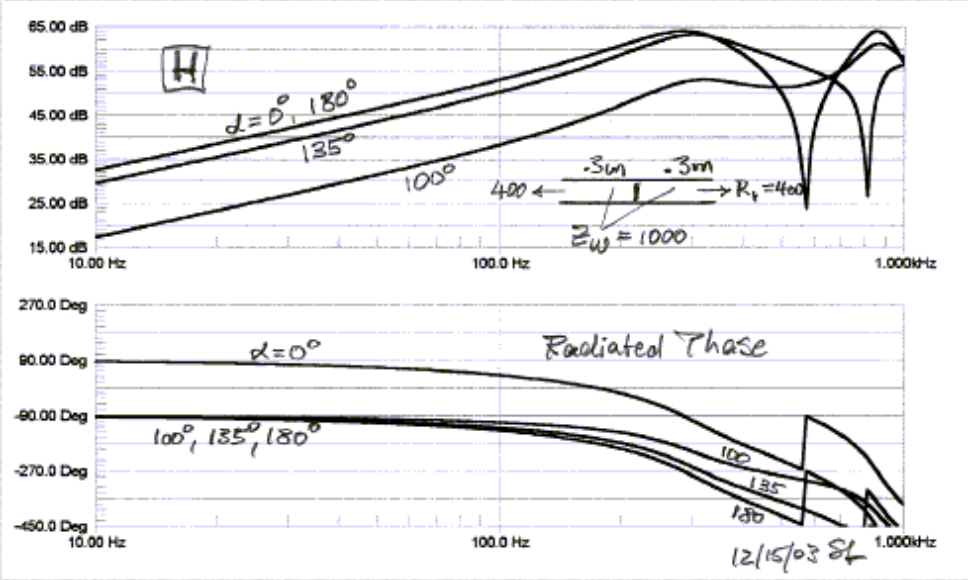


The circuit above describes the acoustic behavior in electrical terms. The two loss-less transmission lines are specified by their impedance Z_w and propagation delay T as function of their length d. The current generator has a finite impedance $R_d >> Z_w$. The flow impedance Z_F is assumed to be zero for the moment. The voltages V_F and V_R across the radiation resistors R_r are summed vectorially using the path length difference L cos(α) to determine the phase shift between front and rear outputs. The summed signal is proportional to the SPL at observation angle α.

It is important to work with representative values of R_r in the electrical model. Assuming R_r = 400 and Z_w = 1000 leads to a 1/4 wave resonance peak of about 6 dB. This is in the range of peak magnitudes that I have measured in the opening plane of H-frame woofers. All modeling is done for L = 0.6 m (24"), thus for the symmetrical H-frame the resonance occurs at 340m/s / (4 x 0.3m) = 283 Hz. The graph below shows magnitude and phase of front and rear radiation.

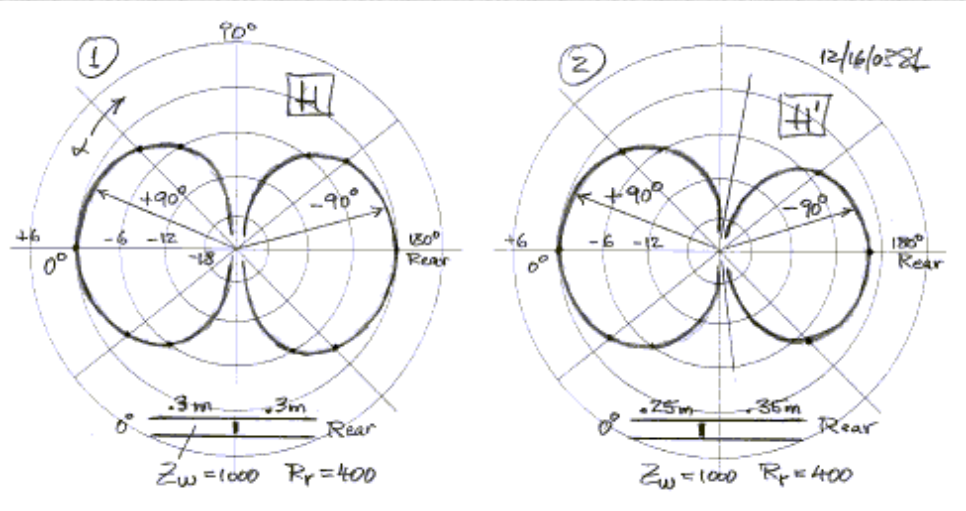


The frequency response at different observation angles α is obtained by adding the front and rear outputs vectorially. The analysis was carried out using CircuitMaker which is a SPICE program. Notice the slight amount of peaking near the resonance. The outputs sum to zero at 90 degrees. This would be off the graph. Instead, the 100 degree response is shown.

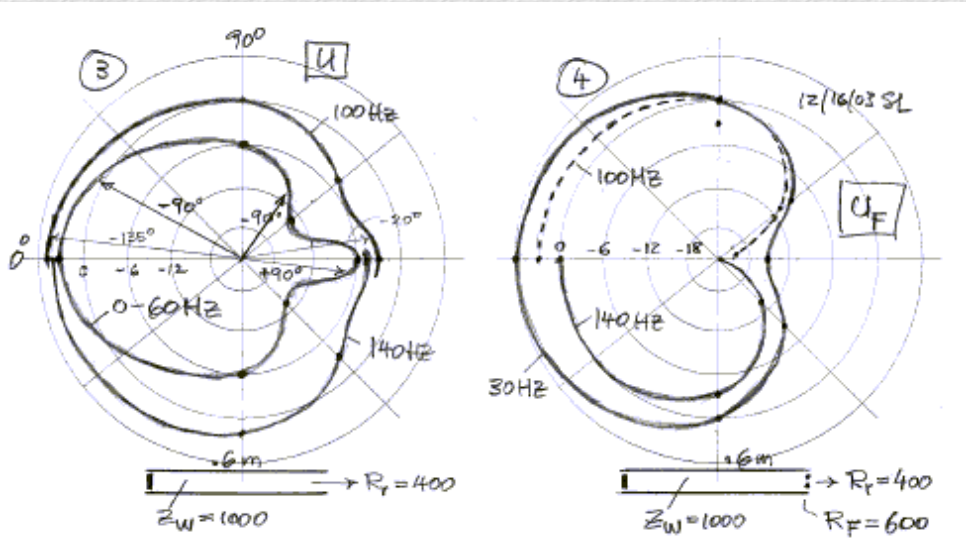


The polar plot (1) below is drawn on a dB scale with 6 dB increments between circles. Note that the phase of the acoustic wave to the left of α = 90 degrees is at +90 degrees, to the right at -90 degrees relative to the volume source excitation.
When the source is shifted slightly towards the front of the H-frame, then the null occurs at an angle α that is greater than 90 degrees. Also,

the rear lobe decreases slightly in magnitude (2).



This trend continues and leads to the polar pattern (3) where the driver has been fully moved to the front of the waveguide. It is the U-frame configuration. The magnitude of the polar response has increased over that of the H-frame in the left half of the polar graph. On-axis and below 60 Hz by 3.4 dB, and at 90 degrees there is no longer any null. The right half of the graph takes on different shape that are quite frequency dependent. There is no null, but there is a small region where the phase of radiation is +90 degrees as compared to -90 degrees for the major portion of direction angles α.



The shape of the polar pattern is strongly influenced by the impedance that the wave sees at the rear opening. By inserting a flow resistor R_F = 600, so that the waveguide is terminated reflection free, the response begins to look more like a cardioid. The on-axis response is 6 dB higher than that of the H-frame of same length L.

A cardioid polar response is obtained by the summation of a monopole response with magnitude 0.5 in all directions of radiation, M = 0.5, and a dipole with magnitude 0.5 on-axis, D = 0.5 cos(α), such that C = 0.5 + 0.5 cos(α) . A monopole, such as a closed box woofer, is a pure pressure source. A dipole, like a H-frame woofer, is a pure velocity source. A cardioid or U-frame woofer would exhibit characteristics of both monopole and dipole. It is not clear what the room acoustic benefits might be. Unlike the H-frame with its 90 degree nulls, these types of mixed radiators can excite room modes at ALL angles. Even the null towards the rear of a pure cardioid woofer does not prevent the excitation of the room mode along the 0 degree to 180 degree axis.

I have heard a rock band that used big Meyer Sound cardioid woofers in a very open setting. Walking around the stage there was a clearly audible reduction of low frequencies in the back. More slam for the audience, less for the neighborhood.
In a normal living room there could be an application for a cardioid above 1 kHz to reduce reflections from objects behind the speaker, if that were an issue.
It's a different story for microphones where the cardioid pattern can provide a clear benefit by reducing unwanted sound pickup.

If you have any interested in further analysis of the U-frame woofer, then go to Appendix L .

Top

Siegfried Linkwitz == webmaster@linkwitzlab.com

Last revised: 05/07/2006 - © 1999-2006 LINKWITZ LAB, All Rights Reserved

Home

Doppler shifts in loudspeaker. Fact or fiction?

Updated Novermber 24, 2004

The document below presents a brief analysis of wave form distortion in loudspeaker dynamic drivers when reproducing both a single tone as well as 2 arbitrary tones such that the different frequencies are related as Fh = n Fl, where n is any number greater than 1.0. The single tone analysis shows that when a driver is subject to a single sine wave input, the acoustic output will under go a distortion of the wave form. The level of this distortion is dependent on the driver displacement. In the two tone case, it is shown that in addition to the wave form distortion present for a single tone, there is also a frequency shift in the higher frequency. The amount of frequency shift is primarily dependent on the excursion of the driver associated with the lower of the 2 frequencies, and completely independent of the excursion related to the higher frequency. If the frequency shift is give by df, then it is shown that the higher frequency tone oscillates about the input frequency Fh as F = Fh + df sin(2*Pi*Fl*T). It is also shown that the magnitude of df asymptotes to a limiting value as the ratio of the frequencies, n, gets large. The frequency shift is shown to result in the creation of intermodulation sidebands, thus rsulting in IM distortion of the input signal. The document is in MS Word format in compressed zip form, about 73k and down loads quickly.

On the next page I present experimental data supporting the results predicted by the analysis.

Download Original Article
(Written Feb. 2001)

As of Noverber 24, 2004 I have added a second paper where I revisit wave form distortion arising from Doppler effects. This new paper, which parallels the analysis below, is much simpler and easier to follow than the previous paper and yields similar results. This new paper is available for download at the link below. The paper is in MS Work format in compressed zip form, about 29k.

Download New Article
(Written Nov. 2004)

11/21/04

In a recent article By Rod Elliott on Doppler distortion in loudspeakers Mr. Elliott examines the frequency shifts associated with the higher frequency in loudspeakers radiating two different frequencies and concludes that the observed frequency shifts are not a result of Doppler effects. He bases this conclusion on the premise that, "If the shift observed is indeed in frequency,and caused by the Doppler effect, then it must occur at the point where the cone velocity is at its peak - at the zero crossoing point of the wave form. However, if the shift is simply one of phase, then the peak shift will occur at the extremes of excursion, where the cone velocity is effectively stationary." Mr. Elliott also states, " Papers such as that by John Kreskovsky are very convincing, but unfortunately work from the wrong premise in the first place." He then goes on to develop his conclusion that the shifts are a result of phase shifts due to the relative position of the driver's cone with respect to the listener. While it is certainly valid to pursue the problem from the perspective of phase shifts rather than Doppler related changes in wave lengths, as I proceeded, a very simple analysis is all that is required to demonstrate the phase based argument is nothing more than a restatement of the Doppler Effect.

In light of these ststements I have prepared the following rebuttal.

Consider the signal

$$S(t) = \sin(\omega_h t + \Phi)$$

Φ is the phase shift and is related to ω_h as,

$$\Phi = \omega_h \Delta t$$

where Δt is incremental the time delay to listener due to the changing position of the cone.

$$\Delta t = (x_o - x) / c$$

Here c is the sound speed and x_o is the distance from cone to driver when the cone is at rest. If the cone velocity due to some low frequency motion is U_l then

$$x = x_o - \text{Integral}(U_l\, dt)$$

and

$$\Delta t = \text{Integral}(U_l\, dt)/c$$

Thus,

$$S(t) = \sin(\omega_h t + \omega_h \text{Integral}(U_l\, dt)/c)$$

If we consider the case where U_l is constant,

$$S(t) = \sin(\omega_h t\, (1 + U_l/c))$$

and we find

$$\omega_h' = \omega_h (1 + U_l/c) = \omega_h + \Delta\omega_h$$

where $\Delta\omega_h$ is the Doppler related frequency shift.

If $U_l = Um \cos(\omega t)$

$$\Delta t = U_m/(\omega_l c) \sin(\omega t)$$

and

and

$\Phi = U_m \omega_h /(\omega_l c) \sin(\omega_l t)$

Thus,

$S(t) = \sin(\omega_h t + ((\omega_h U_m /c)/ \omega_l) \sin(\omega_l t))$

It is evident from the equation for ω_h' above that $(\omega_h U_m /c)$ is nothing more than the maximum Doppler related frequency shift, $\Delta\omega_h$, which will occur over a low frequency cycle and

$S(t) = \sin(\omega_h t + \Delta\omega_h /\omega_l \sin(\omega_l t))$.

This is the same result obtained in my paper where $\Delta\omega_h /\omega_l$ is the modulation index and shows that whether the analysis begins with consideration of wave length compression/expansion or phase variation the result is identical and lead to the generation of intermodulation sidebands due to Doppler related effects.

Finally note that

$(\omega_h t + \Delta\omega_h /\omega_l \sin(\omega_l t))$

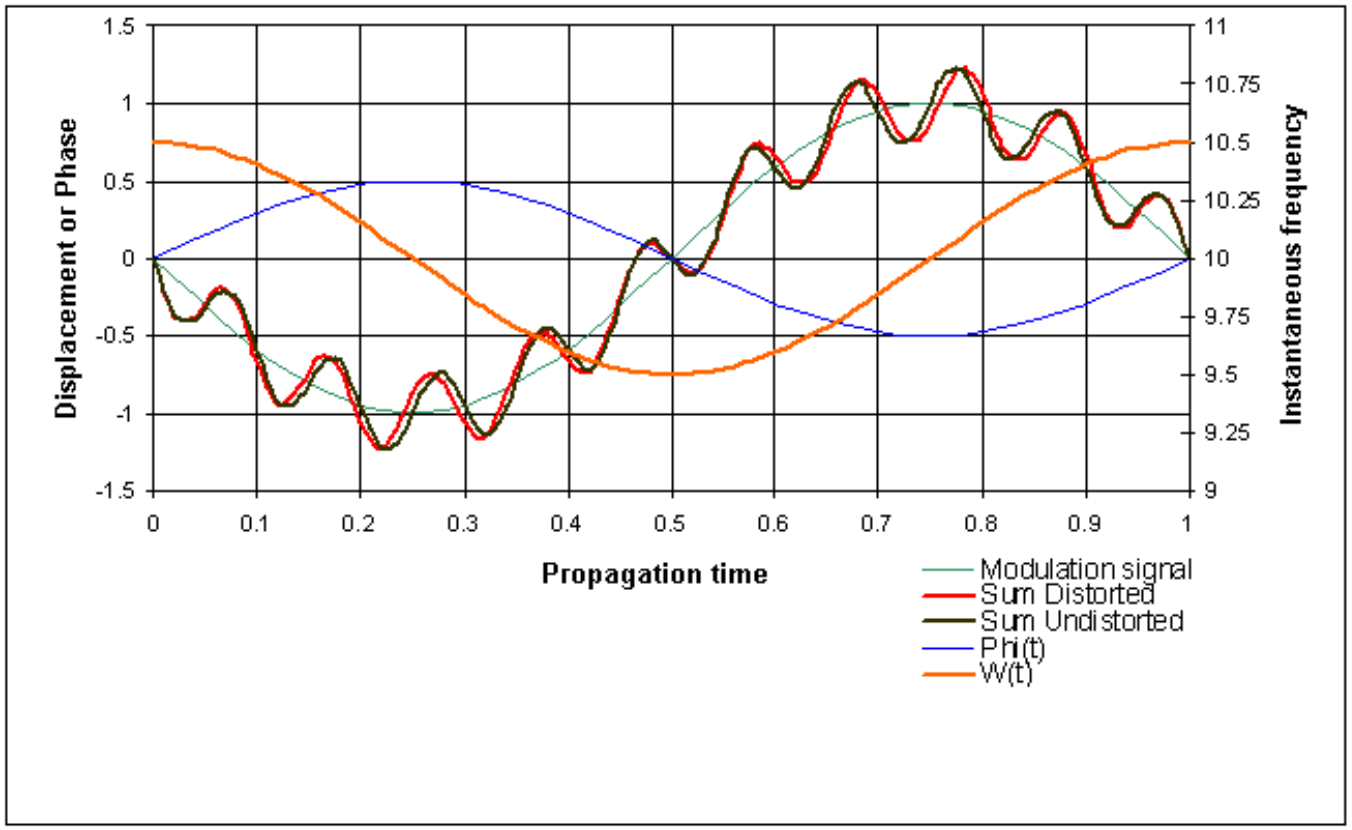
can be considered the instantaneous phase of the signal,

$\Phi(t)$, and the instantaneous frequency is given by

$d\Phi /dt = \omega_h'(t) = \omega_h + \Delta\omega_h \cos(\omega_l t)$

and we note the 90 degree difference between where the phase shift is maximum and where the frequency shift is maximum. Since, as Mr. Elliot measured, the maximum phase shift occured at the the point of maximum excursion, it follows that the point of maximum frequency shift does indeed occur at the point of zero crossoing where the cone velocity is at a maximum as would be expected if it were a result of Doppler effects.

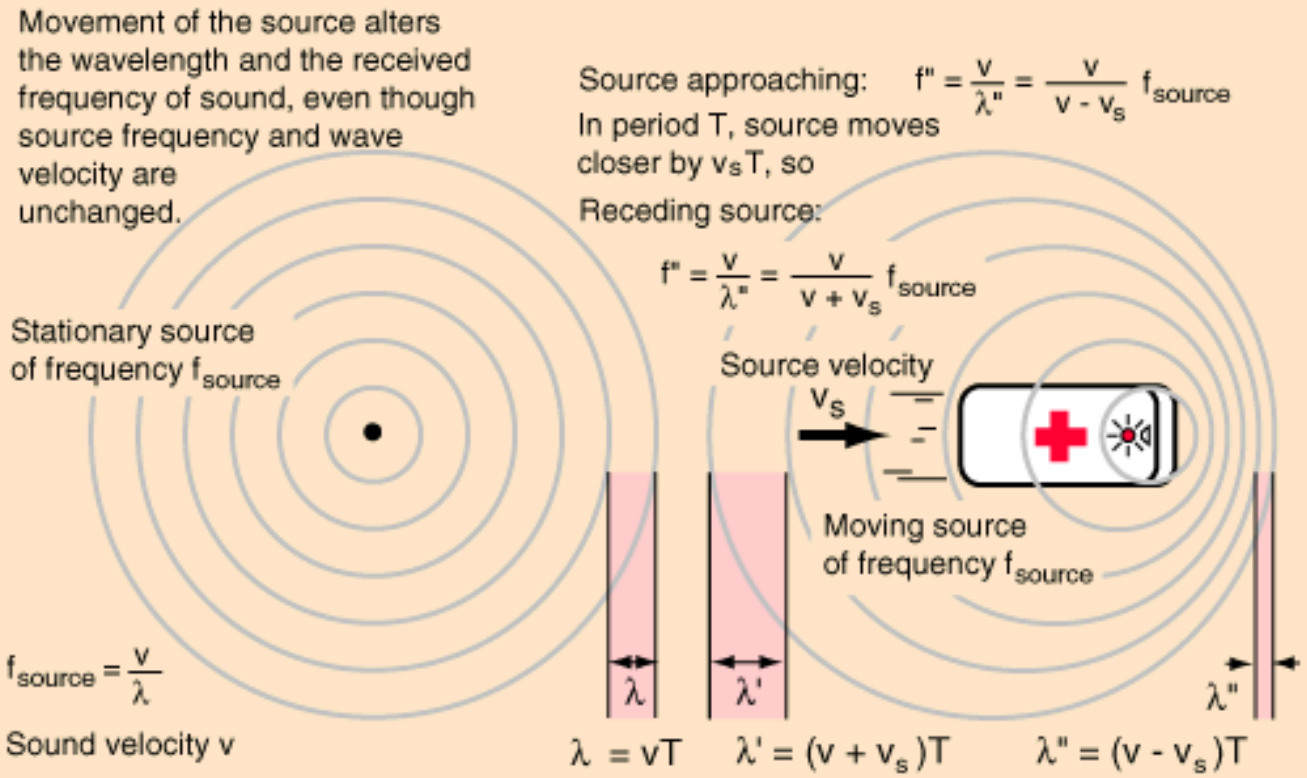
The figure to the left shows the results of the analysis above for a modulation frequency of 1, a frequency ratio of 10 and a modulation index of 0.5. The thin green line is the displacement due to the modulation signal. The black line is the sum sin(T) + sin(10xT). The red line is the sum of sin(T) + the distorted signal. The orange line is the instantaneous frequency of the distorted signal. Note that the frequency is a maximum at the zero crossing of the modulation signal



(green line). The blue line is the time dependent phase of the distorted signal. Note that it is a maximum at the peak displacement of the modulation signal and is consistent with the measured data of Rod Elliott. However, Mr. Elliot's conclusion that this indicates the shift is not a result of Doppler effects is incorrect as the maximum frequency shift does occur at the poit of zero crossoing where the velocity due to the modulation signal is a maximum, a Doppler result.

Doppler Effect

You hear the high pitch of the siren of the approaching ambulance, and notice that its pitch drops suddenly as the ambulance passes you. That is called the Doppler effect.



Discussion | Calculation of wavelength | Calculation of frequency

Police RADAR as an example of the Doppler effect

Ultrasonic Doppler pulse probes

Index

Traveling wave concepts

HyperPhysics***** Sound

R Nave

Go Back

Doppler Effect

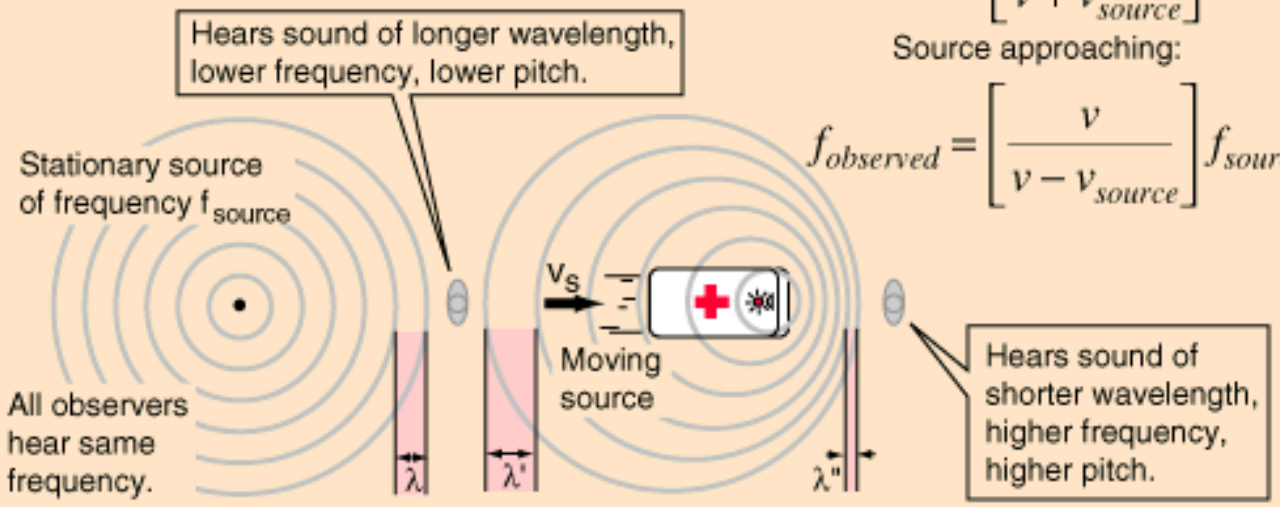
When a vehicle with a siren passes you, a noticeable drop in the pitch of the sound of the siren will be observed as the vehicle passes. This is an example of the Doppler effect. An approaching source moves closer during period of the sound wave so the effective wavelength is shortened, giving a higher pitch since the velocity of the wave is unchanged. Similarly the pitch of a receding sound source will be lowered.

Source receding:

$$f_{observed} = \left[\frac{v}{v + v_{source}}\right] f_{source}$$

Source approaching:

$$f_{observed} = \left[\frac{v}{v - v_{source}}\right] f_{source}$$



Derivation	Calculation of wavelength	Calculation of frequency	Doppler effect for light

Index

Traveling wave concepts

Doppler concepts

Doppler Wavelength Change

The speed of sound is determined by the medium in which it is traveling, and therefore is the same for a moving source. But the frequency and wavelength are changed. The wavelengths for a moving source are given by the relationships below. It is sometimes convenient to express the change in wavelength as a fraction of the source wavelength for a stationary source:

Source receding, wavelength increase.

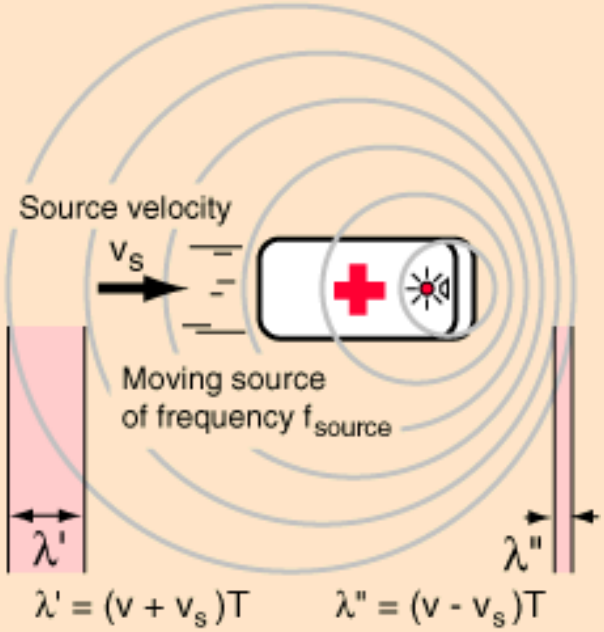
$$\lambda' = (v + v_s)T = (v + v_s)\frac{\lambda_s}{v} = (1 + \frac{v_s}{v})\lambda_s$$

$$\frac{\Delta\lambda}{\lambda} = \frac{(1 + \frac{v_s}{v})\lambda_s - \lambda_s}{\lambda_s} = \frac{v_s}{v}$$

Source approaching, wavelength decrease.

$$\lambda" = (v - v_s)T = (v - v_s)\frac{\lambda_s}{v} = (1 - \frac{v_s}{v})\lambda_s$$

$$\frac{\Delta\lambda}{\lambda} = \frac{(1 - \frac{v_s}{v})\lambda_s - \lambda_s}{\lambda_s} = \frac{-v_s}{v}$$



Derivation | Calculation of wavelength | Calculation of frequency | Doppler effect for light

HyperPhysics***** Sound

R Nave

Index

Traveling wave concepts

Doppler concepts

Go Back

Doppler Effect Frequency Calculation

At temperature [ ] C = [ ] F

the sound speed in air is [ ] m/s.

If the source frequency is [ ] Hz

and the velocity of the source is [ ] m/s = [ ] mi/hr

then for an approaching source the frequency is [ ] Hz

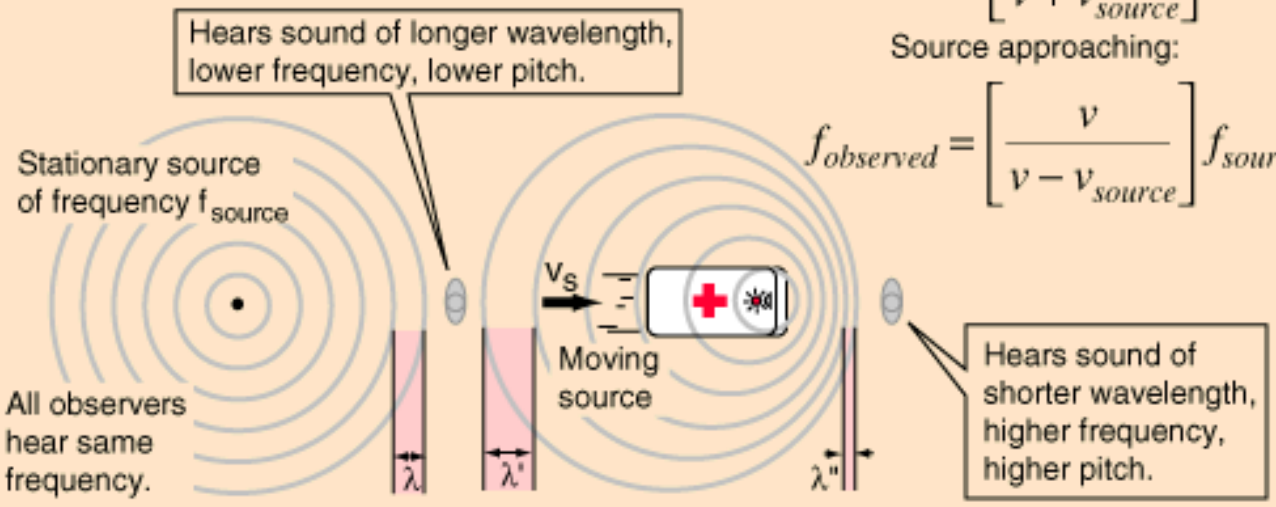
and for a receding source the frequency is [ ] Hz.

Source receding:

$$f_{observed} = \left[\frac{v}{v + v_{source}}\right] f_{source}$$

Source approaching:

$$f_{observed} = \left[\frac{v}{v - v_{source}}\right] f_{source}$$



Note: The frequency will default to A4 (440 Hz) and the temperature will default to 20 C if those values are not entered. Any parameters can be changed.

Derivation | Calculation of wavelength | Doppler effect for light

Index

Traveling wave concepts

Doppler concepts

Pitch

Pitch = frequency of sound

For example, middle C in equal temperament = 261.6 Hz

Sounds may be generally characterized by pitch, loudness, and quality. The perceived pitch of a sound is just the ear's response to frequency, i.e., for most practical purposes the pitch is just the frequency. The pitch perception of the human ear is understood to operate basically by the place theory, with some sharpening mechanism necessary to explain the remarkably high resolution of human pitch perception.

The place theory and its refinements provide plausible models for the perception of the relative pitch of two tones, but do not explain the phenomenon of perfect pitch.

The just noticeable difference in pitch is conveniently expressed in cents, and the standard figure for the human ear is 5 cents.

More detail

Index

Hearing concepts

Place theory concepts

Details About Pitch

Although for most practical purposes, the pitch of a sound can be said to be simply a measure of its frequency, there are circumstances in which a constant frequency sound can be perceived to be changing in pitch.

One of most consistently observed "psychoacoustic" effects is that a sustained high frequency sound (>2kHz) which is increased steadily in intensity will be perceived to be rising in pitch, whereas a low frequency sound (<2kHz) will be perceived to be dropping in pitch. (More detail)

The perception of the pitch of short pulses differs from that of sustained sounds of the

Index

Hearing concepts

same measured frequency. If a short pulse of a pure tone is decaying in amplitude, it will be perceived to be higher in pitch than an identical pulse which has steady amplitude. Interfering tones or noise can cause an apparent pitch shift.

Further discussion of these and other perceptual aspects of pitch may be found in Chapter 7 of Rossing, The Science of Sound, 2nd. Ed.

HyperPhysics***** Sound *R Nave* Go Back

Effect of Loudness Changes on Perceived Pitch

A high pitch (>2kHz) will be perceived to be getting higher if its loudness is increased, whereas a low pitch (<2kHz) will be perceived to be going lower with increased loudness. Sometimes called "Stevens's rule" after an early investigator, this psychoacoustic effect has been extensively investigated.

With an increase of sound intensity from 60 to 90 decibels, Terhardt found that the pitch of a 6kHz pure tone was perceived to rise over 30 cents. A 200 Hz tone was found to drop about 20 cents in perceived pitch over the same intensity change.

Studies with the sounds of musical instruments show less perceived pitch change with increasing intensity. Rossing reports a perceived pitch change of around 17 cents for a change from 65 dB to 95 dB. This perceived change can be upward or downward, depending upon which harmonics are predominant. For example, if the majority of the intensity comes from harmonics which are above 2 kHz, the perceived pitch shift will be upward.

Index

Hearing concepts

References
Terhardt

Rossing
Science of Sound
Ch 7

HyperPhysics***** Sound *R Nave* Go Back

Perfect Pitch

"Perfect pitch" or "absolute pitch" refers to the ability of some persons to recognize the pitch of a musical note without any discernable pitch standard, as if the person can recognize a pitch like the eye discerns the color of an object. Most persons apparently have only a sense of relative pitch and can recognize a musical interval, but not an isolated pitch.

Rossing suggests that less than 0.01% of the population appear to be able to recognize absolute pitches, whereas over 98% of the population can do the corresponding visual task of recognizing colors with no color standard present.

See Rossing, 2nd Ed, p122, sec 7.7

Index

Hearing concepts

Types of Modulation

Amplitude Modulation (AM)

In Amplitude modulation (AM), the information signal is mixed with the carrier signal in such a way as to cause the AMPLITUDE of the carrier to vary at the frequency of the information signal.

Figure 6: Amplitude Modulation

Frequency Modulation (FM)

With frequency modulation, the modulating signal and the carrier are combined in such a way that

causes the carrier FREQUENCY(fc) to vary above and below its normal(idling) frequency. The amplitude of the carrier remains constant as shown in figure below.

Figure 7: Frequency Modulation

As the voltage of the modulating signal increases in the positive direction from A to B, the frequency of the carrier is increased in proportion to the to the modulating voltage. As the voltage reduces from B to C, the frequency of the carrier decreases until it is back to the original value at C. the modulating voltage increases in the negative direction from C to , hence the carrier frequency decreases to a minimum value at D. It will start increasing again from D to E until it reaches its original value again. Thus in one cycle, the frequency has gone up and down (i.e. above and below the idling frequency) as mentioned earlier.

Throughout this process the amplitude of the carrier is not affected. The carrier frequency change above and below that of the unmodulated condition is proportional to sign and amplitude of the modulating signal.

Phase Modulation (PM or Indirect FM)

Phase modulation is a type of frequency modulation. Here, the amount of the carrier frequency shift is proportional to both the amplitude and frequency of the modulating signal.

The phase of the carrier is changed by the change in amplitude of the modulating signal as shown in Figure 8 below.

Figure 8: Phase Modulation

The modulated carrier wave is lagging the carrier wave when the modulating frequency is positive. This can be clearly seen if we concentrate on the peak amplitude along line AB (previous slide). When the modulating frequency is negative, the modulated carrier wave is leading the carrier wave. This can be seen clearly by looking at the peak amplitude along line CD (previous slide).

Integrated Publishing
The Most Informative Site on the Internet
Electrical Engineering Training Series

Order this information in Print | Order this information on CD-ROM | Download in PDF Format

Click here to make tpub.com your Home Page

Page Title: Phase modulation
Back | Up | Next

tpub.com Updates

Google

Search
Web
www.tpub.com

Home

Information Categories
Administration
Advancement
Aerographer
Automotive
Aviation
Combat
Construction
Diving
Draftsman
Engineering
Electronics
Food and Cooking
Math
Medical
Music
Nuclear Fundamentals
Photography
Religion
USMC

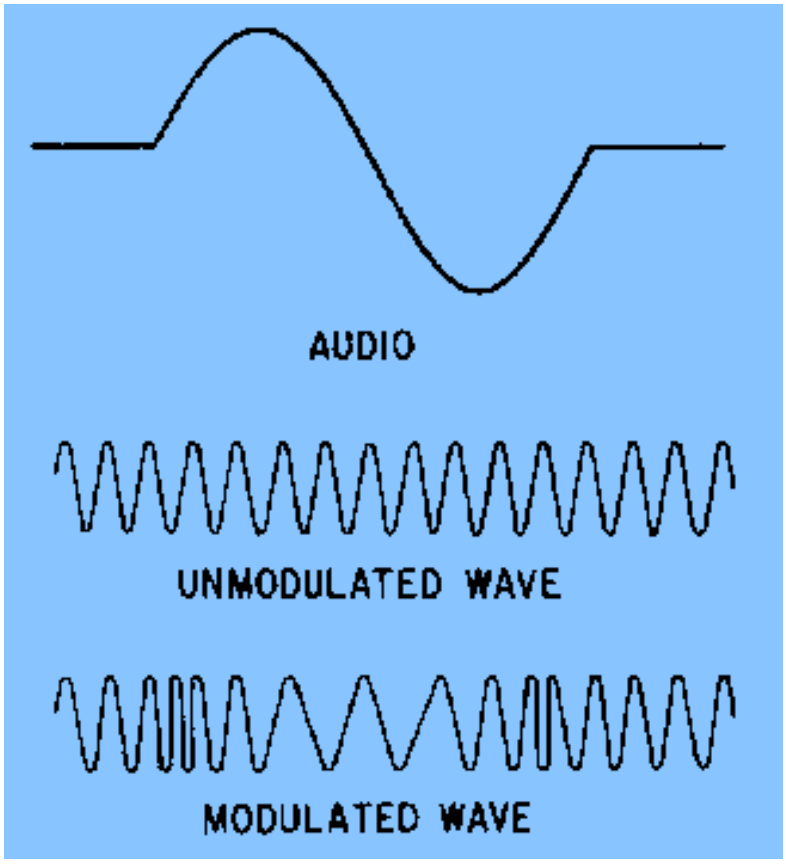
Products
Educational CD-ROM's
Printed Manuals
Downloadable Books

[Back] [Home] [Up] [Next]

PHASE MODULATION

Frequency modulation requires the oscillator frequency to deviate both above and below the carrier frequency. During the process of frequency modulation, the peaks of each successive cycle in the modulated waveform occur at times other than they would if the carrier were unmodulated. This is actually an incidental phase shift that takes place along with the frequency shift in fm. Just the opposite action takes place in phase modulation. The af signal is applied to a PHASE MODULATOR in pm. The resultant wave from the phase modulator shifts in phase, as illustrated in figure 2-17. Notice that the time period of each successive cycle varies in the modulated wave according to the audio-wave variation. Since frequency is a function of time period per cycle, we can see that such a phase shift in the carrier will cause its frequency to change. The frequency change in fm is vital, but in pm it is merely incidental. The amount of frequency change has nothing to do with the resultant modulated wave shape in pm. At this point the comparison of fm to pm may seem a little hazy, but it will clear up as we progress.

Figure 2-17. - Phase modulation.



Let's review some voltage phase relationships. Look at figure 2-18 and compare the three voltages (**A**, **B**, and **C**). Since voltage **A** begins its cycle and reaches its peak before voltage **B**, it is said to *lead* voltage **B**. Voltage **C**, on the other hand, *lags* voltage **B** by 30 degrees. In phase modulation the phase of the carrier is caused to shift at the rate of the af modulating signal. In figure 2-19, note that the unmodulated carrier has constant phase, amplitude, and frequency. The dotted wave shape represents the modulated carrier. Notice that the phase on the second peak leads the phase of the unmodulated carrier. On the third peak the shift is even greater; however, on-the fourth peak, the peaks begin to realign phase with each other. These relationships represent the effect of 1/2 cycle of an af modulating signal. On the negative alternation of the af intelligence, the phase of the carrier would lag and the peaks would occur at times later than they would in the unmodulated carrier.

Figure 2-18. - Phase relationships.

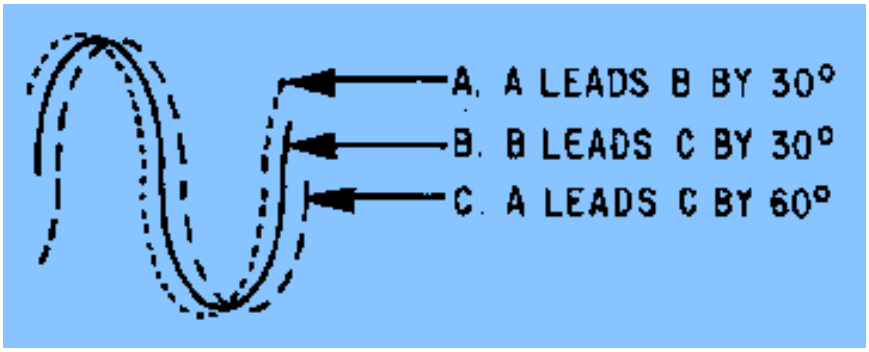
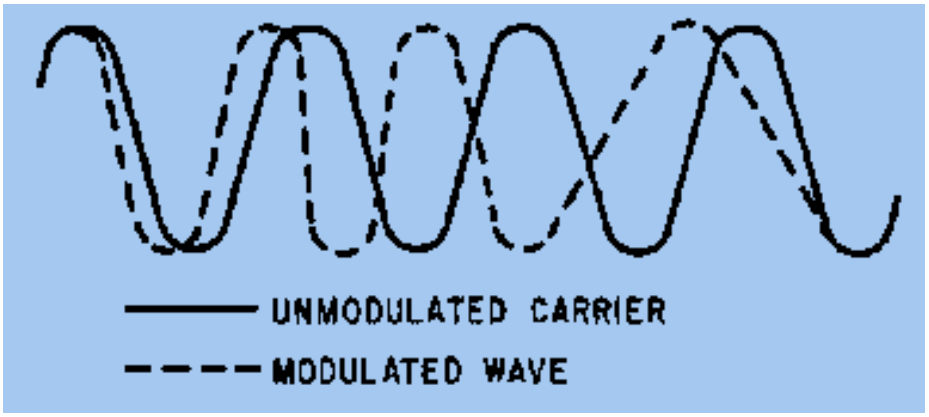


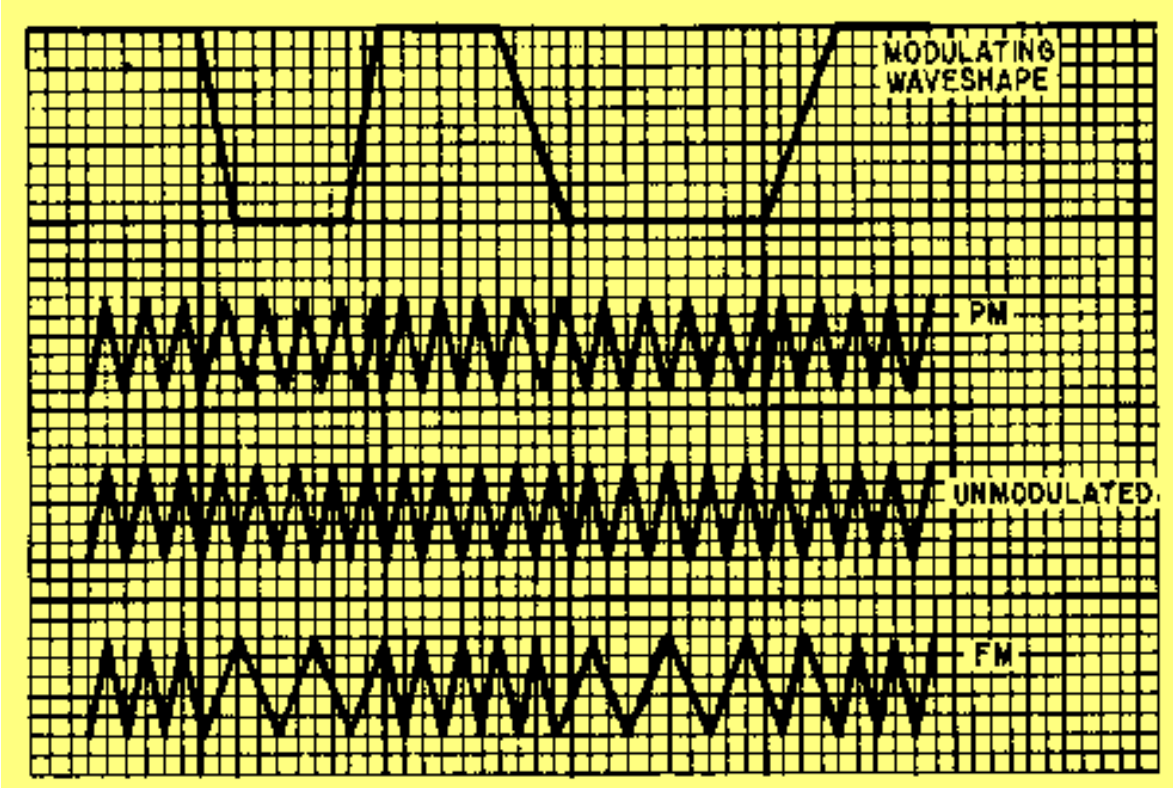
Figure 2-19. - Carrier with and without modulation.



The presentation of these two waves together does not mean that we transmit a modulated wave together with an unmodulated carrier. The two waveforms were drawn together only to show how a modulated wave looks when compared to an unmodulated wave.

Now that you have seen the phase and frequency shifts in both fm and pm, let's find out exactly how they differ. First, only the phase shift is important in pm. It is proportional to the af modulating signal. To visualize this relationship, refer to the wave shapes shown in figure 2-20. Study the composition of the fm and pm waves carefully as they are modulated with the modulating wave shape. Notice that in fm, the carrier frequency deviates when the modulating wave changes polarity. With each alternation of the modulating wave, the carrier advances or retards in frequency and remains at the new frequency for the duration of that cycle. In pm you can see that between one alternation and the next, the carrier phase must change, and the frequency shift that occurs does so *only* during the transition time; the frequency then returns to its normal rate. Note in the pm wave that the frequency shift occurs only when the modulating wave is changing polarity. The frequency during the constant amplitude portion of each alternation is the REST FREQUENCY.

Figure 2-20. - Pm versus fm.



The relationship, in pm, of the modulating af to the change in the phase shift is easy to see once you understand AM and fm principles. Again, we can establish two clear-cut rules of phase modulation:

AMOUNT OF PHASE SHIFT IS PROPORTIONAL TO THE AMPLITUDE OF THE MODULATING SIGNAL.

(If a 10-volt signal causes a phase shift of 20 degrees, then a 20-volt signal causes a phase shift of 40 degrees.)

RATE OF PHASE SHIFT IS PROPORTIONAL TO THE FREQUENCY OF THE MODULATING SIGNAL.

(If the carrier were modulated with a 1-kilohertz tone, the carrier would advance and retard in phase 1,000 times each second.)

Phase modulation is also similar to frequency modulation in the number of sidebands that exist within the modulated wave and the spacing between sidebands. Phase modulation will also produce an infinite number of sideband frequencies. The spacing between these sidebands will be equal to the frequency of the modulating signal. However, one factor is very different in phase modulation; that is, the distribution of power in pm sidebands is not similar to that in fm sidebands, as will be explained in the next section.

Modulation Index

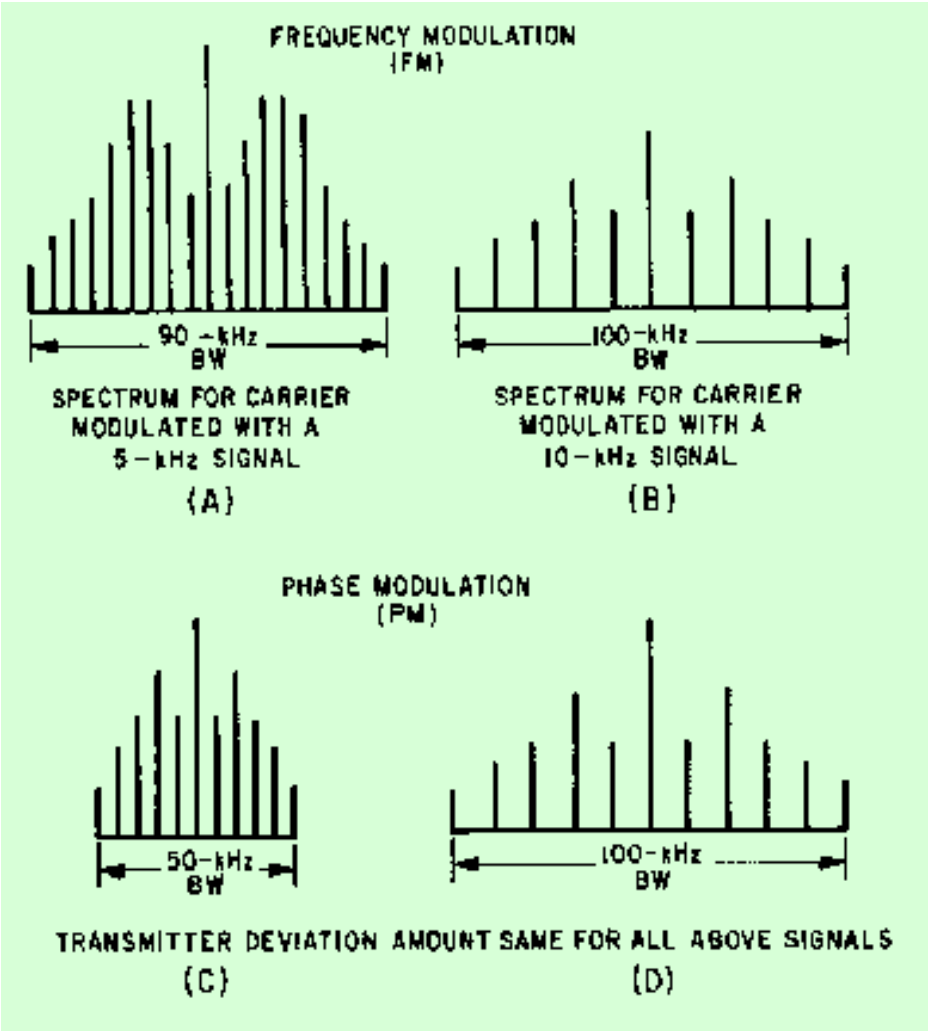
Recall from frequency modulation that modulation index is used to calculate the

number of significant sidebands existing in the waveform. The higher the modulation index, the greater the number of sideband pairs. The modulation index is the ratio between the amount of oscillator deviation and the frequency of the modulating signal:

$$MI = \frac{\text{transmitter deviation}}{\text{modulating frequency}}$$

In frequency modulation, we saw that as the frequency of the modulating signal increased (assuming the deviation remained constant) the number of significant sideband pairs decreased. This is shown in views (A) and (B) of figure 2-21. Notice that although the total number of significant sidebands decreases with a higher frequency-modulating signal, the sidebands spread out relative to each other; the total bandwidth increases.

Figure 2-21. - Fm versus pm spectrum distribution.



In phase modulation the oscillator does not deviate, and the power in the sidebands is a function of the amplitude of the modulating signal. Therefore, two signals, one at 5 kilohertz and the other at 10 kilohertz, used to modulate a carrier would have the same

sideband power distribution. However, the 10-kilohertz sidebands would be farther apart, as shown in views (C) and (D) of figure 2-21. When compared to fm, the bandwidth of the pm transmitted signal is greatly increased as the frequency of the modulating signal is increased.

[Back] [Home] [Up] [Next]

Order this information in Adobe PDF Printable Format

Privacy Statement - Press Release - Copyright Information. - Contact Us -

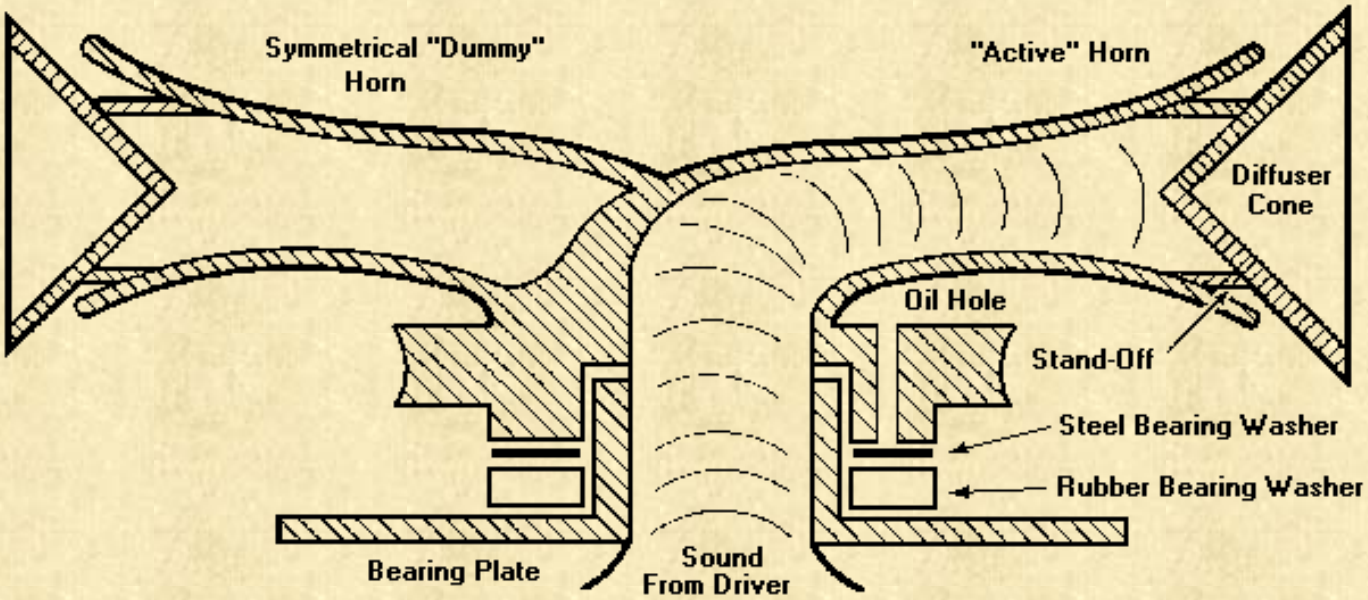


Figure 1. Plan-view of the Leslie Treble Rotor

UNEARTHING THE MYSTERIES OF THE LESLIE CABINET

by Clifford A. Henricksen
Community Light & Sound

First published in Recording Engineer/Producer magazine, April 1981.

The Leslie Rotating Tremolo Speaker System has produced the same unmistakable sound for so many musicians that it has become an international institution for the recording and performing musical arts. Everything from the venerable Hammond B-3 organ, to human voices and electric guitars, have been processed by the Leslie Sound, and have ended up on countless hit records. This article is an attempt to unearth the mysteries of its operating principles, and includes a discussion of the Leslie Cabinet's practical use in both stage and studio situations.

My love-hate relationship with Leslie Speaker began with the purchase of a Hammond organ during the mid-Sixties. The salesman told me I would probably end up buying a Leslie for it, a statement that I dismissed as the usual sales banter. Besides, at the time I had no idea what a Leslie was anyway. The Hammond was purchased because that's what the Animals' organist, Alan Price, used on all the group's early hits, especially "The House of the Rising Sun." I saw the band live in Schenectady, New Youk, where they put a Hammond M-102 direct into a Fender Bandmaster Amp, and I *had* to have this sound. I soon discovered, however, that my new musical heroes had *better* sound. It wasn't long before I knew that Booker T. Jones, Felix Cavaliere, Gary Brooker and

Billy Preston (and many others) were all using Leslies. *This* was The Sound!

My first Leslie, a single-speed Model 45 with a 40-watt tube amplifier sounded great, but it literally got lost in the roar of the Boston-based rock and roll band in which I was then playing. Two big Leslie Speakers were cumbersome to move, they weren't a big improvement in level (plus 3 dB), and the high-frequency drivers broke. From then on, it was guerilla speed shop tactics (mill the heads, oversized pistons, etc.), to the point where I could finally be heard over any guitar player. As a result, I became an expert at Leslie repair, modification and special use, occasionally even letting a few guitar notes into my beloved machines. I've spent the past 15 years in and out of the Music Business, both in studios and on stage. The significant part of this time was spent in some way dealing with Leslie Speaker Systems. As you might guess, this is a rock and roll story. I might as well show my bias here and now. With this in mind, here's what I have learned over the years.

PRINCIPLES OF OPERATION

First and foremost, the Leslie Rotating Speaker is designed as *a sound modification device*. It is not a "Hi-fi" speaker, but rather a part of a musical instrument. You buy a Leslie to *change* the sound of an instrument, not to reproduce it.

The Leslie Speaker System, thoughtfully named after its inventor, Don Leslie, operates on a simple principle: a directional sound source is rotated at constant (or variable) speed around a fixed pivot point.

At a listening point some distance from this whirling affair, three things happen. First, because the source is directional, the intensity of the sound will be at a maximum when it points at the listener (or microphone), The sound intensity will increase as the rotating source approaches dead center, and decrease as it rotates past this point. The resultant effect is called amplitude modulation (AM), which is no big thing, and is a feature on any guitar amp with a "vibrato" or "tremolo" feature. By moving closer to the rotating speaker, the inverse square law will increase the modulation effect.

However, the *big deal* is the Leslie's ability to create frequency modulation (FM). As the source rotates toward the listener, its relative velocity will increase the pitch of any tone it produces; as it rotates away, the pitch well be lowered. This is exactly the same Doppler effect that causes a train whistle (or any other sound on the train, such as grunting pigs, or shrieking passengers), to rise and then fall in pitch as the train approaches and then passes.

Lastly, if you are listening in a room with any significant reverberation, a complete spatial modulation of the sound will happen, as sound is "shot" all around and goes through multiple reflections.

If you want a real treat of all kinds of bizarre whirling horns and things, order copies of Don Leslie's original U.S. Patents: RE#23,323 and 2,622,693, available from The United States Patent Office, Washington, D.C.20231, for a fee of 50 cents per copy. These patent specifications are incredibly entertaining documents, and a must for any Leslie fan.

The Real-Life Leslie

A practical and commercial manifestation of the Leslie principle may take on many forms. Indeed, the manufacturer of Leslie Speakers, Electro Music, Inc. (now a division of Hammond Organ), produces a wide variety. These include models with reverberation, triple channels, and rotating-cone speakers. However, any Leslie afficionado will tell you that the Leslie Models 145, 147, and 122 are the ones with "The Sound". All three Models share the same basic innards: a 40-watt monophonic tube amplifier; an 800 Hz 16-ohm passive crossover; a rotating treble horn and a rotating bass speaker. Both rotating speakers are available with slow and fast A.C. induction motors. (Older Models 45, 47, and 22 are *identical*, except for having single-speed rotors.)

Furthermore, these Leslie models are similar in that they all have components mounted in a three-compartment cabinet. The top compartment houses the rotating high-frequency horn; the middle "box" the high- and low-frequency drivers and crossover (and which also acts as a vented box for the low-frequency driver); and the bottom compartment the low-frequency rotor and amplifier. Louvres located in the top and bottom compartments let out treble and bass sound respectively. All the above speaker systems are virtually identical in terms of their use and sound quality. (The larger Models 147 and 122 allegedly have a better low-end.) Both high- and low-frequency speakers operate on the same principle: a stationary driver (loudspeaker) and a rotating acoustic "projector".

The "Treble Rotor"

The Leslie's high-frequency unit is largely responsible for the "Leslie Sound". Some organists actually find that the bass rotor's slower response to speed changes is distracting, and will actually disconnect the bass rotor drive - especially when playing a bass line. The treble unit consists of a stationary 3/4-inch throat Jensen compression driver, connected to a vertical tube that acts as a thrust bearing (Figure 1). A twin-bell, molded black bakelite horn, which starts vertically but flares horizontally, sits on this bearing/tube, and rotates via a two-speed A.C. induction motor fitted with three (selectable) drive pulleys. This motor drives the treble horn at fast or slow speeds via drive belt and belt tension spring. Direct current is sometimes applied to the "fast" motor so that it will slow down more quickly; this is called a "D.C. brake". The treble horn actually looks like two horn assemblies; in fact, only one is operable as a horn. The other side, a "dummy", acts as a counterweight or dynamic balancer, providing symmetric air drag at high speed. The resulting structure rotates smoothly and without eccentric "wobble" forces.

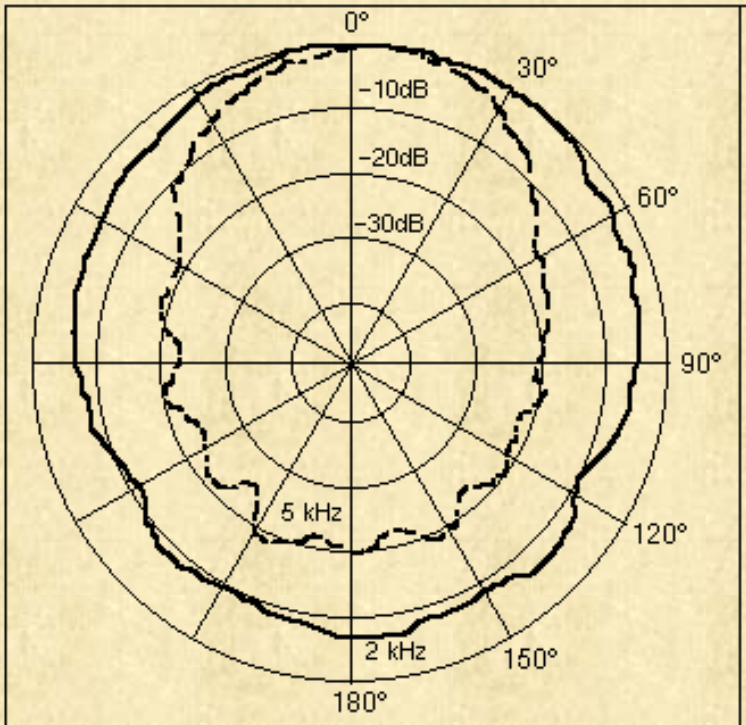


Figure 2. Typical polar response patterns of the Leslie Treble Rotor, with the deflectors removed.

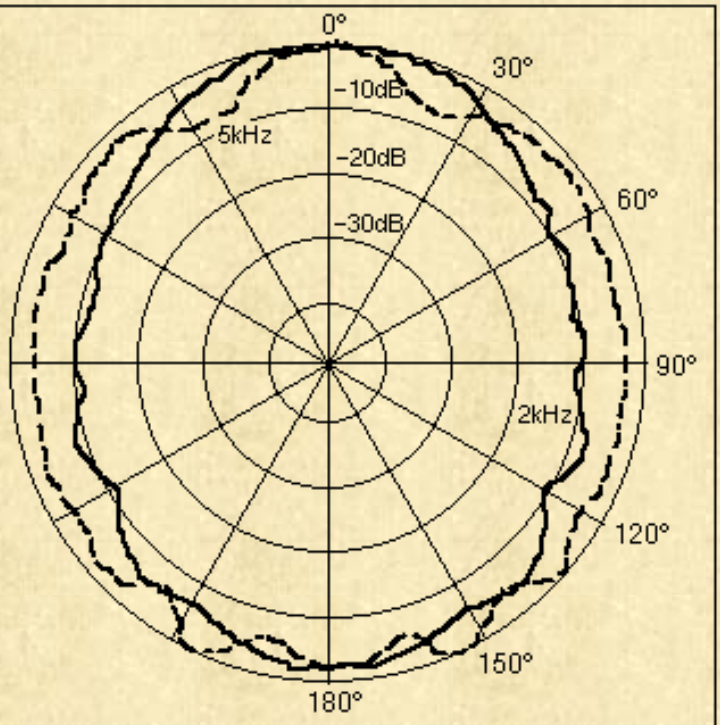
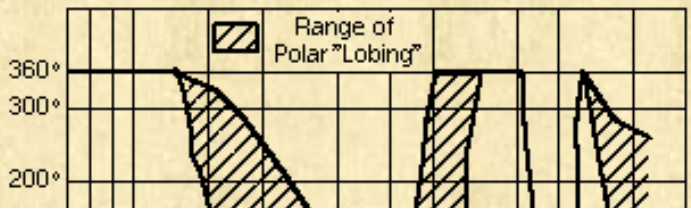


Figure 3. Typical polar response patterns of the Leslie Treble Rotor, with the deflectors in place.



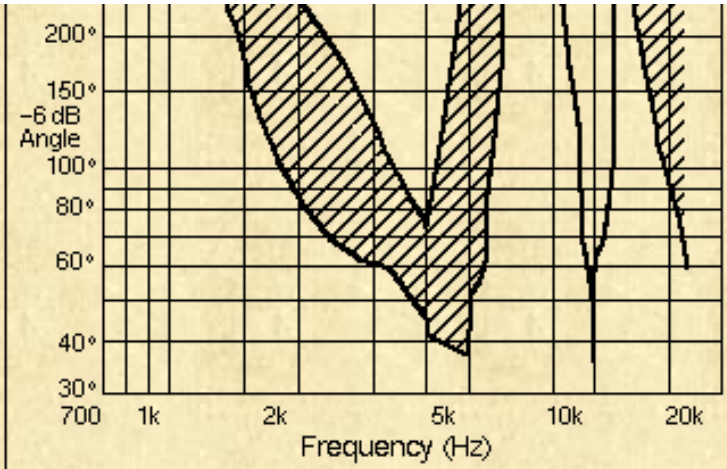


Figure 4. Coverage angle of the Leslie treble horn, with the deflectors in place.

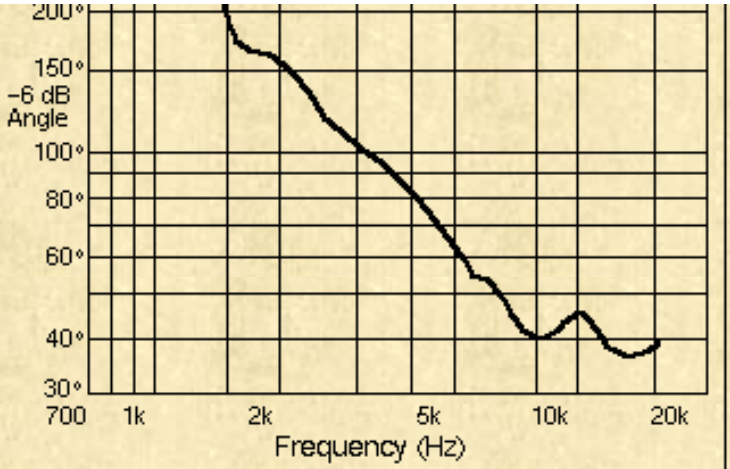
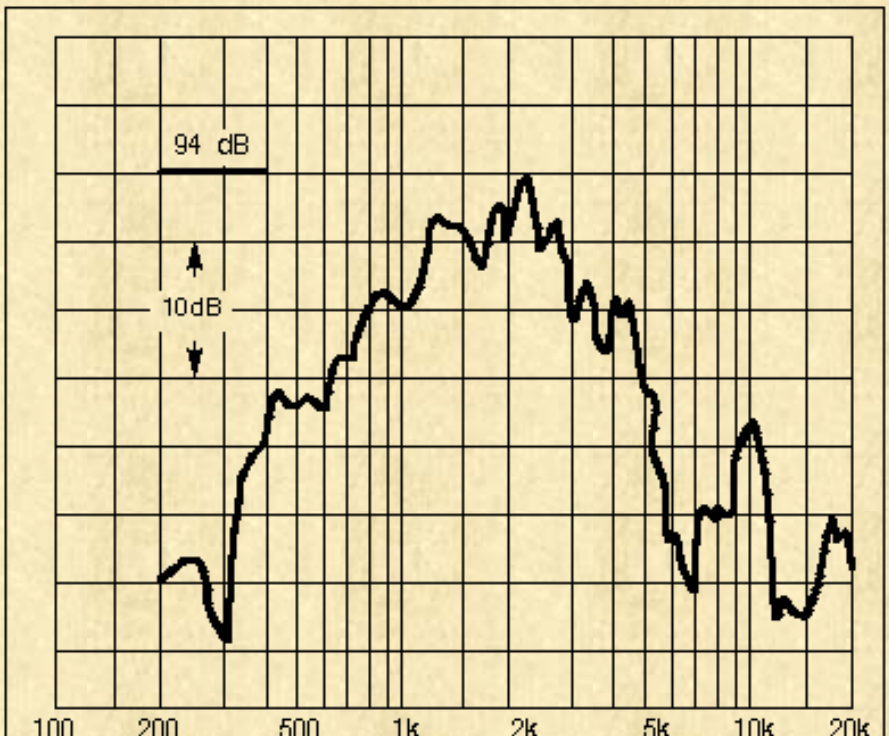


Figure 5. Coverage angle of the Leslie treble horn, with the deflectors removed.

At the mouth of the horn is diffuser cone, which is supposed to widen the dispersion of the horn and make a "more musically pleasing tone". Actually, it does work. Figures 2 and 3 show typical polar response charts of a Leslie Horn with and without the deflector; the results are dramatic. The "without" curve of Figure 2 is typical of a "beamy" straight horn, with a lot of sound concentrated on-axis and very little sound off-axis. The "with" polar response curve of Figure 3 is almost omnidirectional; note, however, the "lobing". As the horn revolves, the sound will actually rise and fall a number of times, giving it an even more characteristic sound.

The horn's so-called coverage angle is taken from polar response charts such as these; it is the included angle between -6dB points from the horn's on-axis level. Figure 4 and 5 show such coverage angles, compiled from a series of third-octave polar response curves. Note how the beamwidth gets smaller and smaller without the deflector. This means that if the deflectors are removed the sound will be much hotter on-axis, but much softer off-axis. Therefore, the AM portion of the Leslie Sound will be much more severe and "choppy", as many a rock and roller will tell you. Many Leslies played in clubs are faced backwards, with the top and bottom rear panels and deflectors removed, because the sound "cuts" better; an effect that translates as, "You can hear it better over the guitar player".

This beaming process will actually raise the on-axis sensitivity of any driver, as can be seen from the frequency-response curves of a University ID-40 driver (a respectable, but typical 3/4-inch throat PA unit) mounted on a Leslie horn with and without deflectors (Figures 6 and 7). It is interesting to note the extra sensitivity above 1.5 kHz available without the reflectors in place. (Note also that the low-end - 800 Hz to 1.5 kHz - response isn't affected by removing the reflector, and that the response is ± 5 dB from 400 Hz to 10 kHz. This latter range could be made even flatter with a real-time analyzer and an third-octave equalizer in the studio.)



Now, this entire discourse is based on the treble horn measured in an anechoic or non-

reflecting, highly absorbent room. In real life, the horn mounted in a louvred wooden box, which means that the picture changes somewhat with resulting internal reflections. The diffuser cone does another important thing: it shifts the apparent sound source position on the horn. With the cone *in*, more of the entire range of the horn will appear to come from the mouth of the horn; with it out, however, while lower frequencies still appear to come from the mouth, higher and higher frequencies will appear to come from progressively further down the throat. They therefore appear to be rotating at a *smaller* radius, which results in less frequency modulation effects. (If the directional sound source was rotating at dead center, there would be no FM effect at all.) So here you are faced with a choice: leave the deflectors in place and the result will be maximum frequency modulation, and a relatively lower amplitude modulation, because of the very wide directional characteristics of the horn. Take the deflectors out though, and the FM will be lost, but you obtain very strong AM - especially at higher frequencies - due to the very narrow beam width of the naked horn. My own ears tell me that the deflectors should be kept on, but you make your choice and you take your chances!

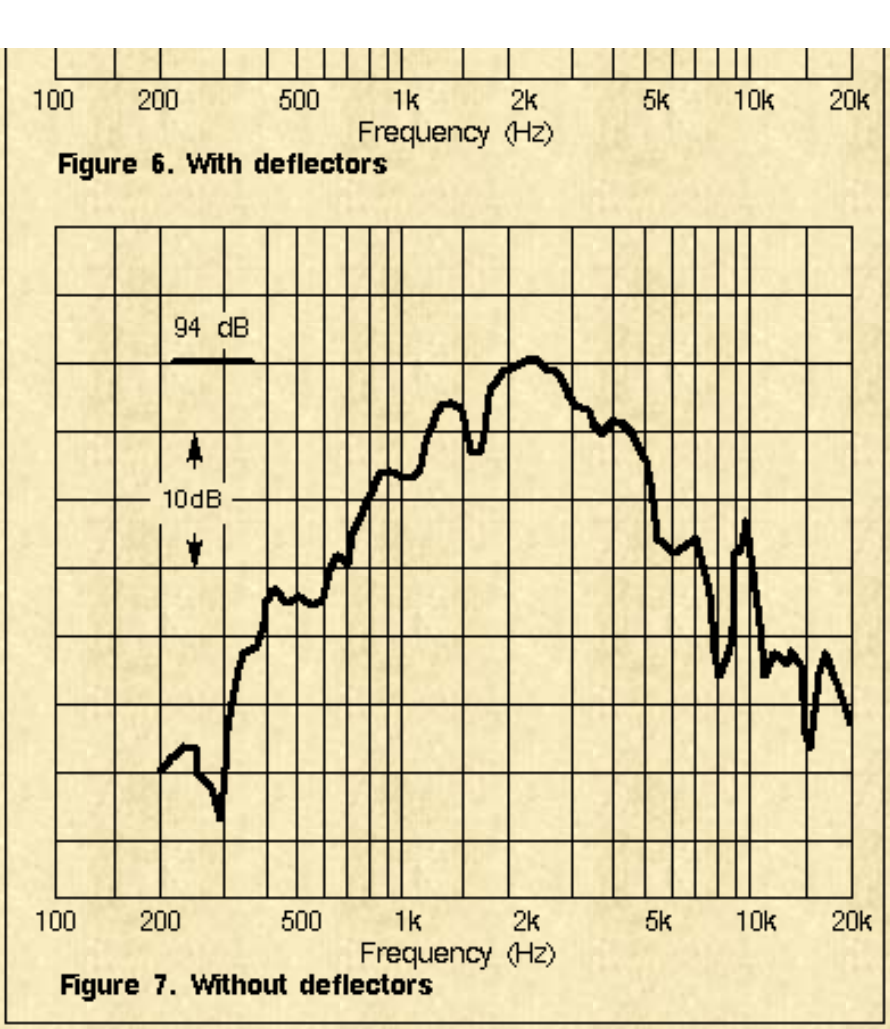


Figure 6. With deflectors

Figure 7. Without deflectors

The "Bass Rotor"

In the lower compartment of the Leslie box is a rotating wooden drum mounted on a vertical shaft, and covered with a black scrim cloth, which provides lower aerodynamic drag on the drum at fast speed (Figure 8). It also works. At the drum's center is a cylinder fitted with a "scoop" which, as in the treble unit, starts vertically (the bass driver faces downward into its entrance) and projects sound horizontally. The drum assembly is driven with a two-speed motor, and ends up at approximately the same rotational speed as the treble unit. The only difference is that the drum's inertia makes it approach final speed over a much longer time period. A D.C. braking voltage can also be applied to the "fast" motor to slow down the speed more quickly when switching from fast to slow. This system works somewhat effectively.

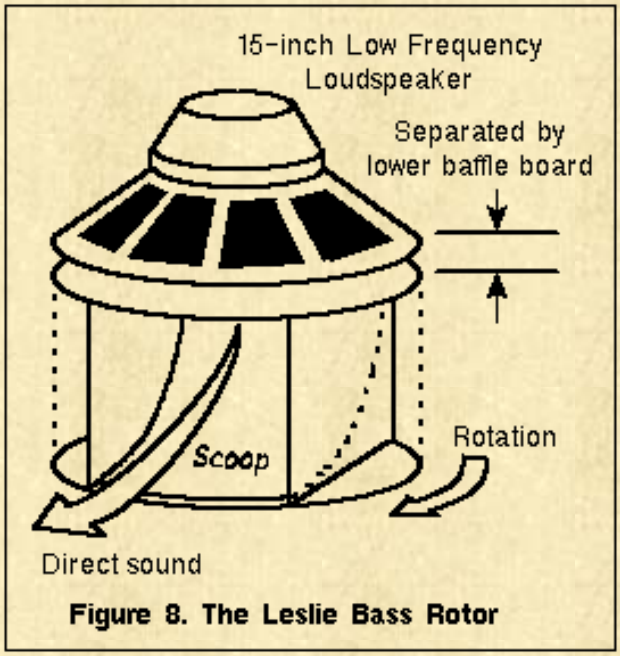


Figure 8. The Leslie Bass Rotor

To my ears and powers of reasoning, the bass rotor works just as an AM device, and only for the upper two octaves or so of the bass section (200 to 800 Hz). Frequencies lower than 200 Hz are probably uneffected by a scoop of this size, since 200 Hz has a wavelength of approximately 5.5 feet. There may be some frequency-modulation effects near the 800 Hz xrossover point, but it sure sounds like AM to me.

The result is a low-frequency "throb", which is very pleasant and especially powerful and beautiful when used in "slow" or "chorus" mode.

The Complete Leslie System

With all these rotating components installed in the box, the System works as follows: an electical signal is sent into the Leslie amplifier driving a 12 dB per octave, 16-ohm crossover, which feeds the 16-ohm bass and treble drivers with the appropriate frequency bands (Figure 9). Input signal, motor controls, and A.C. line voltage (115 V, 60 Hz) are connected to the unit via special plugs and sockets, linked with the "Leslie cable". It is this latter cable that's a major problem for many who would like to use a Leslie. Most self-contained speaker systems have an A.C. cord that plugs into the wall, and an input jack for the music signal. The Leslie, however, has a "Leslie Cord". Just one of the things with which we Leslie fans have to live.

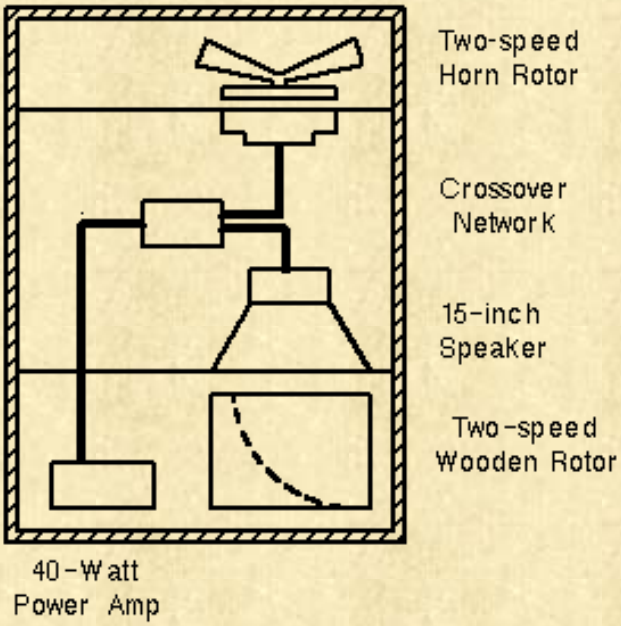


Figure 9. Schematic diagram of the Model 147 Leslie Speaker System

The amplifier chassis contains a 40-watt amp fitted with 6550 tubes - a *wonderful* choice of output tube for instruments - and motor control circuitry. Each rotor is actually driven by two separate motors (fast and slow) mounted in one package. Therefore, there are a total of four rotor motors and four pairs of wires, which all plug into the amplifier chassis. If any of these wires are removed and plugged into a live A.C. outlet, the appropriate motor will turn on safely without any problem, since this is essentially what the motor control does.

Input to the amplifier is via the six-pin plug. The speaker and crossover are both connected by very odd little two-pin connectors that I have never seen in use outside of a Leslie Speaker. They work, however, and seem to be very reliable over long periods of time.

Getting Music In And Out Of A Stock Leslie

Picture the scene. You have just bought a used Leslie; you got a great deal on it just in time for the gig. You excitedly load it up, set up your Farfisa organ and then stare dubfounded at this ugly, black 6-pin Harvey-Hubble *plug*, the only connector visible on your Leslie. Your 1/4-inch phone plug cord hangs limply in your hand... useless... impotent. Mistress Leslie stares back coldly. You can't even plug it in the wall and watch the tubes glow!

Chorus Interruptus... The Horror!

Well, it's a real problem. But, *veehaff vays to do ziss, Fraulein Leslie*. Here's how.

If your studio has a Hammond B3 or C3 organ with a Leslie attached, the latter will probably be a Model 22 or 122. This is a quiet, reliable and troublefree system, although you still can't plug it into the wall and operate the Leslie *independently of the organ*. Most big Hammonds, like the B3 and C3, will have a Leslie plug installed on them. Be careful! Only connect a Hammond Organ to a Model 122, 22, or 122RV; never connect a Hammond via a "Leslie Cord" to a Model 147 ot 145 Leslie. It will basically blow up. Inside the organ a very simple connection can be made via the RCA phono jack fitted to the "expression control box". This jack can be located by taking the back off the organ and merely plugging in. It is a line-level input that will accept most relatively high

impedance signals.

A recent welcome accessory to the Leslie line is the Leslie Combo Preamp. These little beauties actually booost the signal of an instrument to the required level, provide A.C. power for Leslie's amplifier and motors, and have foot switches to change speeds. It also plugs into the wall (via a *real* A.C. cord), connects easily to most instruments (a *real* 1/4-inch phone jack), and attaches the preamp to the Leslie via the standard Leslie Cord.

You may want to build a customized system for your own special use from a stock Leslie 147 or 145. (You'll *probably* want to modify it - this comes later.) Any modifications that involve poking around within the power amplifier should be carried out with *extreme* caution. Remember that the Leslie speaker is fitted with a *tube* amplifier, which has a power supply voltage of over 400 VDC. I've heard that, if you're a player, such a shock can improve 128th note runs, but I wouldn't recommend it. Watch where you put your fingers!

Engineering your own system is actually a simple matter, once you can locate a 6-pin plug that is compatible with the one fitted to the amplifier (or buy a Leslie Cord). Pin 1 is signal ground, and pin 6 signal "hot". Any source capable of delivering over 6 V into a load of 2.5 kohm, or less, will be able to drive the Leslie amp to full power with the "console load resistor switch" set to "open". The switch also inserts an 8-ohm, 10 watt or 16-ohm 10 watt equivalent resistor across the input, which means that a 10-watt power amp could be used to drive the Leslie, with the level control at the right setting. This is the typical setup for many organs equipped with their own power amps. A small guitar amplifier, such as a Fender Champ would do fine. The speaker should be disconnected, and the switch set to "8 ohms". Don't try it with a large power amp, however; your Leslie will sound, and smell, very bad shortly after the input resistors burn up.

Input power (115 VAC) connects via pins 3 and 4. Pins 2 and 5 connect the 115-Volt A.C. coil of the tremolo relay, which is "fast" when off, and "slow" when on. Therefore, a D.P.S.T. switch connecting pin 2 and pin 3, and pin 4 to pin 5 will activate the relay; if you only want "fast" forget pins 2 and 5. (A simpler approach would be to make a direct connection from pins 4 and 5, and use a simple S.P.D.T. switch.)

Modifications To The Leslie

Ah, yes. Audio racing in the streets. Imagine thousands hearing a vocal track through a Leslie... a Leslie that can be heard over a guitar player's Marshall Stack... a clean, Hi-Fidelity Leslie. These things and more can be yours, given the cash outlay in proportion to the desired grandeur, and even more cash if you don't build it yourself. (This country is loaded with local Leslie "speed shops" - and there will probably be 50% more after this article is published.) Electro Music makes Fenderesque Bi-Amp Leslies in vinyl cabinets. However, they are expensive, and are only equipped with amplifiers rated to a maximum of 60-watts. Also, you may very well want to preserve the present walnut beauty and improve its performance. If so, here's how:

External Mono-Amplification

The stock Leslie Models 145, 147 or 122 have a marvelous 40-watt tube amp. When you *really* lean on it, it sounds even lovelier, with a warm rich distortion like nothing you've heard. The show organist at Circus-Circus, Las Vegas has the best sound of this kind I've ever heard. He runs his Hammond B3 on 10, and just keeps the hammer down. However, in a live band situation or a club, or worse, at a concert, even a single 40-watt Leslie just doesn't cut it, even for a keyboard only. The recent rapid expansion of concert sound techniques allows quiter instruments to be miked and simply turned up at the console. I'm told that Keith Emerson and others have actually installed and miked a Leslie in an isolated backstage room during their concerts. (Including Bruce Springsteen's keyboard player, Danny Federici; *R-e/p, February, 1981 - ed.*) The result is a clean Leslie sound with no on-stage bleed from other instruments. In many cases however, a performer prefers more direct level from their own instrument.

The simplest way of enhancing the sound of a Leslie is to disable the built-in amplifier (pull out all four tubes and/

or the fuse) and use the motors only, via a combo preamp or the adapter discussed earlier. The speaker system can then be connected to whatever amplifier you plan to use. This is achieved by disconnecting the crossover from the standard Leslie amplifier, and patching it to a convenient input socket. A 1/4-inch phone jack, or a much-better-but-less standard banana or GR plug would be the answer. A more versatile method would be a 1/4-inch jack which, when accessed, disconnected the crossover input from the Leslie amp, and connected it instead to the jack. The only problem remaining is the power handling capablility of the Leslie Speaker System components, namely the high- and low-frequency drivers and the crossover networks. I'll discuss these as separate modifications.

Treble Driver Modifications

The stock Leslie high-frequency driver is pretty fragile, and is easily overloaded with a stock 40-watt amplifier. Unless for some exotic reason you want to drive it at low levels, plan on a new driver. Choice of drivers is important, and the most expensive ones are not necessarily the best answer. The most expensive professional drivers from Altec, JBL, Electro-Voice, and others, have usable sound reproduction to within the 10 to 20 kHz octave. Such performance probably won't be needed for a Leslie - I'm told that, in fact, a very high frequency harmonic content is actually a deterrent.

Extended-band high-frequency drivers are also rather fragile; a percussion note from a Hammond B3 can easily send a 0.002-inch thick aluminum diaphragm through its full excursion, and into the driver's phase plug. At this point, the musician is off the air, possibly during a now favorite solo, or in the middle of a take. Also, these drivers have large throats (between 1.4 and 2.0 inches in diameter), whereas the entrance hole to a Leslie treble horn is about 3/4-inch. This makes it ideally suited for the 3/4-inch throat PA drivers fitted with screw mounts, such as those made by University, Electro-Voice, Atlas, etc..

Such drivers are relatively cheap, can easily reproduce up to 5 or 6 kHz, and some handle enormous amounts of power. They also have a very similar bandwidth to the stock Leslie driver, which means that the same sound quality will be maintained. Actually, replacing the treble driver would make a stock Leslie much more reliable for "leaned on" use. My own personal preference is the Electro-Voice 1829 driver, a 16-ohm unit ideally suited to the Leslie crossover. I used one on stage for years at full chat with a Bogan MO-

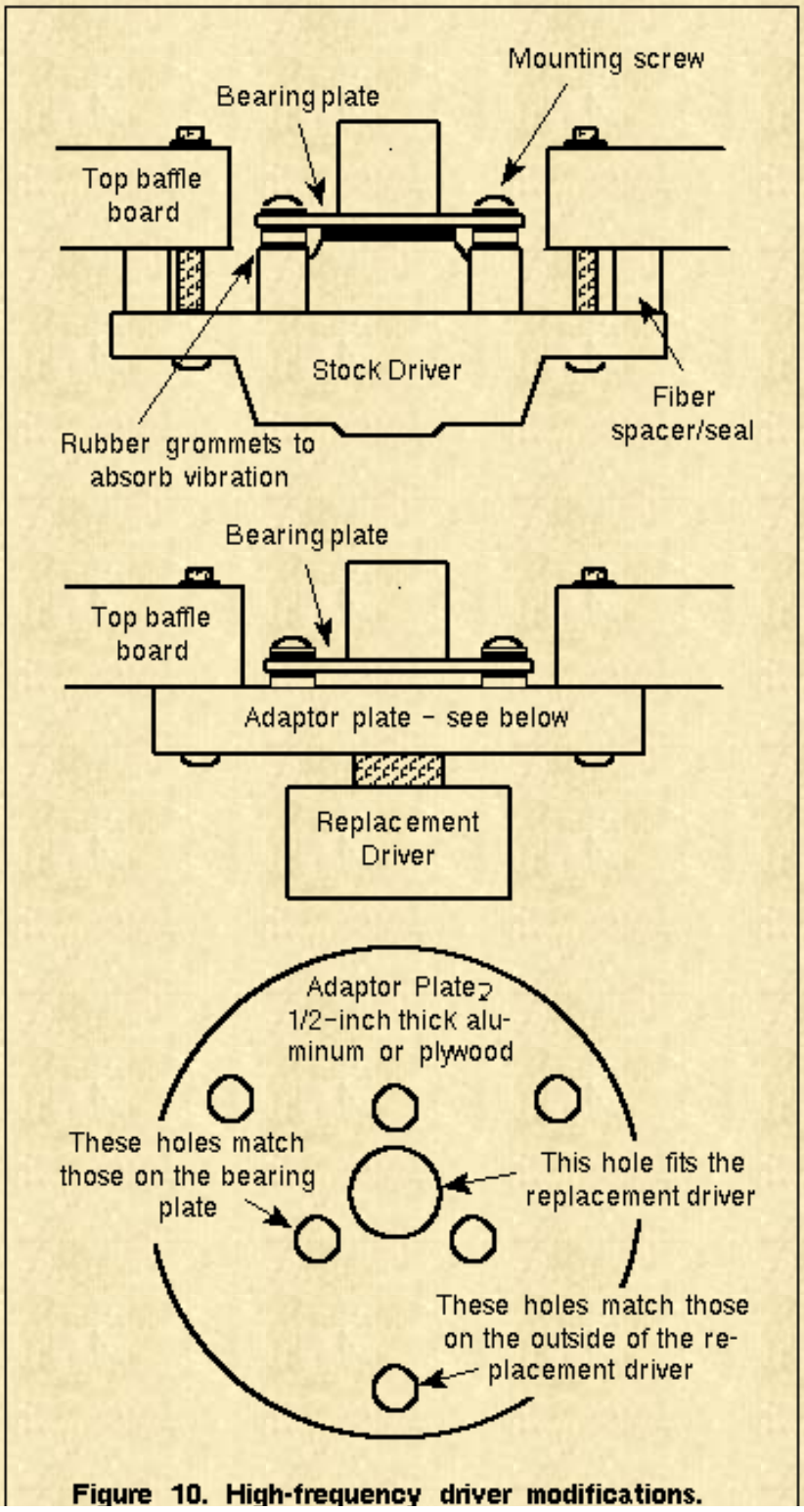


Figure 10. High-frequency driver modifications.

100A 100-watt tube power amp. It's still intact and has a pleasing sound. I'm told by numerous others that the have had similar good luck with an E-V 1829's sturdiness and power-handling ability.

Figure 10. High-frequency driver modifications. Top: standard driver and adaptor plate. Center: changes necessary to accommodate replacement driver. Bottom: layout of special adaptor plate.

Replacement of the Leslie driver involves, first, removing the backs from the upper and middle cabinets to provide access to the driver and horn. A special adapter is necessary to fit the screw throat PA driver to the stock Leslie horn, as shown in Figure 10.

Bass Driver Modification

The stock Leslie Bass Driver is a 15-inch O.E.M. speaker capable of handling about 50 watts. Supply it with 40 watts of a 30 Hz pedal tone and it shakes a lot. You may also be getting dangerously close to the driver's excursion limit, but it seems to be pretty sturdy. I've certainly heard of them breaking with "stock" use. If only a keyboard is used through the system - one without a lot of deep bass - and you protect the driver by filtering out the low-end, the unit may handle lots more as you approach its "thermal limit". If you want to avoid running the risk of breaking the bass driver, the easiest solution is to replace it with another unit capable of handling higher loads. Speaker choice is left up to the individual; since it's a matter of taste I can't really recommend any. An 18-inch speaker will fit with the following modifications: a spacer ring has to be added between the speaker and the Leslie baffle (so that the cone doesn't hit it), and part of the back panel may have to be whittled out for certain larger 18-inch speaker frames.

Crossover Modification

The stock Leslie crossover is a 12 dB per octave, 800 Hz unit requiring both 16-ohm high- and low-frequency drivers for correct performance. Most "mondo-power" 15-inch woofers have an impedance of 8 ohms, which tends to extend the reponse a bit higher than 800 Hz. Although this shouldn't cause any difficulties, it will make an 8-ohm high-frequency driver extend a bit lower, which will definitely cause excursion problems. The answer is to either buy a 16-ohm replacement driver, or perform the following modifications.

If the Leslie is fitted with both 8-ohm high- and low-frequency components, a stock 800 Hz crossover can be obtained from JBL, Altec, and other manufacturers. Community Light & Sound makes a very good crossover, which requires an 8-ohm bass speaker and a 16-ohm treble speaker (if you use it without its attenuation pad). The crossover handles well over 200 watts RMS, and is made from aircore coils and Mylar capacitors. Besides being a true "hi-fi" crossover, it would make an excellent choice for a super-power system. It is also possible to design and build your own crossover network from drawings and circuits published in any of the audio textbooks and cook books.

The best, most versatile and most adjustable solution would be to biamplify the Leslie. Again, it's "art", right? A Crown VFX-2 and stereo power amp would be a great setup with which to experiment.

Lastly, a crossover network consisting of a single series capacitor is a possibility (20 microfarad for 8-ohm units, and 10 microfarad for 16-ohm). It allows full-range sound to pass to the bass speaker, and high frequencies above 1 kHz to the treble speaker with a 6 dB per octave rolloff. Such a crossover network has been used successfully, and produces a very interesting and pleasing sound.

Microphone-Prepared Stock Leslie

This technique has been used very successfully in a live-music situation. A pair of small electret lavalier microphoes are permanently fastened to opposite corners of the treble horn compartment. Output from these mikes is connected to an external stereo mixer equipped with equalization - hopefully by a permanent connector fitted to the back of the Leslie cabinet - and hence to a stereo power amp and a pair of speakers located on either side of the stage. If a stereo PA rig is being used, the mikes can be fed into two channels of the front-of-

house board and panned left and right. As with anything, "good taste" in EQ and general technique is required to achieve a good sound. Actually, bad taste might work well, too; try both.

Wind noise from the horn is not a problem, and minimum mechanical isolation is needed. Lack of wind noise is probably because the cabinet corners are "stagnation" or dead-air spaces. Wind noise, if present, can be reduced by using an open-cell (reticulated Scottfoam) wind screen on the mikes.

An obvious extension of this mike technique is to do the same with the lower rotor, and obtain a mixed top and bottom left and right sound. Apparently, the "top-only" technique works very well in a live situation, but might also serve the purpose for a studio Leslie.

Tandem Rotors

Slower acceleration of the lower rotor is occasionally undesirable - certainly, a matter of personal preference only. To give a more dramatic Leslie sound, the top and bottom rotors can be mechanically linked so that both will turn in sync, and the pair of horns face in the same direction. Silver Sound Systems (337 South Morris Avenue, Crum Lynn, Pennsylvania) built the little screamer with tandem rotors shown in Figure 11, which is an example of a well-prepared custom Leslie unit. It enhances the Leslie action by putting full range sound into the bass unit and only high frequencies to the treble unit. This customized unit was designed for use with electric guitars.

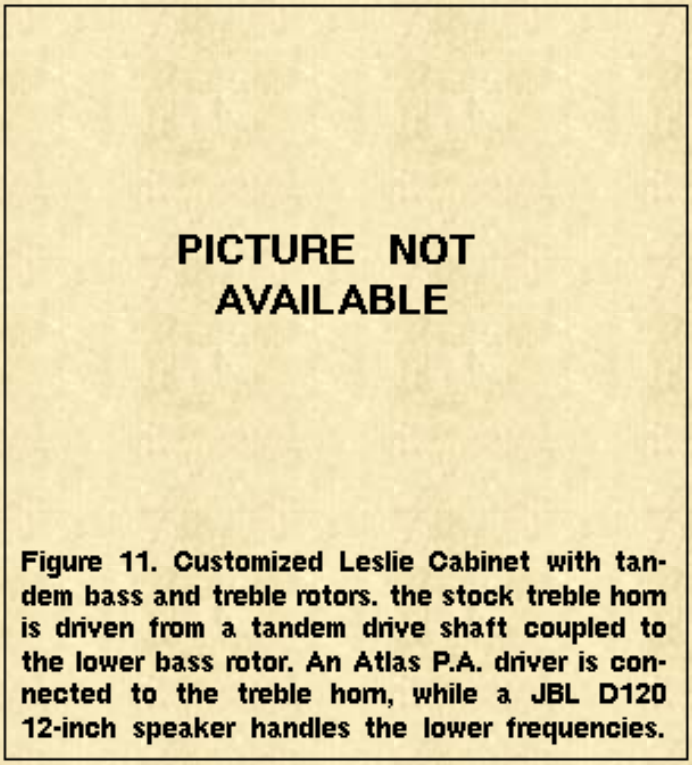


Figure 11. Customized Leslie Cabinet with tandem bass and treble rotors. the stock treble hom is driven from a tandem drive shaft coupled to the lower bass rotor. An Atlas P.A. driver is connected to the treble hom, while a JBL D120 12-inch speaker handles the lower frequencies.

Motor Speed Change

A really effective means of changing the rotor speed is to reach into the treble motors, and move the belt from one of the three drive pulleys to another. (Tricky, eh?) Actually, that's what they are there for; you are *supposed* to do it. My own preference is the middle pulley.

There have been a variety of speed change controls made available for Leslie motors. None have been very successful, however, since they just lower motor voltage, with the result that the motor will usually draw more current and possibly burn out. Watch out for these things. Leslie motors are A.C. induction types and, since they have no brushes, will last practically forever. However, their speed is determined by the 60 Hz frequency input, which is the same problem as varying tape-drive speed. The only solution that makes any sense is to hook-up a 115-Volt variable-speed oscillator. This can be constructed from a generator and tube amplifier that has a 115-Volt output tap. (Just like the old days, Sonny.) I am told it is really not worth the trouble though - the intermediate speed doesn't really sound very good.

STUDIO PRACTICE

As with just about anything, recording a Leslie is a matter of particular taste and purpose. You might want record the best possible recreation of a live Hammond B3/Leslie combination; you might also want to make it sound "better". You migh also want to use it as send and receive device, like an echo unit, with a high level send to the Leslie Amp, and a mono or stereo microphone receive. You could put it in a studio, in a reverb room (or in the bathroom) during the mix, or you might use it live. The possibilities are only limited by your level of creativity and/or insanity. However, the following are some standard techniques and descriptions of behavior.

Mono Recordings

I have recorded Leslies with a single microphone mounted as close as 1 foot from the treble horn, with very pleasing results. If the mike is placed even closer (within several inches) the severe amplitude-modulation effect becomes very annoying, and wind noise from the rotor at high speeds sounds like you have a helicopter in teh room. Which is a sound you might want. The same thing happens on the bass rotor, where up close there is also a lot of mechanical noise. A loose scrim will be real noisy; tighten it up. The sound from both rotors is mellower coming out of the louvres. By removing the back panels and miking the Leslie from the rear, the sound is more "direct" and defined.

Stereo And Multi-Mike Recording

Use of two mikes on the top and bottom of the cabinet is a very effective way of getting a good sound, bearing in mind the AM effects of close-miking. The *best* way I know of recording a Leslie, however, is in stereo. The left and right channels can be recorded with either a top and bottom pair of mixed mikes, or with just single mikes panned between left and right. The stereo image achieved with two pairs panned full left and right is very exciting. Many combinations are possible though, the point being that a Leslie is capable of providng a great deal of spatial information.

I spoke with a few engineers I know about such recording techniques, and here are a few of their preferences:

Jay Mark *(Sigma sound, New York)* has been pleased with a "tight" or "direct" Leslie sound, when recording organ. His principle reason for this is because the organ is used as background and not a featured instrument; the tight sound is needed so that the organ sound is very clear and unmistakable, even when way down in a mix. Jay has used the folloing setup with good results: an RCA 77DX ribbon at back of the top rotor cabinet, about 8 inches from the treble horn, with the high-end rolled of to suit. He also uses, at the back of bottom rotor, a U87 with the lows rolled off, and mixed with the top mike to suit. He remembers experimenting with the top and bottom sound panned left and right, and not liking the effect.

Allen Sides *(Ocean Way Recording, Hollywood)* prefers a pair of tube U67s, located about 5 feet away from the Leslie, aimed midway between top and bottom rotors, and spaced 10 feet apart. The mikes are panned left and right, and recorded on two separate tracks. Allen prefers the U67's natural roll-off on the high-end for de-emphasizing the high-frequency distortion on top. He told me of recording Billy Preston, who played a Fender Rhodes electric piano with stereo vibrato, and sent each channel to two separate Leslies.

Joe McSorely *(Veritable Recording, Ardmore, Pennsylvania)* likes to use a pair of U87s mounted relatively close in top and bottom cabinets. For a "tight" organ sound, he rolls of the lows from the bottom rotor, but records the top flat. Joe echoes a repeating problem - wind noise up close - and he always uses windscreens on the mikes. He says that most organ Leslie recordings done at Veritable are on one track, but a great "fake stereo" mixdown effect can be achieved by panning the dry track to one side, and using a Harmonizer in the "doubling" mode on a second track panned full opposite. Joe describes the resultant sound as "monstrous".

Several Odd Things To Try

An interesting sound results from running a full-range 15-inch speaker with a full range signal into just the bottom rotor. It should be obvious how to do this. The slow acceleration and deceleration of this massive rotor is quite different from the top unit, as is the frequency content. Try it.

There's one other item. Iknow for a fact that certain old Leslie Cabinets have a structural resonance in the wood and probably in the louvres, but it's the real thing. The old wood is simply more resonant, and Hammond percussion notes really set it off. (It's like these guitar players running around in South Carolina pawn shops looking for the ultimate '56 Fender Strat.) I've heard Leslies on records that just have a legendary tone to them,

and others that "just play regular". Go around and knock on a few new and used Leslies - especially the louvres - and see if you think I'm crazy. Anyway, the point to this was, why not try putting a FRAP pick-up on the louvres. Experiment - see if you can bring more of the sound out of the wood.

Specific Examples Of Normal Leslie Use

Normal Leslies are surgically attached to Hammond B3 or C3 organs; usually the former. There are many other brands of organs that have been played through Leslies during recording sessions, but "The Sound" which accompanies this great rock and roll Spirit of America is, pure and simple, a B3 and a Leslie. (Steinway piano, Fender Amp, Gibson, Lesl Paul, Fender Strat/Telecaster/precision Bass - all in the same breath.) The Isley Brothers used the fast Leslie all the way through their hit "Shout". Matthew Fisher combined America Soul, European Bach and B3 lower drawbars through a continuous slow chorus Leslie, and made millions listen to a "Whiter Shade of Pale" with its gorgeous organ theme. Billy Preston, Booker T. Jones and Felix Cavaliere made an art of timely speeding up and slowing down of the Leslie rotor.

These great artists (and many more) pioneered holding single notes (usually the root of the chord) and letting the Leslie provide the dramatics in rotor acceleration. Synthesizer players today *still* haven't found something as exciting. Lee Michaels had hits and toured for a long time with only his B3/Leslie and drummer Frosty. The right players could make a Leslie talk, make it scream, and make you want to dance. The list goes on forever.

...And Deviant Leslie Use (And Abuse)

Steppenwolf organist Goldie McJohn was able to coax an amazing sound from a stock Leslie. He connected the stock bass speaker directly ("full range") to the stock amp and ran the thing flat on 10 (actually it sounds like he found a way to get "11"). You can hear this very clearly on that classic track, "Born To Be Wild". The organ solo features the glorious distortion of a grossly overdriven Leslie amp, and you can hear the big bass rotor slowly speed up and slow down throughout the take. I saw Steppenwolf live at the now physically non-existent Electric Factory in Philadelphia. Goldie used the same setup on stage, only he stacked a bunch of them up right next to him, and miked one through the PA system.

Jethro Tull's *Benefit* album is considered by some to be the prototype Leslie-processed statement, and the hit "Teacher" from the LP is typical. Ian Anderson put both his flute and vocal through a Leslie, as did guitarist Martin Barre. It sounds like the Leslie was kept far away from distortion, a good example of a clean Leslie processing. I wish I know more about the particulars of this record, since it's another classic heavy.

Guitars Through Leslies

Many guitarists have used Leslies to modify the sound of their instrument. One unusual and very successful instance was Rusty Young of Poco. Rusty achieved some marvelous sounds with Poco by putting his steel guitar through a Leslie, producing a very distinctive and attractive effect. The instrumental line in Player's recent hit, "Baby Come Back", and J. Geils' soulful intro to "The Usual Place" are some more typical but outstanding examples of tasty Leslie guitar. I'm sure your lists contain a lot of others.

Personal Madness

I am arranging and putting together this article using a Norelco Dictating Machine, which has a microphone at the end of a coil cord that also acts as a loudspeaker. You can push a button on the recorder and the sound will come out of the microphone end. By grabbing the microphone by its cord and swinging it in circles, this entire article can be played back to me with a very strange and mysterious Leslie-like effect. It's really okay. No... I'm really allright. It's just like a Leslie, you see...

Acknowledgements

Thanks to Ralph Brownlowe, Robert Rankin and Ed Sciaky (WIOQ, Philadelphia) for some priceless and

valuable information used in the compilation of this article.

[This was stated by the author to be Gary Brooker, but he was not the organist for Procol Harum to play "Whiter Shade of Pale" - it was Matthew Fisher. The correction has been made here and will be forwarded to the author. -ed]

Doppler Distortion

Piston Vibrating in a Tube Including the Effect of Excursion

In other sections of this web site the wave produced by a vibrating piston is analyzed, but the effect of the piston excursion is neglected. That is, the piston has velocity, but the piston face is always treated as if its deviation from a fixed location is negligible. Here we obtain an exact solution where the excursion is taken into account. The excursion introduces harmonic distortion even when the piston is vibrating with a pure sinusoidal motion at a single frequency. When two frequencies are present intermodulation frequencies are created. Calling this "Doppler Distortion" may be a misnomer, but that is what people have called it for a long time.

There are people who believe that this is all a myth. As a believer in physics and mathematics, for me the derivation below is conclusive, and ends the debate. For the mathematically challenged among us, there is a heuristic description of the cause of this distortion at the end of this section.

Introduction: Fundamental Cause of Doppler Distortion

Consider a vibrating piston whose position x_p is defined by the equation

$$(1) \quad x_p(t) = -\frac{v}{\omega}\cos\omega t$$

The piston velocity is

$$(2) \quad v_p(t) = v\sin\omega t$$

Now consider an ideal sound wave in the tube, a single harmonic tone at the same frequency ω. We assume that the sound obeys the linear one-dimensional wave equation, i.e. that the sound intensity is within the region where air behaves linearly. Neglecting a multiplicative constant, the mean molecular velocity u, in the x-direction, at any point $x \geq x_p$, and at any time t satisfies

$$(3)\quad u(x,t)=\sin\left(\omega\left[t-\frac{x}{c}\right]\right)$$

Where c is the speed of sound. This form follows directly from the wave equations.

The crux of the situation is this:

At any fixed point x in the tube, the time variation of u is a pure sine wave. The time variation of the piston velocity is a pure sine wave when it is measured at the face of the piston, $x=x_p$,

The time variation of u is not a pure sine wave at $x=x_p$. Thus the piston velocity of equation (2)

does not match the sound wave velocity of equation (3) at the interface of the piston and the air. Since this is the basic boundary condition that must be satisfied, equation (3) is not a valid solution for the assumed piston motion. So at this point we know the piston is not generating a pure tone - i.e. there is Doppler distortion.

This velocity mismatch raises several interesting issues. We will first determine what tones are generated by a piston velocity defined by equation (2). We will return to a discussion of the interesting issues at the end.

Harmonic Vibration at a Single Frequency: Exact Solution

The necessary and sufficient conditions for an exact solution for the sound wave are: (1) the solution satisfies the wave equation, and (2) the wave velocity equals the piston velocity at the piston face. It should be emphasized that this means an exact solution everywhere in front of the piston, from smack dab on the piston face to plus infinity.

As a trial solution, consider

$$(4)\quad u(x,t)=b_0+\sum_{n=1}^{N}a_n\sin\left(n\,\omega\left[t-\frac{x}{c}\right]\right)+b_n\cos\left(n\,\omega\left[t-\frac{x}{c}\right]\right)$$

By construction this trial solution always exactly satisfies the wave equation. Equation (4) defines the wave velocity for any value of x, at any time t. As discussed above, we need to

match the piston velocity at the location of the piston face at any given moment. Therefore we will have a valid solution if we can find coefficients such that at $x=x_p(t)$ the wave velocity matches the piston velocity, meaning

$$(5)\quad v\sin\omega t = b_0 + \sum_{n=1}^{N} a_n \sin\left(n\left[\omega t + \frac{v}{c}\cos\omega t\right]\right) + b_n \cos\left(n\left[\omega t + \frac{v}{c}\cos\omega t\right]\right)$$

Equations (4) and (5) have some superficial similarity, but the crucial difference is that equation (4) gives the velocity as a function of both position and time, whereas equation (5) gives the velocity as a function of time alone at one loci in space, namely the piston face. So equation (5) is used to solve for the coefficients, and then they can be plugged into equation (4) to define the wave everywhere in the tube. If the objective is simply to know the resulting harmonic content, that is given directly by the coefficients themselves.

Being a lazy type, I don't even want to think about solving this nasty equation analytically. But it turns out that a pretty simple Matlab program generates very accurate solutions very quickly (I will send the program to anyone who wants it).

A piston in a tube is ideally loaded, and doesn't need to move much to generate very loud sound levels. For example, at a sound level of 117 dB SPL in the tube the peak piston velocity is only .0488 meters per second, and from equation (5) the first harmonic is 83 dB down from the fundamental. But a piston in an infinite baffle has to move a lot more. A loudspeaker cone with a peak excursion of 5 millimeters at 100 Hz moves at a peak velocity of 3.14 meters per second. At this velocity first harmonic from equation (5) is down 47 dB. A plot of the harmonics for a very high peak cone velocity of 10 meters per second is shown here [34 kB].

There is a constant velocity term b_o in equation (5). Much to my delight the numerical results for b_o from equation (5) agree perfectly with the constant that I predicted a long time ago, based on totally different considerations; see equation (36) in the section on plane waves. Since this is a pretty interesting result by itself, I have appended further discussion of this effect at the end of this section.

In any case, even with a "perfect" piston vibration, and with the air behaving in a perfectly linear manner, there are harmonics. The non-linearity arises strictly from the effect of the piston excursion.

Harmonic Vibration at Two Frequencies: Exact Solution

The numerical solution is in principle easily extended to the two-frequency case. The second frequency term is simply added to equations (1) and (2). To formulate equation (5), the "x" in equation (4), where we need to match the velocity, becomes the new $x_p(t)$ for the combined frequencies. Additional terms are added to equation (5) at frequencies corresponding to the new harmonics and intermodulation frequencies. Note that you can put whatever potential frequency terms you want in equation (5), as long as they satisfy the wave equation. If the frequency is not present, the program simply outputs a zero for that coefficient. In fact many of the coefficients output by the program are totally negligible. The "proof of the pudding" is if coefficients can be found such that the new equation (5) equals the new combined velocity. With the proliferation of terms the numerical solution is a little trickier than the one-frequency case. My current Matlab program only works for a large ratio of the high frequency to the low frequency. But for this situation, and with a little care, accurate solutions can still be found.

As an example, I have evaluated the case of frequencies of f1=50 and f2=1000 Hz, with peak velocities of 1 and .1 meters per second respectively. A plot of the piston velocity overlaid with the wave velocity is shown here [35 kB]. In other words, this is a plot of both the right and left hand sides of equation (5). The first harmonic of f1 is down 56.7 dB relative to the fundamental at f1. The first harmonic of f2 is down 76.7 dB relative to the fundamental at f2. The first sidebands at f2-f1 and f2+f1 are down 30.7 dB relative to the fundamental at f2. Therefore the intermodulation distortion is the dominant effect when these two frequencies are present. A plot of the harmonic content near f2 is shown here. [33 kB].

Harmonic Vibration at a Single Frequency: Approximate Solution

The above analysis is not easy to grasp intuitively. There is an approximate analysis that yields an answer quite close to the exact solution, and is much more intuitive. Given a piston vibrating at a single frequency, imagine a probe at some fixed location in the tube that measures the sound signal. To keep the argument simple, ignore any harmonics at this point, and assume we have a pure sine wave with wavelength λ. If the center of vibration of the piston is moved forward a distance d to a new fixed location, the phase of the signal measured by the probe will increase by a fixed value of $2\pi d/\lambda$ radians. So a forward motion imposed on the piston changes the phase. If we assume that the same phase change occurs when the piston face is displaced dynamically from its center of vibration, then there is a phase modulation of the signal due to the piston excursion. Assuming such a phase modulation, for the single frequency case with the piston location given by equation (1) the resulting signal measured by the probe, neglecting unimportant constants, would be

$$(6)\quad s(t)=\sin\left(\omega\left[t+\frac{x_p(t)}{c}\right]\right)=\sin\left(\omega t-\frac{v}{c}\cos\omega t\right)$$

Equation (6) can be expanded in an infinite series of harmonics of the fundamental

$$(7)\quad s(t)=-J_1\left(\frac{v}{c}\right)+\left[J_0\left(\frac{v}{c}\right)+J_2\left(\frac{v}{c}\right)\right]\sin\omega t-\left[J_1\left(\frac{v}{c}\right)-J_3\left(\frac{v}{c}\right)\right]\cos 2\omega t+\cdots$$

Where J_n is the Bessel function of the 1st kind. For a quite high piston velocity of 20 meters per second, which produces a first harmonic 30.7 dB below the fundamental, the coefficients produced by equation (7) agree with the three equivalent coefficients produced by equation (5) to within 0.2%. The coefficient of the 2nd harmonic term is approximately equal to $v/(2c)$ for $v<<c$.

Harmonic Vibration at Two Frequencies: Approximate Solution

Again to keep things relatively simple, assume that the piston excursion is dominated by the movement at a low frequency ω_1. Then the high frequency signal at ω_2 will have an approximate phase modulation given by

$$(8)\quad s(t)=\sin\left(\omega_2\left[t+\frac{x_p(t)}{c}\right]\right)=\sin\left(\omega_2 t-\frac{\omega_2 v}{\omega_1 c}\cos\omega t\right)$$

So the result is similar, except the effect is magnified by the ratio of the frequencies. The approximate coefficient solution for the 2nd harmonic, $(\omega_2 v)/(2\omega_1 c)$ also gives a result quite close to the two-frequency case presented above.

Interesting Issues

The approximate analysis is so accurate one might ask why I went to the trouble of generating the exact solution. I would turn this around, and say that what interests me is why the approximate solution is not exact! For very high velocities there is a significant disagreement. So the dynamic phase shift argument is approximate and needs verification, which is provided by the exact analysis.

The common method of computing Doppler distortion for the two frequency case is to: (1) compute a dynamic Doppler frequency offset based on the cone velocity at the low frequency, and (2) apply that as a FM modulation of the higher frequency. It is not obvious at first sight, but it is not difficult to show that (with one minor qualification) this is mathematically identical to applying a phase modulation based on the cone position, as done here. That is, the sidebands predicted by both analyses have exactly the same magnitude and phase. But for the single frequency case there isn't a Doppler frequency shift per se, and the phase modulation approach is the only way to get an answer.

Another question is: if an amplifier outputs a perfect sine wave voltage, what is the cone motion of a good speaker? I went back over the details of the Thiele-Small analysis to confirm that it is in fact the sinusoidal motion assumed in equations (1) and (2), as long as the speaker is in its linear region. Mainly this means that the voice coil stays in the linear region of the magnetic field, and the cone excursion is in the linear region of the suspension.

An important question one might ask is if there are similar effects during a sound recording process, and maybe the "distortion" actually washes out. If the musical instrument is a piston in the tube, and if the microphone is a piston in the tube, and if the speaker is a piston in the tube, in principle I think this is possible. But there is still an immediate qualification: the harmonics are amplitude dependent, so if, for example, the playback volume is different than the original "music" volume, the harmonics in the playback will be different - i.e. distorted. In the real world of course the sound source, microphone, and speaker all have very different velocities of motion.

What about Fourier analysis? Certainly the piston motion can be represented by a Fourier transform, and all of the usual spectral analyses can be performed. The same is true for the sound wave velocity generated by the piston motion. However, the two spectra will differ. In most practical cases the difference is small, but the sidebands produced by the modulation of a high tone by a low tone can certainly be significant.

The last question regards beat notes. Since my Matlab program only works for widely separated frequencies, I can't look for these, but you certainly have to guess that they would be there as well.

This type of distortion can only be reduced by minimizing the motion of the surface generating the sound. A larger surface moves less for a given sound level. The large diaphragm of an electrostatic speaker moves very little, and this might at least partially explain the wonderful clarity of some ES speakers. A three-way system is better than a two-way in this regard. A 4th order crossover is better than a 1st or 2nd order. Of course there are other factors involved in all of these design choices, and Doppler distortion may not be the most critical factor.

Heuristic Description of Doppler Distortion

Consider a piston vibrating in a perfect sinusoidal motion at 500 Hz. The wave generated in the tube will look a lot like a 500 Hz sine wave at first glance. But let's consider the zero crossings of the wave. Say the piston position vibrates between x=0 and x=1 centimeter (cm), and begins at x=0 with zero velocity at t=0. So the velocity wave begins with a pulse shape that starts at zero followed by a rising pulse slope. The pulse begins to travel down the tube at the speed of sound, 344 meters per second. At t=1 millisecond (ms) the piston is at x=1 cm and its velocity is zero. At this instant in time the wave consists of a positive half-cycle pulse with the first zero that has traveled down the tube to x=34.4 cm, and the second zero at x=1 cm. The zeros are 33.4 cm apart, and they will remain at that separation as the pulse continues to travel down the tube. At t=2ms the piston is back at x=0 and again has zero velocity. At this instant in time the wave consists of one full cycle, a positive pulse followed by a negative pulse, with zeros that are now at x=68.8 cm, x=35.4 cm, and x=0 cm. This pattern with a wavelength of 68.6 cm will then continue to be replicated as time goes on. The crucial point is that the zero crossings are not equally spaced; on the positive pulse section they are 33.4 cm apart, and on the negative pulse section they are 35.4 cm apart. We have a distorted sine wave.

Mass Transport for a Piston Vibrating in a Tube

To understand this, it is necessary to understand a sound wave as the motion of air molecules. For a pure tone, defined by equation (3), the average molecular velocity forms a perfect sine wave. At the wave peaks, the molecules, on average, are moving forward. At the troughs, they are moving backward at exactly the mirror image of the forward velocity. At the wave crests the molecular density is higher than at the wave troughs. Therefore there is a net flow of molecules in the direction of wave propagation. If you calculate the power flow in watts per square meter, you find that the wave is carrying more power than the piston is putting out. (In this calculation piston excursion is neglected). In other words, you are literally creating energy "out of thin air." Now everyone knows that there really isn't a net flow of molecules, and you can't violate the conservation of energy. What's wrong is that equation (3) by itself is not the right solution. If you add the constant term mentioned above, both problems are perfectly cleared up: there is no net flow of molecules, and the piston puts out exactly as much wattage as the wave carries down the tube. All of the physics and math to back all of this up is given elsewhere on this web site. Follow the link above, and subsequent links.

Acknowledgement: I benefited from technical exchanges in carrying out this analysis. Due to issues beyond my control that is all I am going to say.

Back to the Physics contents

Back to the Main table of contents

3

SOUND

Sound appears to be a topic distinct from motion and heat. However, we now understand sound to be an ordered motion of the molecules of the medium through which the sound propagates. The study of sound provides the opportunity to understand wave motion in preparation for the more abstract study of electromagnetic waves. The subject of sound should be of particular interest to anyone who enjoys music[1-7].

REFERENCES

1. J. Askill, *Physics of Musical Sound*, Van Nostrand: New York (1979).

2. M. J. Moravcsik, *Musical Sound: An Introduction to the Physics of Music*, Paragon House Publishers: New York (1987).

3. J. S. Rigden, *Physics and the Sound of Music*, John Wiley & Sons: New York (1977).

4. T. D. Rossing, *The Science of Sound*, Addison Wesley Publishing Company: New York (1982).

5. H. E. White and D. H. White, *Physics and Music*, Holt, Rinehart, & Winston: New York (1980).

6. A. Wood, *The Physics of Music*, John Wiley & Sons: New York (1975).

7. J. R. Pierce, *The Science of Musical Sound*, Scientific American Books, W. H. Freeman: New York (1983).

Safety Considerations with Sound

Most experiments involving sound are relatively safe except for the hazards not directly associated with the sound as discussed elsewhere. However, it is possible for very intense sounds to permanently

damage the ear especially after prolonged exposure[1,2]. Consequently, in the demonstrations described here, the sound intensity should always be kept at a level well below the threshold of pain (see Table 3.1).

As a general rule, prolonged exposure to sound levels above 85 dB will cause slight hearing loss and above 90 dB will result in mild to moderate loss. Prolonged exposure to sound levels above 95 dB causes moderate to severe hearing loss. Above 100 dB, even short exposure can cause a permanent loss of hearing.

The ear has an acoustic reflex that protects the inner ear from loud sounds in the same way that the pupil of the eye protects the eye by contracting in the presence of bright lights. However, the reflex requires a few hundredths of a second to respond, and thus cannot be relied upon to protect the ear in the event of a short duration sound such as an explosion.

The damage is typically by way of tearing or ripping the microscopic hair cells of the cochlea. Such damage is usually only temporary unless the sound is frequent or sustained. Especially intense sounds are capable of rupturing the eardrum. The damage threshold is frequency-dependent, and the ear is most susceptible to damage by sounds of around 3000 Hertz in part because the auditory canal is a closed tube having a resonance in this region. Prolonged exposure to sounds of a particular range of frequencies can permanently reduce the sensitivity of the ear to those frequencies.

The monitoring of sound levels requires an A-scale sound level meter with a frequency response matched to the response of the ear or one capable of displaying the sound level for all of the various frequency ranges[3]. Protection can be provided by ear plugs and/or ear muffs. Cotton in the ears is not an effective protection unless the cotton is wax-impregnated.

REFERENCES

1. K. D. Kryter, W. D. Wood, J. D. Miller and D. H. Eldredge, Journ. Acoustical Soc. Am. **39**, 451 (1966).

2. E. A. Lacy, *Handbook of Electronic Safety Procedures*, Prentice- Hall: Englewood Cliffs, New Jersey (1977).

3. A. P. G. Peterson and E. E. Gross, Jr., *Handbook of Noise Measurement*, General Radio: Concord, Massachusetts (1963).

Table 3.1

Sound Intensities

Decibels (dB)	Intensity (W/m^2)	Dp/p_o	Example
0	10^-12	2x10^-10	Threshold of hearing
20	10^-10	2x10^-9	Whisper (at 1 m)
40	10^-8	2x10^-8	Mosquito buzzing
60	10^-6	2x10^-7	Normal conversation
80	10^-4	2x10^-6	Busy traffic
100	10^-2	2x10^-5	Subway
120	1.0	2x10^-4	Threshold of pain
140	100	2x10^-3	Jet on carrier deck

3.1

Breathing Helium and Sulfur Hexafluoride

The peculiar sound of one's voice after breathing helium or sulfur hexafluoride provides an amusing demonstration of the variation of the velocity of sound with the density of a gas.

MATERIALS

- helium and sulfur hexafluoride*
- balloon
- funny hats (optional)

*Specialty gases and equipment can be procured from Matheson Gas Products, P. O. Box 1587, Secaucus, NJ 07094.

PROCEDURE

A dramatic illustration of the variation of the speed of sound with the density of a gas is performed by breathing a low density and a high density gas and talking as the gas is expelled from the lungs. Suitable gases are helium and sulfur hexafluoride. Xenon, though more expensive, can be used in place of sulfur hexafluoride. The gases are most conveniently breathed by first filling a balloon from a gas cylinder with a pressure regulator set to about 5-10 psi. One should exhale the air from the lungs and then inhale the gas through the mouth as if sucking a straw. The gas is then slowly exhaled as one talks. A duck hat

for the helium and a cowboy had for the sulfur hexafluoride adds to the humor.

One can also fill a balloon with sulfur hexafluoride and tie it off. It is distinctly heavier than an equivalent air balloon as can be illustrated by dropping the two side by side.

An especially instructive demonstration is to breathe a mixture of helium and sulfur hexafluoride in a proportion such that its density is close to the density of air. This can be done by first filling the balloon with helium and then adding just enough sulfur hexafluoride so that the balloon is slightly heavier than air (like an air-filled balloon). In principle, one's voice should sound normal after breathing such a mixture, but in practice the helium rises first into the nasal cavity and is expelled, leaving the sulfur hexafluoride which is exhausted at a slower rate. Such a demonstration shows that the density of the gas rather than its particular composition is the relevant variable.

DISCUSSION

The speed of sound in a gas is proportional to the square root of its density. For helium the speed is 2.7 times the speed of sound in air. For sulfur hexafluoride the speed is 0.44 times the speed of sound in air. The speed of sound in air is 343 m/s or 750 mi/hr at 20°C and is proportional to the square root of the absolute temperature.

The voice sounds different because the natural resonant frequencies of cavities in the head are raised or lowered in proportion to the speed of sound[1]. Of course the resonant frequencies of these cavities are only one ingredient in the way we sound. The gas does not alter the frequency of the vocal cords nor does it affect how fast we talk. The vocal tract acts as a filter for the sounds produced by the vocal cords in a manner similar to the treble and bass controls on a stereo.

Most people have heard the distinctive yet highly amusing sound that results when one breathes helium. Deep sea divers often breathe an atmosphere that is largely helium so that they can exist in a high pressure environment without exceeding the body's tolerance for nitrogen and oxygen. Most people have not heard the equally amusing sound that results from breathing sulfur hexafluoride. Sulfur hexafluoride has the additional property that, since it is a high density gas, it tends to remain in the lungs for a long time, and thus the effect on the voice persists much longer.

HAZARDS

The major danger in performing this demonstration is hypoxia from not getting enough oxygen while the gas is being breathed. Thus one should not immediately repeat the demonstration but wait a minute or so before breathing more gas. It is a good idea to take a few deep breaths before and after each demonstration. One should point out to the audience that the breathing of most gases (even some air) is potentially harmful. Helium and sulfur hexafluoride are two of a very small number of gases with which this demonstration can be safely done. The substitution of other gases, especially flammable gases such as hydrogen, is not recommended.

REFERENCE

1. A. H. Cromer, *Physics for the Life Sciences*, McGraw Hill: New York (1977).

3.2

Oscilloscope Waveforms

An oscilloscope is used to display the waveforms of various musical instruments in order to show the effect of frequency and wave shape on the sound.

MATERIALS

- oscilloscope
- microphone
- various musical instruments
- tuning forks (optional)
- electronic keyboard (optional)
- function generator (optional)
- noise generator (optional)
- amplifier and speaker (optional)
- inexpensive transistor radio (optional)

PROCEDURE

An oscilloscope with a large screen can be connected directly to a microphone or to the output of an audio amplifier or function generator to illustrate the waveforms produced by various sounds.The oscilloscope should be set to sweep at a rate of about 200 Hz (5 milliseconds per sweep). This rate is fast enough to prevent flicker but slow enough to observe the low frequencies. If the oscilloscope has a triggered sweep, it should be set on "internal" and the triggering level set to a low value to make the resulting pattern as stationary as possible. A storage scope allows the waveform to be saved so that it can be discussed. In a large room, a TV camera and monitor(s) can be used to improve visibility of the oscilloscope, and in such a case the oscilloscope and camera can be backstage and connected to the PA and TV system.

Sine waves can be produced with tuning forks or organ pipes or with an electronic function generator. One could first show the sounds corresponding to sine waves of different frequencies. Anyone in the audience with perfect pitch should be able to identify the musical notes (see Table 3.2). The way in

which musical scales are constructed makes an interesting digression[1,2]. The standard of A-440 is not universally accepted, nor is the even-tempered scale. Some scientific manufacturers once adopted a standard of 256 (2^8) for middle C, but musicians ignored it. With a function generator, one can show the frequency limits of hearing (approximately 20 Hz to 20 kHz).

Musical instruments such as a violin and a trumpet have distinctive waveforms[3,4]. The clarinet somewhat resembles a square wave. One could show examples of string, wind, and percussion instruments. One can illustrate frequency modulation (vibrato) and amplitude modulation (tremolo). If someone in the audience can play the instruments, it adds a touch of interest to have them do so. Alternately, an accomplished musician can be planted in the audience who when called upon pretends minimal skills until suddenly launching into "The Flight of the Bumblebee" or some similarly intricate piece.

Modern electronic keyboards can be used to emulate a wide variety of musical instruments. Demonstrating the difference in waveform between a real instrument and a synthesized instrument can be quite interesting. The complex waveform required to transmit speech and music is especially dramatic.

Square waves and other waveforms can be produced with a function generator connected to an amplifier and speaker as well as to an oscilloscope. One could illustrate the different sounds produced by sine waves and square waves of the same frequency. The difference results from what musicians call "overtones" and physicists call "harmonics." Harmonics are integral multiples of the fundamental frequency, but overtones may or may not be related to the fundamental in a simple way. In most string and wind instruments, the overtones form a harmonic series, but in percussion instruments such as the drum, the overtones are more complicated and less "harmonious" (not a single discernible note). The combination of overtones is what gives each musical instrument its characteristic quality, or timbre.

This demonstration offers the opportunity to exhibit unusual musical instruments and other sources of sound. One might try blowing over the top of a soft-drink bottle and alternately drinking from it to change its pitch. The same can be done with a flexible plastic tube (1/2" diameter Eastman Poly-Flo) partially filled with colored water, fruit juice or wine, which one can drink when done. The bottom can be cut out of a Coke bottle, and different notes played by blowing over the top while it is dipped different depths into a beaker of water. Wine glasses stuck with a small mallet or rubbed along the rim with a clean, grease-free finger emit good sine waves. Music has been written for such a glass harmonica. A bent saw blade, whistles, horns and sirens are other possibilities. A long, hollow pipe, open on both ends, lowered over a Bunsen burner can also be made to emit a sound[5], as was first done by Sir Charles Wheatstone (1802-1875) in a series of public lectures at the Royal Institution in London [6].

An especially impressive demonstration can be done with an aluminum rod about a centimeter in diameter and at least a meter long. Hold the rod at the center in one hand. Put powdered resin on the fingers of the other hand and stroke the rod along its length to excite longitudinal vibrations. These

vibrations have a high pitch and vibrate for a very long time after one stops stroking. Thus they are extremely sinusoidal. The sound can be stopped abruptly by touching one of the antinodes at either end of the rod. Then grip the rod at a point one quarter of the length from one end and repeat the stroking to excite the second harmonic. It helps to place a mark at the right spot.One can compare these longitudinal vibrations with the transverse vibrations that occur when the rod is struck with a mallet. The transverse vibrations are at a much lower frequency unless the rod is very short as in a xylophone.

One can illustrate that the perceived loudness of a sound depends upon its wave shape. The square wave sounds louder than a sine wave of the same frequency because of the harmonics, even when the amplitude of the square wave is lowered by $(pi/2)^{½}$ to account for its larger average power. Also the perceived pitch of a sound depends slightly on its intensity[7]. Usually the pitch change is downward at low frequencies and upwards at high frequencies as the loudness increases, but the effect varies considerably with the individual[8]. Furthermore, the effect is noticeable only for sine waves[9,10]. These perceptions can be demonstrated with a pair of function generators and a switch to alternate between the two. If the sources are sounded simultaneously, the presence of beats can obscure the effect.

An inexpensive transistor radio can be connected to the oscilloscope to illustrate the distortion that results when the volume is turned up too high and the waves are flattened due to saturation of the transistor amplifiers. This could lead to a discussion of high fidelity audio equipment and the need for high power amplifiers to minimize distortion due to the generation of harmonics.

By tuning a radio or television between stations, one can demonstrate what noise sounds like. There are many types of noise such as white noise (all frequencies present in equal amounts), 1/f noise (sometimes called pink noise), $1/f^2$ noise (sometimes called brown noise since it is characteristic of brownian motion), and others. Commercial noise generators are available that produce the various types. Alternately various kinds of noise can be simulated by filtering white noise through an audio equalizer (adjusted to unequalize).

DISCUSSION

It is interesting to note that the shape of a waveform is determined by its Fourier frequency components. The phase as well as the amplitude of the individual frequency components will affect the shape of the waveform, but since the ear is insensitive to the relative phases of the components of a wave[11], waveforms of quite different shapes may sound identical to the ear.

HAZARDS

There are no precautions other than keeping the sound intensity down to a safe level.

REFERENCES

1. J. Backus, *The Acoustical Foundations of Music*, W. W. Norton & Company: New York (1969).

2. E. E. Helm, Scientific American **217**, 93 (Dec 1967).

3. H. Lineback, Scientific American **184**, 52 (May 1951).

4. C. A. Culver, *Musical Acoustics*, McGraw-Hill: New York (1956).

5. J. S. Miller, Am. Journ. Phys. **27**, 367 (1959).

6. J. Tyndall, *Sound*, third edition, Longman: London (1975).

7. S. S. Stevens, Journ. Acoustical Soc. Am. **6**, 150 (1935).

8. A. Cohen, Journ. Acoustical Soc. Am. **33**, 1363 (1961).

9. W. B. Snow, Journ. Acoustical Soc. Am. **8**, 14 (1936).

10. D. Lewis and M. Cowan, Journ. Acoustical Soc. Am. **8**, 20 (1936).

11. G. von Bekesy, *Experiments in Hearing*, McGraw-Hill: New York (1960).

Table 3.2

The Notes of the Even-tempered Scale for the Octave Above Middle C[a]

Note	Frequency
C	261.63 Hz
C#	277.18 Hz
D	293.66 Hz
D#	311.13 Hz
E	329.63 Hz
F	349.23 Hz
F#	369.99 Hz
G	392.00 Hz
G#	415.30 Hz
A	440.00 Hz

A#	466.16 Hz
B	493.88 Hz
C	523.25 Hz

[a]A-440 is used as the standard, and the interval between each note is $2^{1/12} = 1.05946$.

3.3

Doppler Effect

A reed mounted on the end of a rotating arm produces a tone whose pitch wobbles up and down as the arm rotates.

MATERIALS

- sound reed
- motor with rotating arm*
- strobe lamp (optional)
- tape recording of train whistle or car horn (optional)

*Available from Fisher Scientific and Sargent-Welch Scientific Company

PROCEDURE

The Doppler effect (named after the Austrian, Christian Johann Doppler, 1803-1853) can be illustrated in a device which consists of a reed mounted on the end of an arm that is spun by means of a motor with suitable gears to control the speed[1]. The rotation causes air to pass over the reed producing a sound whose pitch wobbles up and down as the reed successively moves toward and then away from the observer. It helps if the speed of rotation can be adjusted. If a microphone is being used, it should be turned off or placed on the side of the apparatus toward the audience to reduce the interference between the direct and amplified sounds. The spinning arm can be viewed with a strobe lamp to freeze its motion. Alternate techniques for exhibiting the Doppler effect are also available[2-4]. A smoke detector wired so as to emit a continuous tone can be swung around on the end of a rope.

One can illustrate the effect verbally, but a tape recording of a train whistle or a car horn is more effective. Recordings of the Doppler effect can often be found on sound effect albums at record libraries or made quite easily with a tape recorder and automobile on a country road. This makes a good weekend family project for students who can then share their results with their classmates.

DISCUSSION

The frequency heard by a fixed observer for a source in motion is given by $f = f_o v/(v \pm v_s)$, where f_o is the frequency of the source, v_s is its velocity and v is the velocity of sound (about 343 m/s in air at 20° C). The plus sign holds for motion away from the observer, and the minus sign holds for motion toward the observer. Thus if one wants to produce a ±5% frequency shift for a reed at the end of a 25-cm-long arm, a rotation speed of about 650 rpm is required.

The Doppler effect was tested by a rather bizarre experiment in Holland a few years after Doppler derived his formula in 1842. For two days a locomotive pulled a flat car back and forth at different speeds. On the flat car were trumpeters sounding a particular note. On the ground, musicians with a sense of absolute pitch recorded the note as the train approached and receded.

There are many examples in nature of the Doppler effect. In addition to the usual train whistle and car horn, the Doppler effect also occurs for electromagnetic waves, but since the velocity of light is much larger than the velocity of sound, the effect is much smaller for a given source velocity. Even so, the police use Doppler radar to catch speeders. Doppler radar is now used to measure the speed of baseballs and tennis balls. In air traffic control, the Doppler effect is used to discriminate against stationary targets and to detect wind shear. The Doppler effect is also important in the reception of radio signals from earth-orbiting satellites. The Doppler effect is used in the determination of the distance of galaxies[5,6] (the Hubble constant) and in the Möessbauer effect.

HAZARDS

The main hazard in this demonstration is coming in contact with the rapidly rotating arm. A switch and speed control should be provided that permits operation from a safe distance. The apparatus should not contain parts that could come loose and be thrown out into the audience.

REFERENCES

1. H. A. Robinson, ed., *Lecture Demonstrations in Physics*, American Institute of Physics: New York (1963).

2. F. S. Crawford, Am. Journ. Phys. **41**, 727 (1973).

3. H. F. Meiners, *Physics Demonstration Experiments*, Vol I, The Ronald Press Company: New York (1970).

4. J. S. Miller, *Physics Fun and Demonstrations*, Central Scientific Company (1974).

5. E. Hubble, Proc. Natl. Acad. Sci. **15**, 168 (1929).

6. E. Hubble and M. Humason, Astrophys. J. **74**, 43 (1931).

3.4

Bell in Vacuum

An electric bell in a jar makes a sound that decreases in intensity as the air is evacuated from the jar.

MATERIALS

- electric doorbell
- battery or power supply
- bell jar*
- vacuum pump
- pressure gauge (optional)
- sound meter (optional)

*Available from Carolina Biological Supply Company, Central Scientific Company, Fisher Scientific, Frey Scientific Company, Nasco and Sargent-Welch Scientific Company

PROCEDURE

The necessity of a medium for the propagation of sound can be demonstrated by means of an electric bell (a doorbell type) in an evacuated bell jar. The electrical leads for the bell can be brought out of the bell jar through vacuum feedthroughs. Transmission of sound by the electrical leads can be kept to a minimum if they have a small cross-section and are rather long and coiled up. They can be led through a rubber stopper in a hole in the glass to further dampen the transmitted sound. The bell should be suspended by its leads so as not to touch the bell jar. The bell is turned on, and the bell jar is evacuated to a pressure in the millitorr range with a mechanical pump while the sound slowly diminishes until it is almost inaudible. With a sound meter and pressure gauge, one could plot the sound intensity as a function of pressure.

In a simpler version, a small bell with a clapper is mounted inside a jar small enough to be held in the hand and shaken. The jar is equipped with a rubber stopper, hose and shut-off valve.

DISCUSSION

The sound propagation should cease when the mean free path for air molecules becomes longer than a

few centimeters. At room temperature and 760 torr of pressure (1 atmosphere), the mean free path is about 2×10^{-5} cm and is inversely proportional to pressure. Thus the pressure should ideally be reduced to about 10 millitorr to reduce the sound to the lowest possible level.

HAZARDS

The only significant hazard with this demonstration is implosion of the bell jar. If the jar is designed for vacuum use (heavy flint glass), it should be safe so long as it is not struck with a heavy object or otherwise cracked. The bell should be of the type that operates from a low voltage (0-6 volts) so as not to pose an electrical hazard. A variation of the demonstration, in which one places a battery inside the bell jar to operate the bell, is not recommended since the battery may explode when exposed suddenly to a vacuum.

3.5

Wave Speed on a Rope

The difference in wave propagation speed for transverse waves on ropes of different masses and tensions is illustrated with a stick and two, 3x5-inch index cards.

MATERIALS

- various ropes
- various weights
- 2 pulleys
- 2 3x5-inch index cards (preferably different colors)
- stick

PROCEDURE

One end of a rope is attached to the wall at about eye level. The rope is stretched across the room, and the free end is given a flick with the wrist. One can show that the speed of the wave on the rope increases with the tension in the rope but is independent of the shape of the wave. Dispersion (change in the shape of the wave as it propagates) can be observed. The reflection of the wave at the fixed end can be discussed. Note that the wave is inverted upon reflection from the fixed end.

Two identical ropes are strung across the front of the lecture room. The tension is made different by placing one end of each rope over a pulley and hanging different weights on each with the other end of the ropes anchored securely in the opposite wall. Two, 3x5-inch index cards (preferably different colors)

are folded and placed one over each rope about a foot from the anchored end. The opposite ends of the ropes are then struck simultaneously with a stick while the audience is told to watch the cards. The card over the rope with the greatest tension will jump off the rope slightly, but very noticeably, before the other.

To illustrate the variation of velocity with mass density, two ropes are used, with obviously different masses (say a quarter-inch-diameter and a half-inch-diameter rope). For ropes of the same material, the mass per unit length is proportional to the square of the diameter of the rope. The ropes are attached together and passed around a single pulley so as to produce the same tension in each. Cards are placed over the ropes near the ends that are anchored separately to the wall, and the ends that are passed around the pulley are struck with a stick as described above. The card over the smaller rope will jump off before the other.

One can combine the two demonstrations by using two ropes with different mass densities under different tensions, but with the same ratio of tension to mass density. In such a case, the cards should jump off simultaneously.

DISCUSSION

Fundamental to understanding the operation of musical string instruments is the velocity of a transverse, traveling wave on a string or rope under tension. Usually the velocity is so great that it is difficult to observe directly. The wave velocity is given by $(T/\mu)^{1/2}$, where T is the tension and μ is the mass per unit length. Under the same tension, a string or rope of the same material but twice the diameter as another will propagate sound a factor of two slower. String instruments have thick strings (large μ) for the low notes and thin strings (small μ) for the high notes as well as some means to adjust the tension in the strings.

HAZARDS

The ropes should be strung across the room above head level to prevent walking into them. Care should be exercised to ensure that the weights do not fall on a foot if a rope breaks, comes loose from its anchor in the wall or slips off its pulley.

3.6

Beat Frequencies

Two sound sources of nearly equal frequencies and amplitudes are made to exhibit beats if the frequency difference is less than about 10 Hertz.

MATERIALS

- 2 tuning forks or organ pipes with adjustable pitch
- 2 sine wave generators with amplifier and speaker (optional)
- oscilloscope (optional)
- 2 glass tubes to fit in organ pipes (optional)
- nichrome wire and power supply (optional)

PROCEDURE

The beating of two tones of nearly equal pitch sounds to the ear like a single tone whose intensity varies at the beat frequency. The demonstration can be performed using any two sound sources, but the effect is most striking if the the sources are sinusoidal and of equal intensities. The cleanest effect results from using two sine wave generators of equal amplitude connected to the same loudspeaker, in which case the amplitude varies from zero to twice the amplitude of each source[1-4]. The voltage applied to the speaker can be displayed on an oscilloscope. Two identical tuning forks also work well if a small piece of wax is placed on one of the forks to alter its frequency. Similarly, two identical organ pipes can be used if one is detuned by partially covering its mouth or by adjusting a movable plug if the pipe is so equipped.

The beating of two identical organ pipes can be used to illustrate the variation of the velocity of sound with temperature[1]. Two identical glass tubes are placed one in each organ pipe, and the pipes are tuned for zero beating. A nichrome wire is lowered into one of the glass tubes and heated to 250-300°C by an ac current. A noticeable beating should occur as the resonant frequency of the heated pipe rises with the increased velocity of the sound. The wire can then be cooled until the beats disappear.

The explanation of the beating can be presented graphically by drawing two sine waves of slightly different frequencies on the blackboard and showing how they constructively interfere at certain times and destructively interfere at others. For an audience versed in trigonometry, a single trigonometric identity yields the result: $\sin(w_1 t) + \sin(w_2 t) = 2\sin[(w_1+w_2)t/2]\cos[(w_1-w_2)t/2]$. Thus two sounds of frequency w_1 and w_2 will be perceived as a single frequency $(w_1+w_2)/2$ but with an intensity that reaches a maximum twice during each cycle of the frequency $(w_1-w_2)/2$. Thus the beat frequency is w_1-w_2 (assuming w_1 is the larger frequency). If the two sounds are of unequal amplitude, the cancellation will not be perfect when the waves destructively interfere, and the intensity variation of the beat will be less pronounced. Note that there is no wave present at the beat frequency. This can be illustrated by using two frequencies above the audible range of hearing but adjusted so that the beat frequency is within the audible range. Actually, if the sounds are sufficiently intense, it is sometimes possible to hear the beat frequency of two inaudible sounds because of nonlinearities in the ear.

DISCUSSION

Beats provide a means for musicians and piano tuners to tune their instruments. The tension in a violin string is adjusted until a given note gives a zero beat frequency with a standard note being played by the concert master. Other notes can then be adjusted so that their overtones beat with the overtones of the standard. In fact, the just diatonic scale is arranged so that the different notes are related in frequency by the ratios of small integers so as to minimize the beating of the overtones when two notes are played together. Such a scale seems to give the most pleasing sound, although instruments are more typically tuned to the even tempered scale in which the notes have twelve even logarithmic intervals within each octave and thus produce some low frequency beating among their harmonics.

Beats appear to play an important role in our perception of the quality of music[5]. A low frequency beat (a few Hertz or less) can sound pleasing, in that it emulates the trembling of the human voice. Rapid beats can be very unpleasant to the ear just as a flickering light is unpleasant to the eye and a fingernail scratched across a surface is unpleasant to the touch. Thus when one singer in a chorus or one instrument in an orchestra is off key, the resulting beats render the sound highly obnoxious. The famous german physicist and acoustician (and chronically bad lecturer), Hermann von Helmholtz (1821-1894), had a reed organ built that was based on a 24 note scale optimized to reduce the beating between the harmonics of the various notes. The sound was reported to be very pleasing but apparently not sufficiently so to revolutionize the way in which music is written and played.

HAZARDS

There are no precautions other than keeping the sound intensity down to a safe level.

REFERENCES

1. H. A. Robinson, ed., *Lecture Demonstrations in Physics*, American Institute of Physics: New York (1963).

2. R. E. Miers and W. D. C. Moebs, Am. Journ. Phys. **42**, 603 (1974).

3. S. H. Vegors, Jr., Am. Journ. Phys. **43**, 1103 (1975).

4. H. F. Meiners, *Physics Demonstration Experiments*, Vol I, The Ronald Press Company: New York (1970).

5. H. L. F. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, Dover: New York (1954).

3.7

Breaking a Beaker with Sound

A glass beaker exposed to a sufficiently intense sound wave at its natural resonant frequency is made to shatter.

MATERIALS

- 2000-ml, clean, glass beaker or large wine goblet
- small mallet
- small, high-power loudspeaker
- audio amplifier (at least 50 watts)
- variable-frequency, sine wave generator
- oscilloscope (optional)
- microphone (optional)
- digital frequency meter (optional)
- strobe lamp (optional)

PROCEDURE

In this classic demonstration[1,2] the input of the audio amplifier is connected to the variable-frequency, sine wave generator, and the output is connected to the loudspeaker. The glass beaker is set on some material such as a styrofoam pad that allows it to ring with minimal damping and is placed a few inches from the speaker. The beaker is tapped near its open mouth with a small mallet to ensure that it rings with a slowly decaying tone. With a microphone and oscilloscope, the resonant frequency can be measured and shown to the audience. Alternately, the speaker can be connected directly to the input of the oscilloscope and used as a microphone. The sine wave generator must be tuned to exactly the same frequency as the natural resonant frequency of the beaker.

Without an oscilloscope, one can adjust the frequencies to be equal by repeatedly tapping the beaker with the mallet and listening for the beats with the sound from the speaker as the generator frequency is slowly varied. This works best if the sound level is about the same from the speaker and from the beaker. Alternately, one can put the speaker near the beaker at a low sound level and listen for the increase in sound intensity as the resonant frequency is approached. Still another method is to watch or feel the vibrations of the beaker with the fingertips as the frequency is varied. Strips of paper folded over the edge of the beaker give an indication of the amplitude of the vibration. It is best to find the proper frequency beforehand using a digital frequency meter and to record the value (to a precision of at least one Hertz). This is so that the resonance condition can be quickly found for the demonstration, although there is considerable pedagogic value in showing how the resonance is found if time permits.

The members of the audience are then cautioned to cover their ears, and the sound is turned up to a high level. A slight readjusting of the frequency may be required. The frequency must be adjusted *very*

slowly, and one must wait a few seconds for the oscillations to build to a value sufficient for the glass to break.

The demonstration can also be done with a wine goblet or any other glassware with vertical sides that exhibits a high quality resonance in the acoustical range. All that is required to test the quality of the resonance is to tap the glass with a fingernail or a pencil and listen to how long the resulting ringing lasts. Any glassware that can be made to ring for more than a few seconds should suffice. Glass with a high lead content works best. In a variation of the demonstration, one can use a fixed frequency sound source, tuned slightly below the natural resonant frequency of the glass, and then slowly add water or other liquid to the glass until its resonant frequency is lowered to match that of the sound source.

A strobe lamp tuned near the resonant frequency of the beaker or one of its harmonics allows one to observe the vibrations of the glass, although visibility is difficult except for those relatively close. A television camera and monitor can be used to improve visibility. Prior to breaking, the glass may deflect by as much as a centimeter.

DISCUSSION

This demonstration illustrates that a sound wave is a compressional wave with a characteristic frequency. The wave carries significant energy from the source to the point where it is absorbed.

More importantly, the demonstration illustrates the phenomenon of resonance. The Q (quality) of the resonance is defined as 2(pi) times the number of cycles (number of radians) required for the energy of an oscillation to decay to 1/e (37%) of its initial value. Thus the highest Q resonance is the one for which the ringing lasts the longest. A piece of glassware with a resonance at 1000 Hz that decays to 37% of its initial energy in one second has a Q of about 6000. The Q is also a measure the amplitude which an oscillator driven at its resonant frequency will acquire. Thus the glass with a Q of 6000 will vibrate with about 6000 times more energy when driven at resonance than when driven well off resonance with a sound of the same intensity. However, it takes about Q radians (one second in this example) for the oscillation to build to near its full amplitude. The driving frequency also has to be matched to the natural frequency to a precision of about 1/Q (one part in 6000 in this example). Since the sound intensity has to be high (in excess of 90 dB), precisely of the right frequency, and sustained for a few seconds, there is little danger of breaking one's wine glass while casually listening to music on the stereo!

Dorothy Caruso denies rumors that her husband Gigli could shatter wine glasses with his unamplified voice[3]. Although it might be possible for a singer to do so, there are no documentated cases of any having succeeded. The Memtek company in Forth Worth, Texas produced a television commercial showing Ella Fitzgerald and others breaking glasses with recordings of their voices on Memorex tape.

HAZARDS

Unless the demonstration is designed very carefully, the sound intensity required to break a beaker can

damage one's hearing. The sound should be turned up slowly to allow the audience to take protective action. The members of the audience should be encouraged to cover their ears. When the glass shatters, the glass fragments tend not to be ejected with great force but rather to fall straight down. Even so, it is best to perform the demonstration at least ten feet away from the audience or to use a transparent protective screen and to wear safety goggles. A tray underneath the glass will catch most of the glass fragments, but the broken pieces should be collected promptly and disposed of.

REFERENCES

1. W. Walker, Phys. Teach. **15**, 294 (1977).

2. H. Kruglak, R. Hiltbrand and D. Kangas, Phys. Teach. **28**, 418 (1990).

3. H. Kruglak, Phys. Teach. **17**, 49 (1979).

3.8

Flame Pipe

A pipe several meters long filled with natural gas and connected to a loudspeaker produces a flame whose intensity varies with position along the length of the pipe.

MATERIALS

- specially constructed pipe
- source of natural gas, methane or propane
- matches
- loudspeaker
- audio amplifier (few watts)
- sine wave generator
- meter stick (optional)

PROCEDURE

A brass or copper pipe about 5 cm in diameter and 2 meters long contains several hundred small-diameter holes in a line along the length of the pipe. A downspout pipe will also work. One end of the pipe is sealed and fitted with a nozzle through which natural gas, methane or propane can be admitted. On the other end of the pipe is mounted a loudspeaker so as to make a reasonably tight seal. The loudspeaker is connected to the output of an audio amplifier whose input is connected to a sine wave

generator of variable frequency.

The gas is turned on, and after waiting a few seconds for the air to be expelled from the pipe, the gas is ignited with a match where it exits the holes along the pipe. This produces a nearly continuous wall of flame several meters long. A considerable throughput of gas is desired, and the effect is best when viewed in subdued illumination. The frequency and amplitude of the sound emitted by the speaker is adjusted until a clear standing wave pattern is seen in the flame. At certain frequencies, the wave will resonate in the pipe, and the required amplitude will be quite small. More amplitude is usually required as the frequency is increased, but one should be careful not to burn out the speaker. As the frequency is raised, the distance between adjacent wave crests will decrease. The demonstration can also be done with other wave shapes such as square waves or the waves corresponding to speech and music.

DISCUSSION

The fundamental principle illustrated in this demonstration is the compressional wave nature of sound. Where the average pressure is greatest, the gas emitted through the holes is greatest, and the flame is most intense. The flame intensity is low where the wave displacement is maximum as a result of the Bernoulli effect. The information resembles that displayed on an oscilloscope but is more direct and less mysterious.

The variation of wavelength with frequency (or pitch) of the sound is easily illustrated. With a meter stick one can show that the product of frequency and wavelength is a constant equal to the velocity of sound (about 460 m/s in methane at 20°C). Note that the distance between the nodes of the flame is half a wavelength.

This demonstration also illustrates the resonant frequencies of a closed organ pipe. The closed end of the pipe is a displacement node (the air cannot move there) and thus is a pressure antinode (maximum) since the displacement and the pressure are 90° out of phase with one another. Conversely, the end driven by the speaker is a displacement antinode and a pressure node.

A resonance will occur in the pipe when the length of the pipe is an odd multiple of a quarter wavelength. At resonance the pressure variation within the pipe becomes quite large, and a relatively low sound intensity results in a large variation in flame intensity. As the pipe heats up, the sound velocity changes (proportional to the square root of the temperature in Kelvins) and a slight readjustment of the frequency is required to maintain resonance. Variation in the air-to-gas ratio also causes the resonance to drift.

HAZARDS

The flame should be placed well away from any materials that it might ignite. The pipe becomes quite hot after prolonged operation and requires some time to cool off. If the flame is lit too quickly after the gas is turned on while there is still air in the pipe, the flame could be sucked into the pipe, destroying the

speaker. The substitution of other flammable gases such as ethane or hydrogen is not recommended.

3.9

Ultrasound

Various sources of sound with frequencies above the range of audibility are used to illustrate the distinction between a physical sound wave and the perception of sound.

MATERIALS

- ultrasonic dog whistle*
- variable-frequency audio oscillator with speaker (optional)
- ultrasonic cleaner (optional)**
- ultrasonic microphone and oscilloscope (optional)

*Available from most pet supply stores

**Available from Carolina Biological Supply Company, Cole Parmer Instrument Company, Edmund Scientific and Sargent-Welch Scientific Company

PROCEDURE

This demonstration can be introduced with the old question of whether a tree falling in a deserted forest makes a sound. The answer depends upon whether sound is defined as a physical wave or the perception of that wave by the ear and brain of a living creature. Sources of ultrasound provide a means to produce a sound wave that is not perceived by humans as being a sound.

Probably the least expensive and most convenient source of ultrasound is a dog whistle. A dog trained to come to the sound of the whistle who interrupts the presentation when it is first demonstrated will get the attention of the audience. Such whistles can usually be adjusted from a frequency that is barely audible to most people to well above the range of hearing. Ultrasonic pest repellers are sold that emit a very intense burst of ultrasound every few minutes.

A variable-frequency audio oscillator with a high-fidelity audio amplifier and good quality speaker (tweeter preferred) can be used to explore the whole range of audible frequencies. One can point out that high frequency audible response deteriorates with age (a condition known as "presbycusis"). If the audience consists of many ages, one could perform an experiment on this. One reason that children and

small animals are more responsive to high frequencies is a simple matter of size scaling. An ear with smaller components is more sensitive to the shorter wavelengths of a high-frequency sound.

An oscilloscope with microphone can be used to prove that there is a sound present. Most microphones have low sensitivity to ultrasonic frequencies, but with a sensitive oscilloscope they can usually be made to indicate the presence of a sound. A high-pass filter between the microphone and oscilloscope is of considerable help. A small capacitor will often suffice. Ultrasonic microphones can also be purchased.

Another common source of sound above the audible range is an ultrasonic cleaner which usually consists of a vat of liquid with an ultrasonic transducer that creates sound waves in the liquid. Typically the liquid is some organic compound capable of dissolving oil and grease. Jewelers often use ultrasonic cleaners to clean rings and other jewelry. A freshly opened plastic bottle of carbonated beverage partially emersed in the liquid can be made to erupt and spray its contents several feet into the air.

Because of the short wavelengths, ultrasonic transmitters and receivers are well suited for the display of interference and diffraction effects usually demonstrated with microwaves or light. This helps to dispel the mystery of why, if light and sound are both waves, we can hear around a corner but we cannot see around one.

DISCUSSION

A sound wave is a compressional disturbance that will propagate in most materials. The range of human hearing is usually taken to be about 20 Hz to 20 kHz, although it is more accurate to say that the ear is most sensitive to frequencies around 4000 Hz and becomes progressively less sensitive at higher and lower frequencies requiring greater sound intensity to produce a detectable response. The product of the frequency and the wavelength is equal to the speed of sound in the medium. In air at 20°C, the speed of sound is 343 m/s, and so the corresponding wavelengths are 17.15 meters to 1.715 cm, respectively. Typical ultrasonic sources produce frequencies from about 20 to 100 kHz, corresponding to wavelengths as short as a few millimeters.

The reason it is difficult both to produce and to detect ultrasound is that something has to move to create or to respond to the sound wave. The inertia of the object (eardrum or whatever) prevents it from moving appreciably during the short period of the wave. Even so, the human ear is a remarkable instrument, responding to displacements of the ear drum that are hardly larger than the diameter of an atom!

HAZARDS

It is possible to damage one's hearing by exposure to sufficiently intense ultrasound even though the sound is inaudible, but it is unlikely unless one places an ultrasonic transducer directly in the ear, which is definitely not recommended. Internal damage can occur if one places a finger or other part of the body in the vat of an ultrasonic cleaner that is turned on.

[Next Chapter] [Previous Chapter] [Introduction] [Server Home]

J. C. Sprott

Christian Andreas Doppler

1803 - 1853

Click the picture above
to see two larger pictures

Full MacTutor biography [Version for printing]

List of References (13 books/articles)

Mathematicians born in the same country

Show birthplace location

Honours awarded to Christian Doppler
(Click below for those honoured in this way)

Lunar features **Crater Doppler**

Previous	(Chronologically)	Next	Main Index
Previous	(Alphabetically)	Next	Biographies index

JOC/EFR © April 1998

The URL of this page is:
http://www-history.mcs.st-andrews.ac.uk/Mathematicians/Doppler.html

Elliott Sound Products **Project 71**

Subwoofer Equaliser
(Using The Linkwitz Transform Circuit)

Rod Elliott (ESP)

PCBs are available for this project. Click the image for details.

Introduction

The Linkwitz transform circuit is a hugely flexible way to equalise the bottom end of a sealed loudspeaker enclosure. Unlike the "EAS" (Electronically Assisted Subwoofer) project or the ELF™ systems, a speaker that is corrected using this method is flat from below resonance to the upper limit of the selected driver. The low frequency rolloff point is determined by the parameters of the transform circuit. Should the enclosure size be too small and cause a lump in the response before rolloff, this is also corrected. A conventional active crossover network is then used to divide the subwoofer signal from the main channel signals.

For a detailed look at how the circuit works, please click here to see the article that describes the operation of the circuit.

The original Linkwitz Transform spreadsheet was presented by TrueAudio (2), and is reproduced here with their permission. One of my readers added an extremely useful extra page that calculates the driver response in the box, and I added the ability to use litres instead of cubic feet if desired. The transform circuit was invented by Siegfried Linkwitz (1), and is another of his gifts to the audio world (the other is the Linkwitz-Riley crossover).

The Linkwitz Transform spreadsheet can be downloaded here, or from the Downloads page.

Driver Selection

A quick word is warranted here, to allow you to determine if the speaker you have will actually work in a small sealed enclosure. The Linkwitz transform circuit (or EAS principle) will allow any driver to extend to 20 Hz or even lower. A good quick test is to stick the speaker in a box, and drive it to 50 or 100W or so at 20 Hz - you should see a lot of cone movement, a few things will rattle, but you shouldn't actually hear a tone. A "bad" speaker will generate 60 Hz (third harmonic) - if you don't hear anything, the speaker will

work in an equalised sub.

If a tone is audible, or the speaker shows any signs of distress (such as the cone breaking up with appropriate awful noises), then the driver cannot be used in this manner. Either find a different driver, or use a vented enclosure.

Description

The schematic for a complete equaliser is shown in Figures 1 and 2, and considering the flexibility of the circuit, it is not at all complicated. The complex part would normally be determining the component values based on the speaker driver's response in a given enclosure. This is simplified by the spreadsheet, which requires only a small amount of input from you to be able to calculate the box response, and the equalisation needed to achieve the desired response.

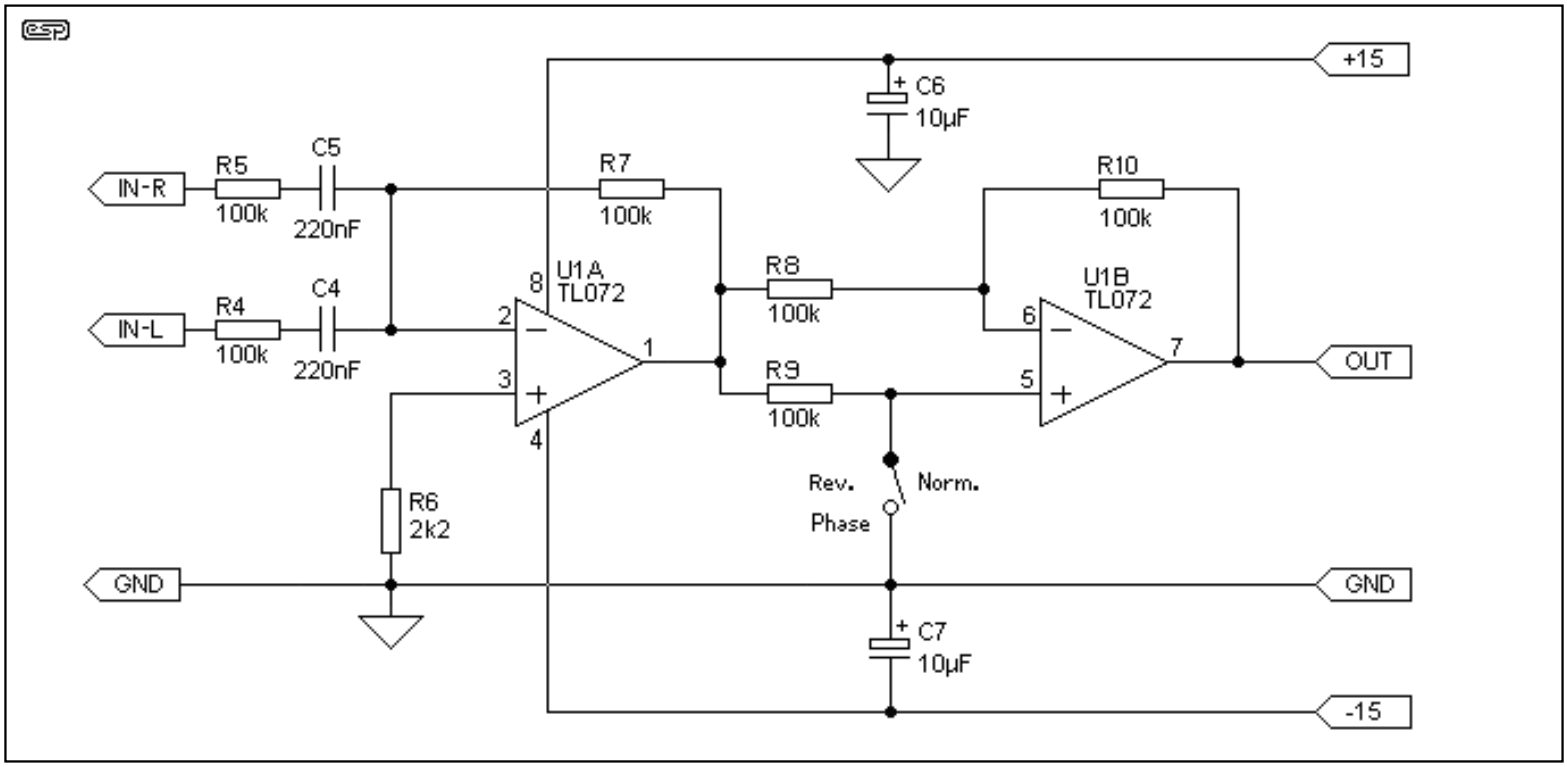


Figure 1 - Input Mixer and Phase Switch

A simple mixer circuit is shown in Figure 1, and this is simply be added to the front end of the circuit in Figure 2. The mixer stage completely isolates each channel from the other, since the summing point is a "virtual earth" because of the feedback. The second opamp is used only as an inverter, so the polarity of the signal may be switched to find the optimum setting. When the switch is open, the phase setting of the equaliser is non-inverting (both the input and Linkwitz transformer are inverting). Closing the switch reverses the phase.

Figure 2 shows the circuit of the Linkwitz "transformer" (for a single channel). The values of the unmarked components are determined using the Linkwitz Transform calculator spreadsheet (see below for more information). This circuit must be driven from a low impedance, such as the output from an opamp (as shown in Figure 1). The output capacitor as shown is a bipolar electrolytic, but a conventional polarised electro can be used if you prefer. The polarity will be different with different

opamps, so measure the DC output voltage (it will only be a few millivolts) to determine the correct polarity.

If you don't want to use the mixer of Figure 1, to create a mono signal from a stereo source, a simple resistive mixer may be used. A resistive mixer requires that the source impedance from each preamp output is quite low to prevent crosstalk. If you are using a typical valve preamp, I suggest the use of buffer amplifiers prior to mixing, as the output impedance of most valve preamps is too high to allow resistive mixing without introducing crosstalk. This is not a problem if the summing inputs are used. In any case, the Linkwitz transform circuit will need to be driven from an opamp to ensure low source impedance.

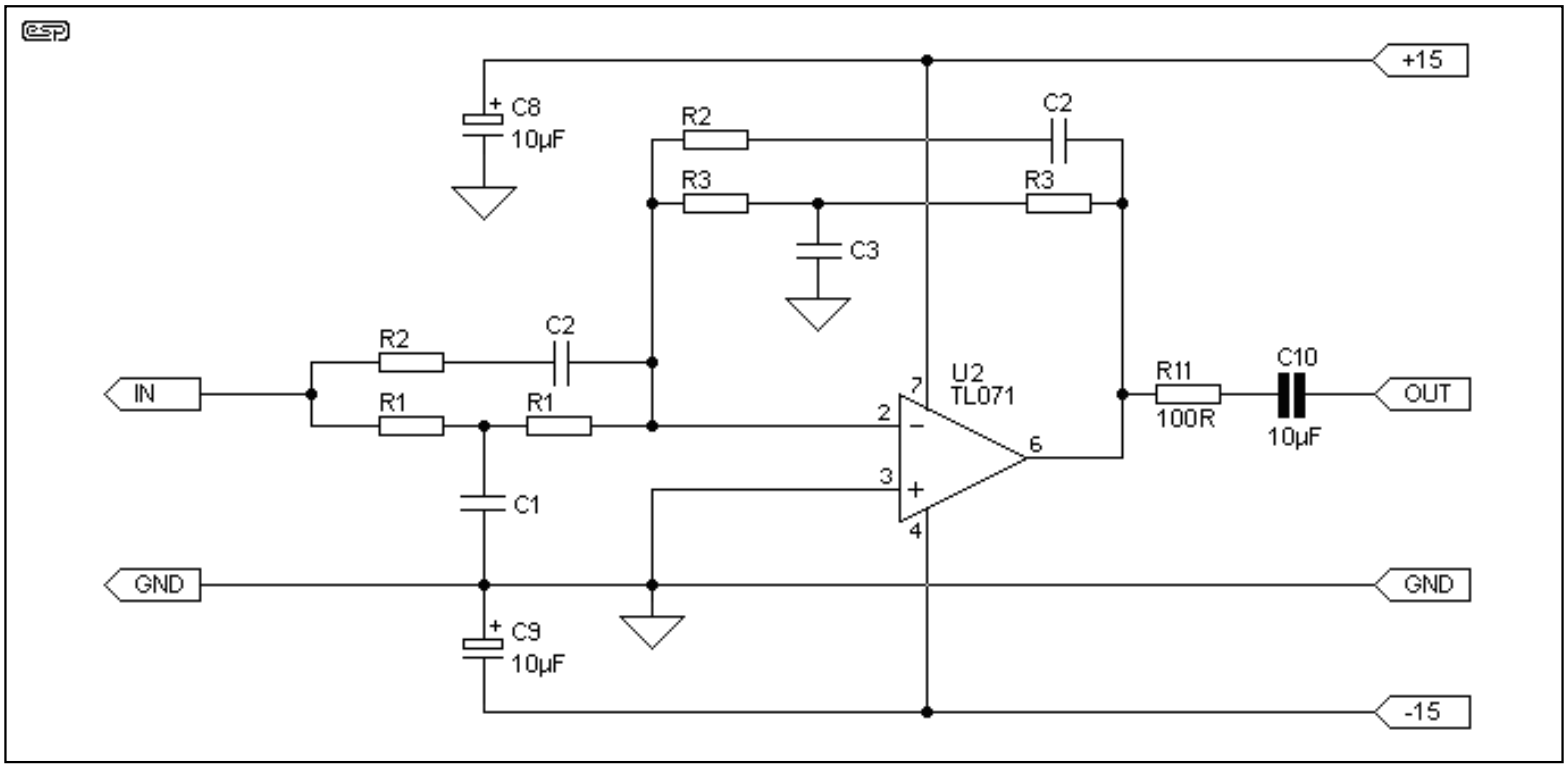


Figure 2 - Linkwitz Transform Circuit

When designing the circuit from the spreadsheet, the dire warnings about DC gains above 20dB may be ignored, as the circuit shown has zero gain at DC. The input circuit has been designed to roll off (i.e. is 3dB down) at 7Hz, with a 6dB/Octave slope. The spreadsheet does not show this, but it must be considered regardless.

I will instead issue my own dire warnings about excessive (greater than 20dB) gain at any frequency. Not allowing for the reduced radiating efficiency of any driver at very low frequencies, this is a simple case of determining the power requirements to obtain a flat response down to (say) 20Hz.

Let's assume that the power needed at 100Hz is 25W to keep up with the main system. if a given design shows that there is a 3dB gain at 50Hz, this means that 50W is needed for this frequency. As frequency reduces, the gain increases, and each 3dB of additional gain demands double the power. A mere 6dB of gain requires 100W, 9dB means 200W, and 12dB of gain means that 400W will be needed. For a gain of 15dB, we are now into serious power requirements, with a system power of 800W.

If you make the box as large as practicable, the amount of boost is reduced considerably - as always, there is a tradeoff between the allowable size of the enclosure, and how much power you are willing (or can afford) to use. The power limits of the driver itself must also be considered, and although most loudspeakers can tolerate transient signals higher than their rated power, they will have excessive distortion if excursion limits are exceeded.

Fortunately and perhaps surprisingly, the very low bass signals typically recorded are not that loud, so you will never need the full amount determined from the estimations above. One thing that is essential is to use a speaker with high efficiency - the higher the better.

Loudspeaker Driver Selection

Selecting a high efficiency driver with a low free air resonance and a relatively small equivalent volume (Vas) means that the minimum amount of equalisation can be used. Again, each 3dB increase in efficiency means half the power is needed for the same SPL, so a speaker with 92dB/W/m is preferred over one with 89dB/W/m (all other things being equal), as only half the power is needed for any given SPL with the more efficient driver.

I will not suggest drivers, and it is entirely up to you to find suitable components based on manufacturer's details. I suggest that you look at loudspeakers intended for car sound systems, as these usually have a reasonable efficiency and low Vas. The range is enormous, and there are many excellent drivers that will not cost the earth. There are also many other excellent drivers that *will* cost the earth - your choice.

Apart from efficiency, Vas, resonance (etc), also consider the excursion limits of the driver. A speaker that has Xmax of 20mm will perform much better (and with a lot less distortion) than another similar driver with an Xmax of 10mm. In much the same way, a 380mm (15") speaker will produce more SPL with less excursion than a 300mm (12") speaker can. Also consider using multiple smaller drivers - two 300mm drivers have more cone area than a single 380mm speaker, and can move more air.

A simple test for drivers is to apply a signal at 20Hz from a clean audio oscillator. Using an amp of about 100W or so, and with the driver mounted in an enclosure of no less than about 28 litres (1 cubic foot), increase the level until the amp starts to clip - this will be immediately audible! Just below the clipping level, listen carefully for any audible frequency that is not 20Hz (which itself is virtually inaudible). In particular, if you hear 40Hz or (more likely) 60Hz, the speaker is distorting, and generating harmonics. With a good driver, you should be able to see the cone rushing back and forth, but should only feel the air movement - no audible harmonics should be heard.

An Example Calculation

Let's look at a hypothetical driver with the following characteristics (these are inserted in the "Box" page) ...

Measurement	Symbol	Value
Free air resonance	fo	24 Hz

Total Q	Qts	0.38
Equivalent volume	Vas	134 litres
Box volume	Vb	28 litres

We would like to get to 20Hz (-3dB), so insert these values into the spreadsheet. For Q(p) (total Q of transformed" system) of 0.8, we get a very small rise before the rolloff starts, which extends bass response marginally.

Now, select a value for C2 - this will determine the values for the remaining components, and is a compromise. As its value is increased, the resistances needed are reduced (this is good), but the value of C1 increases (sometimes to rather large values - this is bad).

I have been asked for definitions of "good" and "bad" in this context, so here they are ... "good" is resistance in the range of about 2.2k up to around 220k or so. Lower resistance loads the opamp, and higher resistance may increase noise levels. "Bad" capacitance values are those that are big and expensive, and if possible, I suggest that you keep the maximum capacitance below 2.2uF or so. Higher values will not hurt anything, but are bulky and expensive, or (in the case of non-polarised electros) may be small and cheap, but have a wide tolerance so accuracy will be compromised.

In the end, you will have to settle on values that are probably outside these goals, and in the end, noise is not really a big problem at low frequencies. The primary aim is to get values that are available - designing the perfect circuit is pointless if it won't do the job.>/p>

With the selected value of 56nF for C2, the following values are calculated for the remaining components ...

Component	**Calculated Value**	**Use ...**	**Actual Value**	**Error**
R1	8.44 k	8k2 + 220R	8.42 k	0.24%
R2	36.98 k	33k + 3k9	36.9 k	0.22%
R3	70.32 k	68k + 2k2	70.2 k	0.17%
C1	1.905uF	1uF + 1uF	2 uF	5.0%
C2	56 nF	56 nF		0%
C3	228 nF	220 nF		3.5%
Maximum gain	18.41 dB			

It will be necessary to use resistors / capacitors in series (or parallel) to obtain the required values - this can often be quite irksome, but only needs to be done once. A final error of up to 5% is almost certainly going to be OK, as the speaker characteristics will usually vary much more than this, and our hearing is incapable of resolving such small errors - especially when the room acoustics interact with the total system!

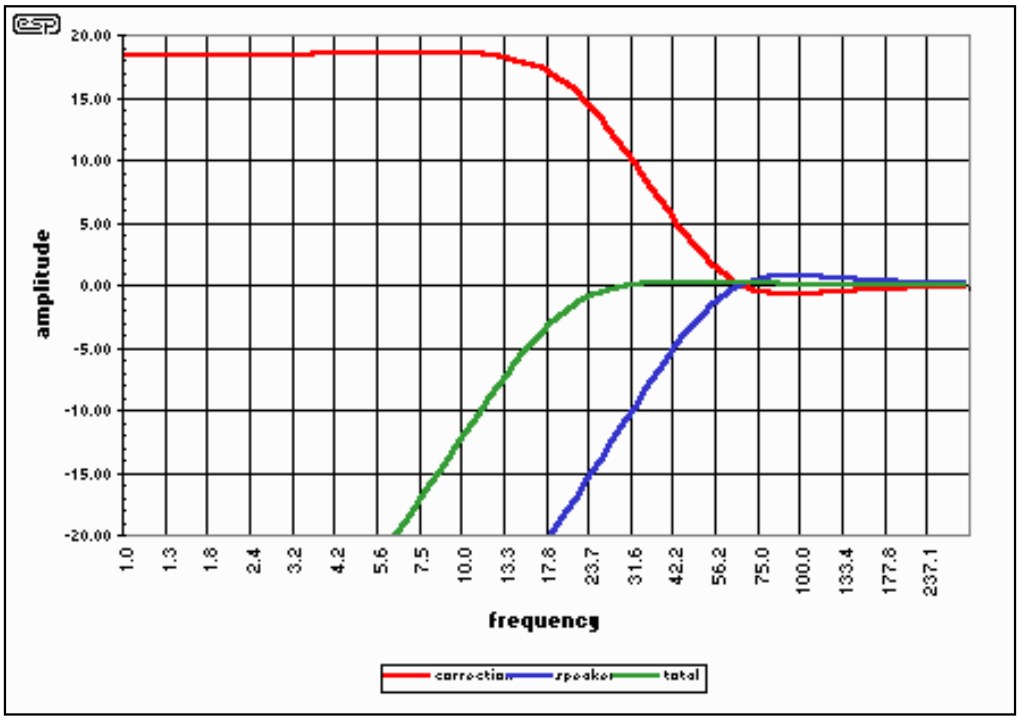


Figure 3 - Response Curves

Note that the maximum gain is over 18dB - this is rather high, but the energy levels at such low frequencies are actually very low as well, so although it may seem that an enormous amount of power will be needed, this is probably not really true in reality.

To work this out, let's have a look at the speaker sensitivity. We will make an assumption that it remains constant, although this is rarely true as the frequency falls, due to the radiating efficiency of the driver itself. The assumption will work out well enough in practice (he says from experience).

Assume a quoted efficiency of 90dB/W/m as a starting point. For "normal" high level listening, an SPL in the listening room might be around 100dB, which represents a power of 10W. Remember, this is worst case, and assumes that the LF energy level is the same as at upper frequencies. For all practical purposes, the lowest frequency is about 25Hz - our ears respond very poorly to anything lower.

> For what it is worth, a good source of a 25Hz signal is the heartbeat at the beginning of Pink Floyd's "Dark Side of the Moon". This should rattle windows, and be felt as much as heard.

So, 10W to get 100dB (which is pretty loud), and we have a 15dB boost at the 25Hz point. Since each 3dB increase involves doubling the power, we get the following ...

First, the power needed to get to 100dB ...

Power	1W	2W	4W	8W	10W
SPL	90dB	93dB	96dB	99dB	100dB

Now, the power to maintain 100dB as the speaker rolls off and the equalisation compensates ...

Power	10W	20W	40W	80W	160W	320W	640W

Boost	0dB	3dB	6dB	9dB	12dB	15dB	18dB

This is not entirely correct with music signals, because the above figures are for steady state signals, and music is transient in nature (well, most of it, anyway :-). Peak power levels may be much higher at some frequencies, typically around 70-80Hz where kick drums make most of their noise, but the amp has copious reserve power at these frequencies. The underground railway rumble in that otherwise wonderful orchestral recording will be at a very low level, and most music has very little below 30Hz. If you enjoy pipe organ music, I suggest two to four 380mm drivers in a reasonable sized box (around 50 litres per loudspeaker), and equalised as described here. To get realistic response to 16Hz will require about 300W per speaker, for a quoted sensitivity of 90dB (actual figures will vary depending on the drivers selected, of course). This will be truly awesome, and if your dwelling survives, you may be able to defray some of the system costs by acquiring adjoining properties at an advantageous price :-)

The biggest benefit of a subwoofer that goes to such depths is that you will find yourself listening at lower levels than before, since the "feel" that used to require an SPL of perhaps 110dB or more, is now attainable at maybe 95dB SPL. There is no loss of realism - indeed it is enhanced - and your ears will love you for it, and for all the right reasons.

PCB Version Of Circuit

The PCB for this circuit is available. It has been simplified from the circuits shown above, and retains a rudimentary phase switch but no mixer. Mixing can easily be done using a pair of (say) 22k resistors from each channel to the input, provided the impedance is low enough from the preamp to prevent crosstalk.

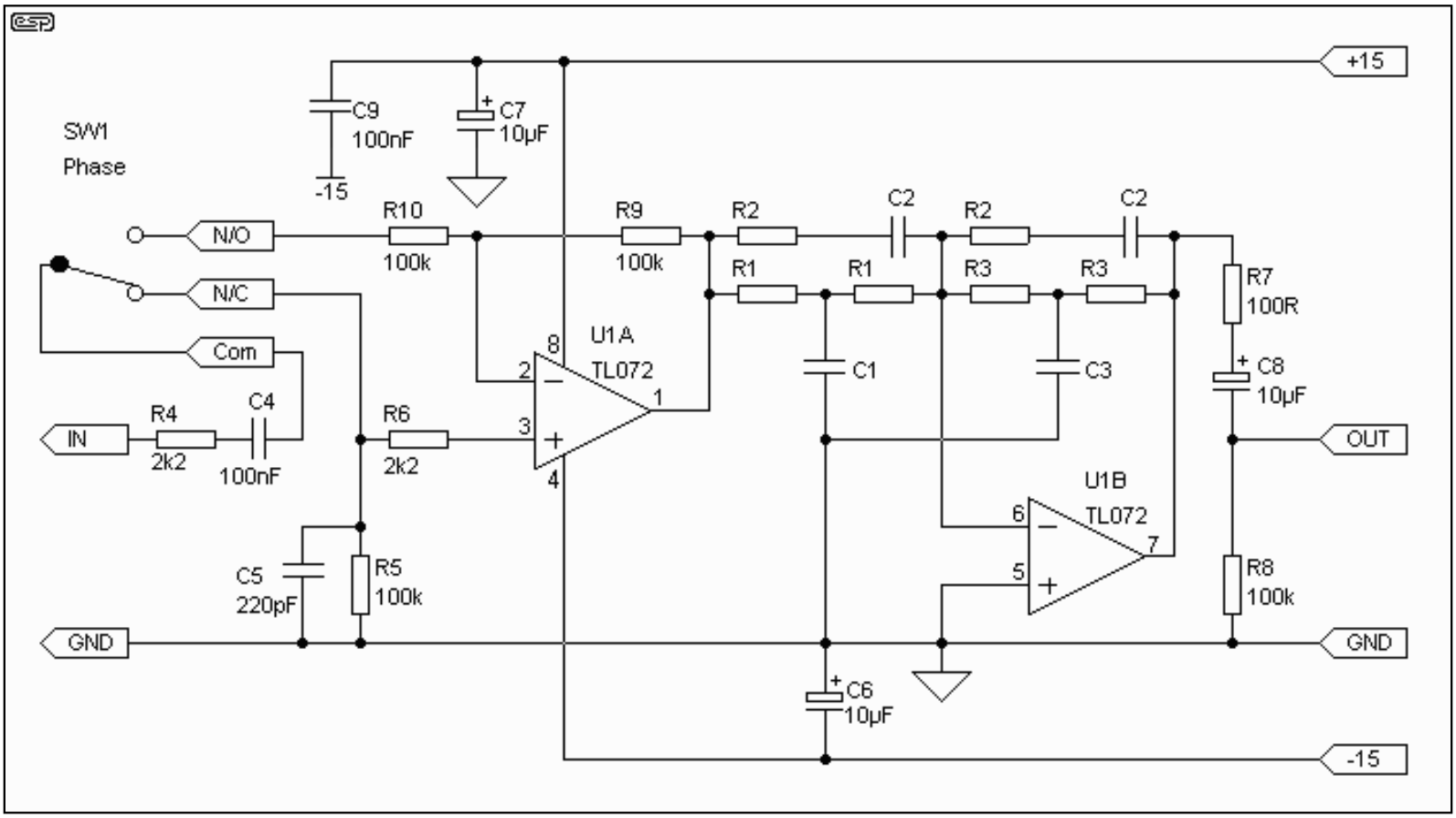


Figure 4 - PCB Version, Complete Circuit

If you don't want to use the phase switch, the Com and N/O (Common and Normally Open) pins can be linked on board to give a non-inverting configuration (note there are two inversions in the circuit). Link Com and N/C for inverting operation.

Because of the DC gain in the circuit, you may find that C8 is incorectly polarised when the circuit is powered on. I suggest that you either use a non-polarised cap, or measure the DC, and determine the correct polarity before this cap is installed. Because of the different way opamp inputs are configured, the output will be positive or negative, and I cannot predict which way it will go, since constructors will want to use different opamps in many cases. The polarity is correct for the TL072, but will be reversed if (for example) an LM1558 were to be used, since the latter has a bipolar input instead of FETs. The exact offset voltage is also difficult to predict, but using the TL072, should be less than 300mV. Other opamps can be expected to be similar.

This circuit was devised to be as cheap as possible to build, but provides for parallel connection of caps for C1 and C3, and series resistances for R1, R2 and R3, since these will usually be odd values (as shown above).

References

1. Siegfried Linkwitz - Active Crossover Networks for Noncoincident Drivers, JAES, Vol. 24, No. 1, January/February 1976
2. True Audio - The original of the Linkwitz Transform spreadsheet program

Projects Index

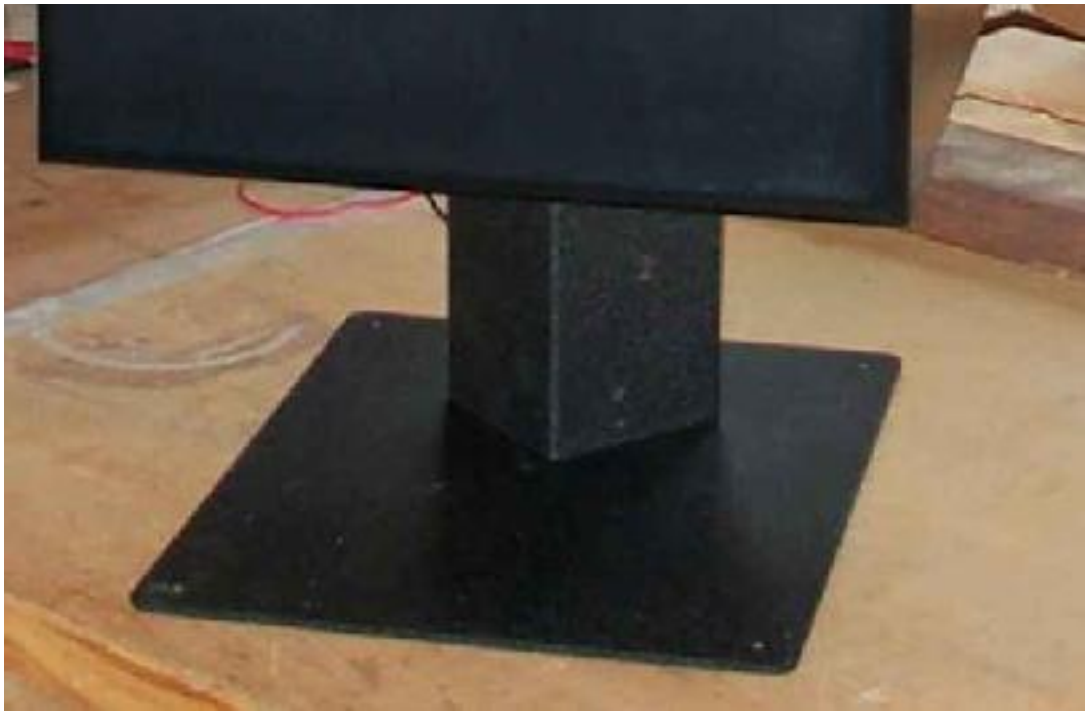
Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000/2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright © Rod Elliott 14 Oct 2000

WALT
FAULKNER

VISATON



BAILEY'S
GEL
GRIP

Elliott Sound Products
PO Box 233
Thornleigh NSW 2120
Australia

Elliott Sound Products **VP103 Hi-Fi Valve Preamplifier**

The **ESP VP103 Hi-Fi Valve Preamplifier** is a premium vacuum tube preamplifier, designed for discerning hi-fi enthusiasts who will appreciate the unique sound of valves. (See photo of prototype.)

Main Index

Contact ESP

VP103 Hi-Fi Valve Preamplifier

Price - AU$1,800 (+ 10% GST for Australian residents)

The design is no compromise, with the following features:

- Massive overload margin - input signals up to 10V rms will not clip the input amplifier
- A timer changes the front panel power indicator as the unit warms up
  - Red - Should not use, not temperature stable
  - Orange - May use, but not completely stable yet
  - Green - Unit is ready for use
- The power supply uses massive capacitance (for a valve circuit), completely eliminating hum on the DC supply, and providing excellent rejection of short-term mains variations as well as providing an extremely low impedance voltage source
  - Fully regulated 12.6 Volt DC heater supply to eliminate induced hum and noise and provide optimum heater operating voltage at all mains voltages
  - Current limited heater supply, to reduce inrush current (typically a major cause of valve heater failures)

- Standard valves are used at the lowest practicable voltage and current, ensuring maximum valve life · Local feedback is used on each amplifying stage to reduce the effects of valve ageing, and reduce output impedance and distortion
- Very low distortion. Even at an output level of 5 Volts rms and a load impedance of 10k, distortion remains below 0.2%
- Frequency response from 10 Hz to 150 kHz (-1dB)
- Channel balance better than 0.5 dB
- Auxiliary mains output for VR102 Phono preamp or other low current device (this must not be used for power amplifiers). NOTE: Aux mains output is switched but not fused
- Rear panel level controls for CD and Tuner inputs, to allow their level to be matched to other components in the system
- All signal paths use traditional point-to point wiring, and not a printed circuit board
- All signal paths are shielded by a ground plane, made from un-etched fibreglass PCB to ensure minimal hum and noise pickup
- Valves are matched to ensure optimum channel balance is maintained
- High quality 1% metal film resistors are used exclusively (except for one 1 watt resistor in the power supply), to provide low noise and long term stability
- All electrolytic capacitors in the signal path are low leakage types, and are bypassed with polyester caps to ensure optimum high frequency performance

---

Please Note

The ***VP103*** is not a production item, and for the time being will be built to order only. If the unit goes into regular production, this will be determined by the response I get to various market probes (this being one of them). A photograph of the interior of the prototype is shown at the end of this section. The final version uses a high quality custom made case with CNC engraved front and rear panels.

This preamp sounds superb - open, with a stable stereo image, and detailed highs. The

specifications are as follows:

Specifications (all measurements relative to 2V rms output unless otherwise stated)

Distortion	< 0.2% (10k load)
Signal to Noise Ratio	> 70dB
Maximum input level	10V rms
Maximum gain	18dB / 10dB*
Tape Output Gain	6dB
Crosstalk (at 1kHz)	- 65 dB
Crosstalk (at 10kHz)	-50 dB
Frequency Response	10 Hz - 150 kHz (-1dB)
Channel Balance	Within 0.5dB **
Operating voltage	115 / 230 / 240 Volts AC

* I have found it necessary to provide two gain settings (switch selectable), since a gain of 18dB is too high for many power amplifiers.
** Channel balance is expected to be better in the production version, but it is not possible to predict by what margin at this stage.

The unit follows the minimalist approach, with the only front panel controls being the power switch and indicator, a 6 position rotary input selector switch and a volume control.

Inputs are provided for:

- CD Player
- Tuner
- Phono (external preamp required)
- DVD Player
- Tape
- Auxiliary

And the outputs are

- Tape (to Cassette, MiniDisk, DAT, etc)
- Main (to amplifier)

Note that the phono input is a simple line level input, since the ***VP103*** has no inbuilt RIAA equalisation amplifier. The ***VR102*** Phono Equaliser is likely to become available shortly after the ***VP103*** is in production, or the user may choose any other RIAA equaliser which has an output voltage of around 300mV to 1V rms.

Rear panel level controls are provided for the CD and Tuner inputs, since these devices rarely have output level adjustment. This allows the gain for each device to be set, so there are no radical differences in sound level when changing signal sources.

In anticipation of the ***VR102*** Phono Equaliser, there is a switched rear panel mains output. This is not to be used for the power amplifier (however tempted one may be), as it is low current only (<2 Amps max.).

All input and output RCA connectors are gold plated.

The case design for production is not complete, but will be all aluminium, powder coated in dark grey, with ventilation slots on the top and bottom. The front panel is black anodised aluminium, with all markings engraved.

Dimensions are fairly standard for hi-fi equipment, being:

- 44mm (1 3/4") High (1RU)
- 420mm (16 1/2") Wide
- 300mm (11 3/4") Deep

Photo of the interior of the prototype VP-103

The 3 dual triodes can be seen on the right, and are mounted on a piece of PCB for shielding only. The power supply occupies the left hand side, and the heatsink is for the 12.6V DC heater regulator.

Input switching is all performed at the rear (hence the long extension shaft), close to the input connectors. The two pots at the rear are for the tuner and CD inputs, to allow them to be adjusted to match the level of other sound sources.

As this is the prototype, the case is not a final design, but an off the shelf 1 Unit rack cabinet (with the ends cut off). The final design is completed (but not yet available), and is far more robust and looks better, too.

Some Musings

There is something of a resurgence of valve (or tube) amplifiers of late, and it is worth while taking a look at the reasons. There are those who will simply buy whatever is in fashion, but they are a small minority, so what is about valves that has experienced hi-fi buffs waxing lyrical about technology that is (or was thought to be) well past its use-by date?

I don't have the answer (unless 42 is acceptable), but during the design of the ***VP103***, some very interesting facts came to light (or to be more precise, I was reminded of a few things which had faded somewhat).

Distortion

Everyone knows that valves have greater distortion than transistors or integrated circuits, but is this a bad thing? Apparently not, provided the levels are kept low and are predominantly second harmonic. This is a characteristic of valves - if properly designed - and it seems to be accepted as a part of that 'valve sound'.

Where it gets interesting is in overload. All transistor and IC amplifiers are limited by relatively low supply voltages (typically +/- 15 Volts), and at some point just below the power supply voltage they clip - not gently or progressively, but hard and fast. So fast in fact, that harmonics are immediately generated well beyond the limits of our hearing.

"Ah, but a preamplifier should never clip - there should be plenty of headroom for all practical input sources", I hear. I agree with this, and in my own testing have never seen a properly designed preamp clipping internally (unless you do something silly, like turn the power amp gain right down and then drive the preamp into distortion).

However, it stands to reason that 20dB of headroom might just help with transients, and perhaps this is one of the deciding factors?

Slew-Rate

Then of course there is 'slew-rate', which is the speed that the output can change. This is typically measured in volts per microsecond (V/us), and most modern IC devices have a slew rate which is greater than 10 V/us. This is far greater than is needed for audio, where a measly 0.5 V/us is actually sufficient even for a low power amp - except it sounds awful.

Valves have no slew-rate limit as such, since there is no compensation capacitor used (although the internal grid to plate capacitance does have a similar effect, but is much less savage). Instead, the signal simply rolls off smoothly once the upper limit has been reached. At no frequency or amplitude will a valve convert a sine wave into a triangular wave, as will many opamps and power amplifiers.

Negative Feedback

When a preamp is built using IC or discrete opamps, there are actually tens to several hundred active devices in the signal path. Each of these contributes it own little bit (gain, noise, distortion, etc), and the overall signal is tidied up using negative feedback. This works so well that with the latest opamps, distortion cannot be measured with a simple noise and distortion measuring set. Indeed, the signal source is likely to be an order of magnitude (or several) worse than the device under test.

Perhaps it is this clinical treatment of our sounds that is somehow to blame - can an amplifier be too good, too clean? I don't know, but there is a marked difference between a 'solid state' and a good valve preamp - with the latter having more 'air', and a sense of openness that IC and transistor units don't. Even here, it is difficult to be objective, since I used an opamp preamp for quite a few years, and it is quite possible that the difference is imagined.

Most valve circuits use little or no negative feedback in the preamp stages. The ***VP103*** is an exception, but even here, the level of feedback is very low. Valves have a relatively low gain (especially compared to opamps, which have a low frequency gain of over 100,000), so massive feedback is not an option, since there is not enough gain to spare.

Class-A

A valve preamp is pure Class A - it has to be, since it is too difficult (and a completely useless exercise into the bargain) to build a Class B valve preamp. Virtually all IC opamps have a push-pull output stage which spends some of its time in Class B. Perhaps this generates some artefact which I have been completely unable to measure (along with many others), but which is audible. The various laws of acoustics would indicate that any such artefacts would be inaudible (due to masking, among other things), but there seem to be many things in audio which don't seem to obey the law! (At least according to some reviewers and the like.)

Open-Loop Frequency Response

This is one area where the valve is measurably vastly superior to opamps. The gain of an ordinary opamp from DC to (about) 100 Hz is typically 100,000 - but after the compensation circuit is added, this rolls off at 6dB/octave after the 100 Hz point is reached. This means that at 200 Hz, the gain is down to 50,000 and by the time you get to 25 kHz, the gain will be down to a little under 400 (or less). Premium opamps are better, but the same thing still happens - its just not as severe.

What this means is that the input signal gets less and less feedback compensation as the frequency increases, so the distortion and output impedance will both increase in proportion.

With a valve preamp circuit, the gain is (or should be) pretty constant from the lowest operating frequency (determined by coupling capacitors, which cannot be avoided), up to 50 kHz or so, and will roll off very gently after this. Even if feedback is used, it will be constant over the audio range and beyond, which may account for some of the perceived difference.

Conclusion

I don't know! What I do know is that my valve preamp really does sound better than (or maybe just different from) a transistor or IC preamp. As little as 10 years ago I would have probably taken the opposite view, but upon reading reviews and then deciding to build one just to hear it has changed my thinking.

ESP Home Page

Contact ESP

Page last updated 18 May 2001 - new photo, additional input, and minor changes to text.

The Audio Pages

Elliott Sound Products **Project 82**

Loudspeaker Test Box

Rod Elliott (ESP)

Introduction

Testing loudspeakers is covered in some detail in the passive crossover article (see Design of Passive Crossovers), but it is irksome at best to have to fiddle about with clip leads and components lying all over the workbench. This simple project is intended to make life that little bit easier when you are trying to determine the optimum frequency compensation network for woofers and midrange drivers.

Description

They don't come much simpler than this. A suitable plastic box, a couple of switches, 4 combination binding posts / banana sockets, a handful of capacitors and a wirewound pot complete the entire project.

Of all the components, the pot is most likely to cause problems. The suggested value of 30 Ohms was used in my case because I just happened to have one lying around in my junk box, but they don't seem to be easy to get. A very basic tweeter attenuator may work OK, and the value is not that important. Even a resistance as low as 8 Ohms could be used, with a switchable 8.2 Ohm resistor in series so that the range will be from 0 to 16 Ohms (near enough). I shall leave it to the individual to determine the best way to achieve the required range, which is typically 0 to 20 Ohms.

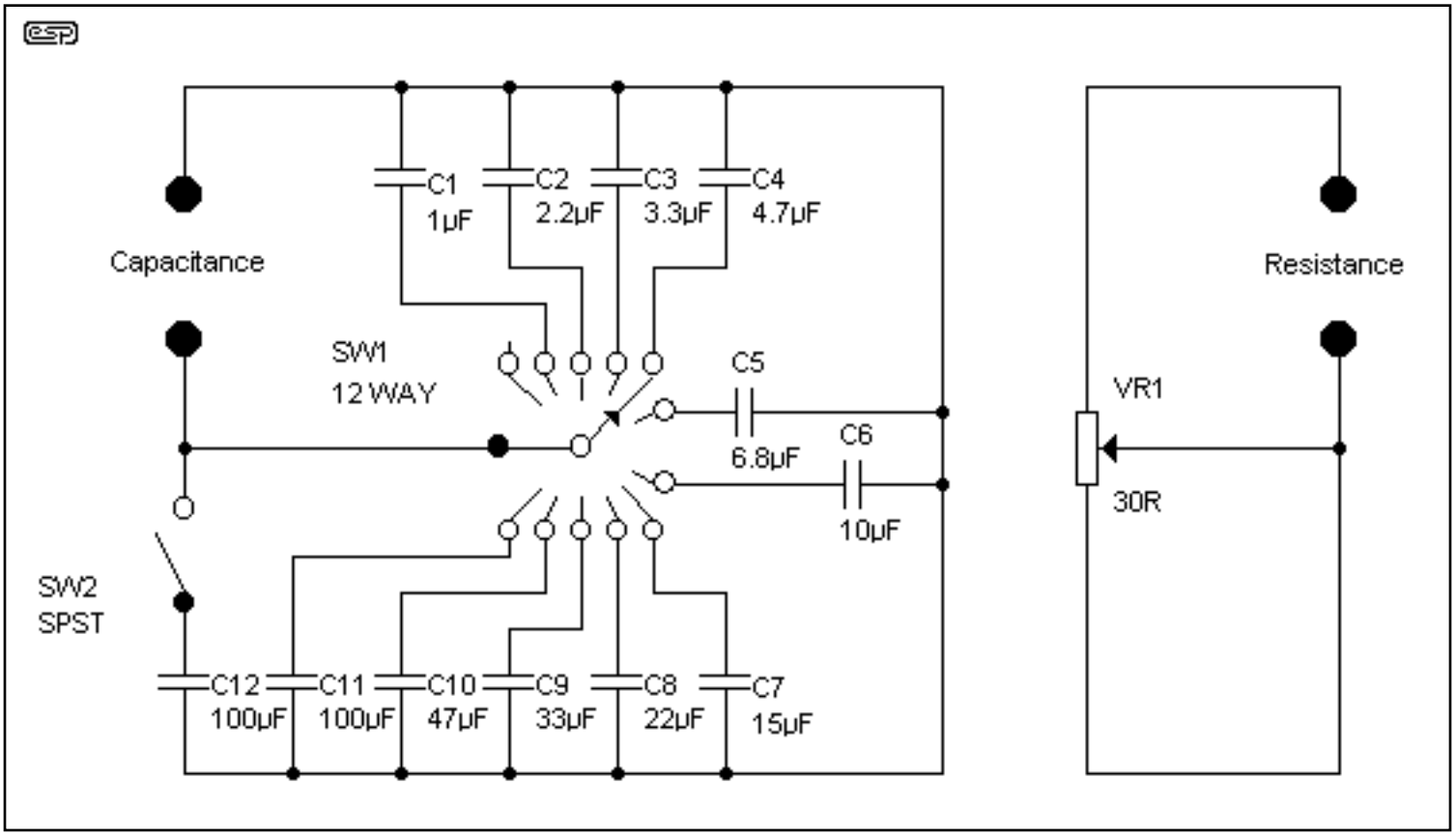


Figure 1 - Complete Circuit of Test Box

The switch SW2 and associated 100uF cap is optional, and increases the range up to 200uF - in most cases this won't be necessary though. The caps are all bipolar electrolytics, and this is perfectly OK, since they will not be used at high power, and their sonics are not important (assuming that you can hear the difference anyway).

To use the box, simply connect it across the speaker - mostly you will want the caps and pot in series, so a lead between the two (or even a switch) makes this easy. From the output of suitable small amplifier, place a 100 ohm or so resistance in series with the output. Sweep the signal frequency across the speaker. Resonance will not be affected, but there will be a "magic" combination of capacitance and resistance that will make the impedance above resonance completely flat. No maths, no spreadsheets, just a quick frequency scan and twiddle a couple of knobs to get the values needed to ensure that the crossover actually will work at the design frequency.

The network derived from the testing will be placed in parallel with the speaker, and prevents the rise of impedance with increasing frequency. This rise is caused by the inductance of the voice coil, and plays havoc with the performance of any passive crossover network.

Impedance Measurements

One of the most difficult measurements to make in audio is impedance, but this is also made relatively simple. Carefully measure the voltage across the speaker at the crossover

frequency - this will (of course) be the same as the voltage anywhere else outside of the resonance affected area of the impedance plot.

Dissconnect the speaker, and without changing anything, connect the pot (all by itself) instead. Adjust the pot until the voltage is exactly the same as you measured before, when the speaker and parallel network were in place. Disconnect the amp, and measure the resistance of the pot - this is the speaker impedance, and it is this - ***not*** the "nominal" impedance - that must be used to calculate the crossove network components.

Don't change the frequency from the generator or the amp level just yet ...

In many (OK, all) cases, it is highly recommended that the tweeter is also measured for impedance, and a resistance used in parallel so that the impedance at crossover exactly matches that of the woofer or midrange driver. Simply connect the tweeter across the amp, with the pot in parallel. Carefully adjust the pot until the voltage is exactly the same as that measured before (this is where a pot that goes up to about 30 Ohms is useful). When the voltage is the same, disconnect the amp and tweeter, and measure the setting of the pot. This is the resistance that should be used in parallel with the tweeter to make everything work properly.

This is a simple procedure that works extremely well, and my test box has been used quite a few times in the short time I have had it. The measurements described in the Design of Passive Crossovers article work well, but this is much quicker, and is "real", in that you can see the effect of varying resistance or capacitance easily. There will nearly always be a compromise in the crossover design, and this little test box lets you see exactly what happens, and select the best possible combination for the speaker drivers you are using.

Return to Projects Index Page

Back to Home Page

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright © Rod Elliott 30 Sept 2001

Cross-Over Network; Air Core Inductor Calculator

6/98 - Jens Moller - http://www.colomar.com/Shavano/inductor_info.html
updated 5/02, 3/04, 5/04 and 07/05

Many people have asked me for a copy of this program. It is not for sale (it remains a free service of this Web Site). This Web Site's goal is provide solutions for Musicians who are making their own gear, and in general, are not interested in the math; rather, they simply need working solutions.

The program is based on a series of equations that are available in many Electrical Engineering texts dealing with Inductors, along with a knowledge of wire characteristics. If you want to duplicate the effort for your own purposes, please consider discussing it with an Electrical Engineer or a College Professor who teaches Electrical Engineering, and see if they will assist you in your efforts. You will need to be fairly good at math and programming to recreate the functions of this application.

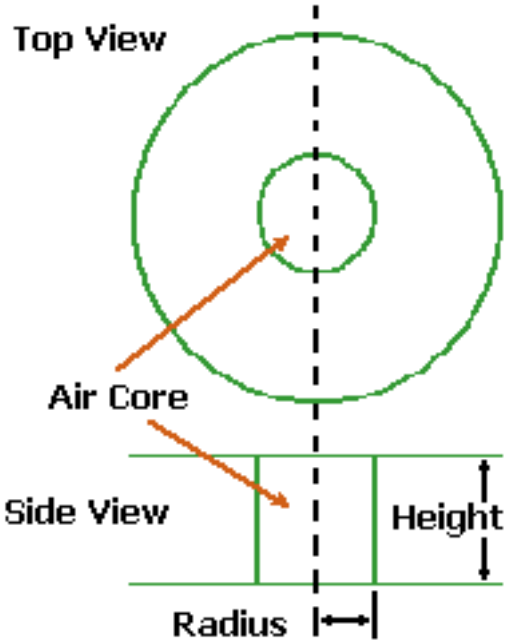
If you are building your own cross-over network, you'll find that the inductors used most frequently for high powered systems are simply large coils of wire. You might be able to save some money by winding your own. You also might be able to more specifically design your cross-over by creating an inductor that is specifically the value you need rather than a standard value.

The following application will generate a design, if you tell it what the inductance needs to be (in milli-Henries - mH). It is intended for solutions within Audio frequency ranges.

Please fill in the following Information.

Inductor mH: [ ] (milli-Henries)

Generate Results | Reset



When you make the form, it should allow you to create an inductor that has this general shape. It may be possible to find pre-made forms, but you really don't need one once the wire has been wrapped and secured.

You'll need to create a form (this could be made of plastic or wood; what ever you have available) to wind your inductor in. You'll need to be able to have the correctly sized opening to wind the core onto (the program gives that as Radius - multiply this by 2 to get the diameter), and you'll need to be able to brace the form to keep the windings together at the right width.

Once wound, use plastic/ nylon Wire Ties, or electrical tape, to hold the inductor together. These can become quite warm when in use. The DC Resistance is an indicator of how much power these will soak up - the lower the DC Resistance the better. For example, if you create a 10 mH inductor using 26 gauge wire, its DC Resistance is 8.27 Ohms - which is approximately the same as the 8 ohm impedance of the Woofer - this implies that the Inductor will absorb an equal amount of power as the Woofer will. If you try to pump 200 Watts into this, the Inductor will dissipate 50 percent (or 100 watts) and will most likely fail in short time. Using 14 gauge wire changes the DC

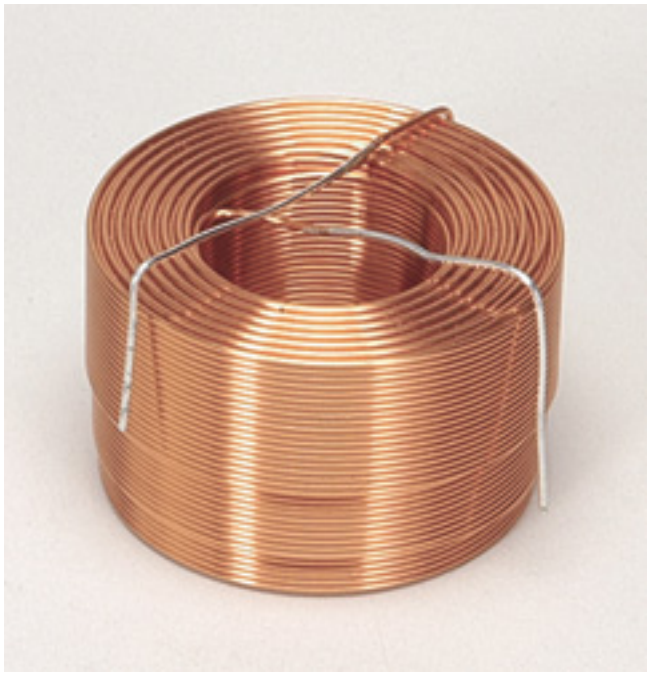
Resistance to .90 Ohms, and instead, the Inductor absorbs only approximately 10 percent of the power (or 20 watts). It still gets warm, but it won't fail catastrophically.

Use transformer wire that has a lacquer finish. Do not use wire that has plastic sheathing around it - this is not appropriate for winding these Inductors.

The program rounds up or down to the next full turn.

For assistance in determining cross-over points, See: First Order Crossover Networks and Second Order Crossover Networks

The end results will looks something like these:



You may need a source of this type of wire (it needs to be solid wire that has a thin coating that insulates it electrically). This is normally called 'magnet wire' or 'bell wire'. For your reference, here are the first three sites that came up when I looked for 'magnet wire 18 gauge' in a search engine (I have no association with these at all):

- http://www.parawire.com/18gaugemagnetwire.shtml
- http://www.powerwerx.com/magnet-wire.html
- http://www.elexp.com/cbl_mg30.htm

This one was suggested by someone who buys lots of wire:

- http://www.mwswire.com/

I often buy my wire at electronic surplus houses - you can get exceptional deals there if you are lucky.

All Wire gauges are AWG - see Comparing SWG to AWG - Wire Gauge for more information.

Return to Shavano Music Online Home page

Problems? Comments? Contact Jens Moller.

Barry's Inductor Simulation

LC Time Simulation Home Up Site Map Wire Loop

This is a Java simulation of an air-core inductor. You should see an applet (below) with slider controls to select the coil's dimensions and wire size. (Trouble? See below.) Note this is only the coil and does not include a projectile.

Inductor Simulation

This program calculates inductance using Wheeler's Formula:

$$L\ (uH) = \frac{0.8 * (N*A)^2}{(6A) + (9B) + (10C)}$$

where N = number of turns, A = average coil radius, B = coil length, and C = coil thickness. All dimensions are in inches and the result is microhenries. The simulator handles all necessary conversions to metric and millihenries.

The simulator will incidentally give you other handy information, such as the number of turns and the length of wire needed.

Your goal in this simulator is to design a coil with a certain amount of inductance. You should have **already chosen** a target inductance value somehow, perhaps by experimenting with the RLC Simulator program. Here you can try various wire sizes and physical dimensions to see what happens. The idea is to design a coil which can be physically built, and which has a reasonably low amount of resistance.

Q. What is a circular mil?
A. Unit of measure used to define the area of a wire. The area of a circle one one-thousandth (0.001) inches in diameter.

Trouble?

This program requires Java Runtime Environment (JRE) 1.4 or above. You can download it by visiting http://java.sun.com/j2se/1.4.1/download.html and scroll to the section "**Download J2SE v 1.4.1_06**". Look in the "**JRE**" column for your own operating system.

Download and install the JRE, or if you write Java code then get the software developer's kit (SDK) instead.

If my applet still doesn't work, try to open your Java Console window in your web browser. Copy and paste its contents into a message and e-mail it to me for trouble-shooting.

About the Program

The program was written with NetBeans (www.netbeans.org) and IntelliJ (www.intellij.com) using AWT classes. My Java source code is open and **freely available** in these files:

- InductorSim8.java runs the user interface and simulates the coil.
- CopperWire.java is a helper class that describes characteristics of copper wire, using values graciously supplied by WireTron, a manufacturer of wires for electromagnets. Any errors are solely my own.
- CopperWireTest.java is a unit test (based on JUnit) to help keep it working while I make changes.

I want to keep the Internet clean so please report bugs, request features, make suggestions and send me compliments about this simulator. Of course I am also grateful if you send suggested updates in Java source code form!

Last update September 20, 2005 by Barry Hansen ©1998-2005

hits

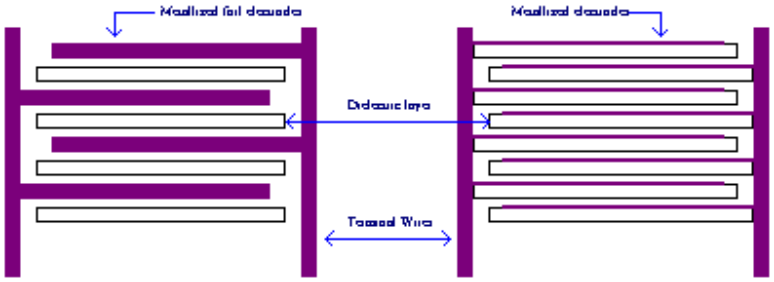
This data sheet is a compendium of facts and recommendations on the usage of capacitors readily available at Caltech. Important parameters to consider and their effects: ESR (power dissipation in the capacitor and useful frequency range), tolerance (ability to plug it into frequency sensitive circuits), temperature/aging drift (capacitance changes in sensitive circuits), ESL (useful frequency range).

Film capacitors: A relatively large family of capacitors, they differ pretty much just in their dielectric properties. Available capacitance ranges from 10pF - 15uF depending upon the actual type of capacitor. Members include polyester (polyethylene terephthalate or Mylar™ from DuPont), polystyrene, polypropylene, polycarbonate, metallized paper, etc.

Temperature coefficients of selected dielectrics:

Capacitor Type	Typical Temperature Coefficient
polyester (Mylar)	600 to 900 ppm/deg.C
polypropylene	-200 ppm/deg.C
polystyrene	-125 ppm/deg.C
polycarbonate	+100 ppm/deg.C

Two distinct subfamilies of the film capacitors are film and foil and metallized film .



film and foil metallized film

Quoting from The Circuit Designer's Companion:

There are two common methods of providing the electrode; one has a separate metal foil wound with the film dielectric, the other has a conductive film metallized onto the dielectric directly. The film and foil construction requires a thicker dielectric film to reduce the risk of pinholes, and therefore is more suitable to lower capacitance values and larger case sizes. Metallized foil has self-healing properties - arcing through a pinhole will vaporize the metallization away from the pinhole area - and can therefore utilize thinner dielectric films, which leads to higher capacitance values and smaller size. The thinnest dielectric in current use is of the order of 1.5um.

Generally, higher power and more precise applications will require film and foil, but the metallized film capacitors are very nice.

Metallized paper capacitors hold a niche high voltage application because when arcing occurs, less carbon buildup occurs, keeping the risk of fire low.

Look at temperature and voltage limitations to aid in trimming down your choices.

Ceramic: The common form is the multi-layer or stacked ceramic (monolithic); single layer also exists (ceramic disk). Physically, the multi-layer looks like the film and foil above, a dielectric stuffed between metal plates. The multi-layer is marginally more expensive than the single layer.

Ceramic capacitors are separated into three common grades of dielectric. The three grades go by the following names:

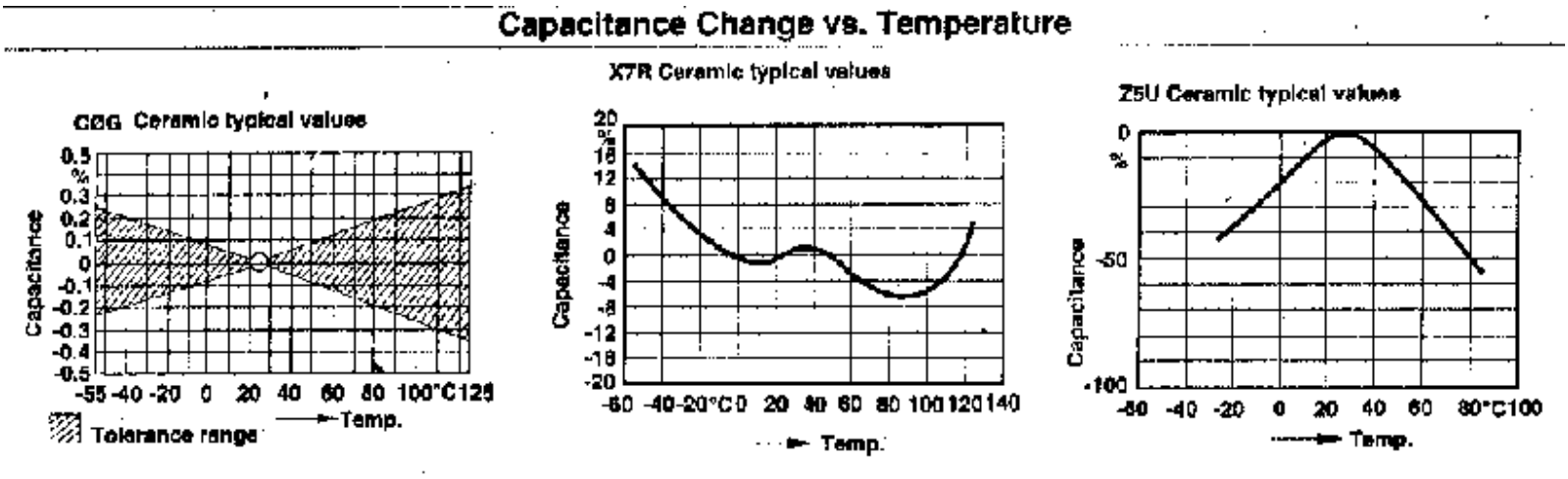
C0G or NP0: best in all features except permittivity; due to the low K (dielectric constant) of the dielectric, these are physically larger than equivalent values in the other grades; also, typically range of values 4.7pF to .047uF. Typical tolerance 5%.

X7R: just a bit more expensive than the low grade Z5U, but improved in tolerance and temperature characteristics. Temperature coefficient is

non-linear, however. Typical range of values 3300pF to .33uF. Typical tolerance 10%.

Z5U: Suffers from a relatively large temperature coefficient. Good points include price and size. These are the classic bypass capacitors. Typical range of values .01uF to 2.2uF. Typical tolerance 20%.

Temperature behavior of the 3 ceramic dielectric grades:



How to use these capacitors?

Z5U: bypass, coupling

X7R: notice the lack of small values; similar to cheaper film capacitors; useful for non-critical timing, coupling.

C0G: very good capacitors; tightly specified in tolerance and temperature. Tradeoff is size.

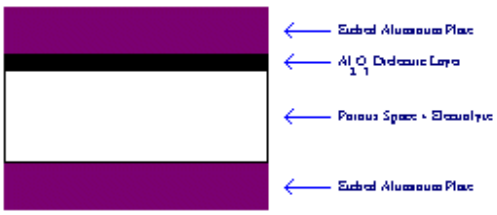
Ceramic disk capacitors are simpler to make than the monolithics. They hold a niche area as a cheaper and physically smaller (for the same capacitance value) alternative to silver mica. They are available with higher voltage ratings than monolithics. Since they only have a single slab, they are found with lower capacitance values (typical range 1.5pF to 2.2uF) than the monolithics (which can have 60-80 layers), but can still get large with a noticeable increase in physical size. At the upper end, too, the tolerance and temperature coefficients get extremely large.

Silver-mica: Another stacked, low K capacitor. Mica is really the general family name (mica is the dielectric); silvered mica is just the most popular form. Good points: low ESR, temperature coefficient between 0 and +100ppm/deg.C. These are very similar to C0G ceramics. They suffer, however, from high dielectric absorption. They are popular for their high frequency characteristics (up to 500Mhz). Typical values range from 2pF to 1500pF

Electrolytics: A "breakthrough" in capacitor technology in the early 1900's. Instead of placing a solid wedge of something (which can be quite thin), an electrolyte solution is used. The electrolyte serves as the 2nd electrode. The electrolyte is not the dielectric. The dielectric is a very thin layer of oxide which is grown electro-chemically in production. The thickness of this oxide layer is on the order of .01um, much smaller than any piece of plastic or ceramic that could be used as a separator. To contact the electrolyte, another piece of foil is used, but it is the electrolyte that is truly the plate.

There are 2 common electrolytic capacitors:

Aluminum Electrolytic: A side view of an aluminum electrolytic capacitor:

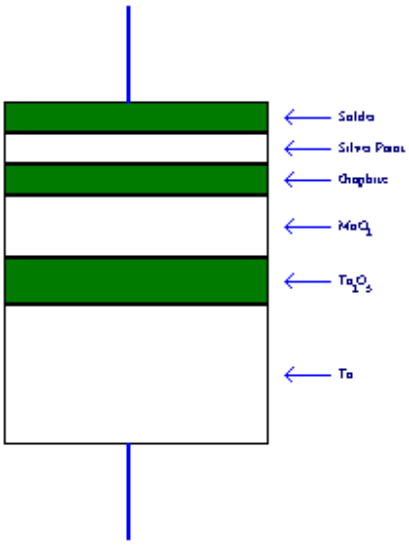


The electrolyte is held in a porous spacer. The Al plate is finely etched. Since the oxide is so thin, this etching increases the surface area, increasing the capacitance. A slab capacitor like ceramic would not benefit from this approach.

The porous spacer and both plates serve to thicken a single layer, but the important distance is the dielectric thickness, which is, of course, extremely small. Thus, electrolytics enjoy a huge capacitance density advantage over other capacitor technologies. One limitation to this technology is its polarized nature. With just a small reverse voltage, the oxide breaks down and current will freely flow. Another byproduct of this technology is the reduction in capacitance as you approach the working voltage of the capacitor. This is caused by a growth in the thickness of the oxide layer as a high voltage is placed across the capacitor. All the common aluminum electrolytics are prone to have the electrolyte dry out. This means that their lifetime is shorter than other capacitors, anywhere between 3 and 20 years depending upon the quality of the line, according to The Capacitor Handbook. If you really want your design to last, check out the aging information for your aluminum electrolytic. Typical values range from

Tantalum Electrolytic: Comes in both wet and dry electrolytes. The dry, or solid tantalum is the most common, and can't truly be called an electrolytic capacitor. Electrolytic is normally reserved for those capacitors in which one of the terminals is an electrolyte. Solid tantalums use manganese dioxide as the second terminal.

A side view of a solid tantalum capacitor:



The capacitor starts with solid Ta powder which is worked onto Ta wire. The pellet is immersed in an acid bath and attached to a DC supply. Current flow encourages the Ta2O5 oxide growth. The MnO2 layer is created by dipping the pellet into Mn(NO3)2 solution and then heated. The rest of the processing is just to get a wire connected. The solid tantalum shares the extremely thin dielectric, high capacitance behavior of the aluminum electrolytic with an added bonus that since it is dry, its lifetime in much longer, and it also has a lower leakage current (higher Rdc). Also, tantalum capacitors can tolerate a little reverse voltage (as much as 15% of the working voltage according to some sources). Typical values range from 0.047uF to 330uF. Tantalums also do not come in as high a working voltage rating (typical max voltage = 50V) as aluminum electrolytics (available with at least 450V, probably higher)..

Bipolar electrolytics are available when you need to have large capacitance but cannot maintain a unipolar bias. Electrolytics generally suffer from larger ESR and Rdc than other capacitors.

Terminology:

Capacitor Model Parameters:

ESR: Equivalent series resistance. A resistive element of the capacitor model found in both the ac and dc domains. Contributing factors: electrodes, leads, dielectric. This value can change with frequency, time, whatever.

ESL: Equivalent series inductance. An inductive element of the capacitor model; effect only seen in ac or transient domains. Contributing factors: electrodes, leads.

Rdc: a dc leakage current through the dielectric. This value varies with temperature and age.

Rac: a parameter to describe ac losses in the dielectric; may vary nonlinearly with frequency and temperature

Rd, Cd: parameters to describe dielectric absorption

dielectric absorption: phenomenon where after a capacitor has changed value, stored charge will migrate out of the dielectric over time, thus changing the voltage value of the capacitor. This is apparently an important issue in sample and hold circuit applications; perhaps someone might want to investigate this on their own.

dissipation factor (D.F.), tan(Á): This is a measure of the lossiness of the capacitor. It takes into consideration the ESR and the capacitance

through the ratio ESR/Zc = (2[[Eth]]fC)(ESR). The tangent expression comes from the phasor diagram for impedance (Á = loss angle).

K: relative permittivity, dielectric constant

Q: 1/D.F. = cot(Á)

Miscellaneous quotes from High-Speed Digital Design, A Handbook of Black Magic:

A power supply provides low impedance at low frequencies.

Local bypass capacitors provide low impedance at higher frequencies.

The best way to get very low inductance is to parallel a lot of small capacitors.

Lead inductance acts like an inductor in series with a capacitor.

ESR acts like a resistor in series with a capacitor.

Together they degrade a capacitor's effectiveness as a bypass element.

For large-valued capacitors, smaller packages have higher series inductance and ESR than larger packages.

Capacitor performance varies widely.

Higher-dielectric-constant materials pack more capacitance into a smaller space but have poor temperature coefficients and aging instability.

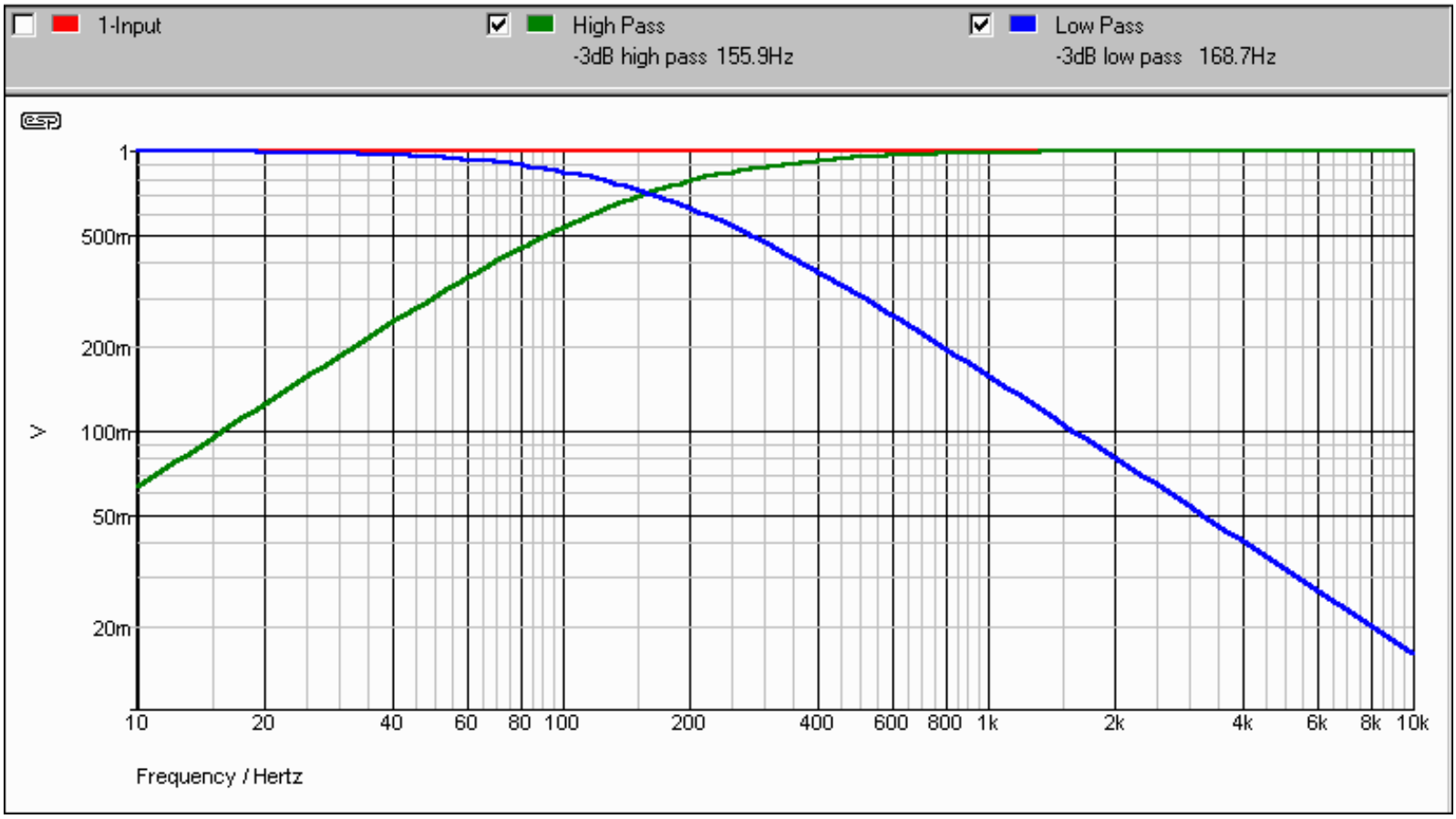
Aluminum electrolytics do not work well in cold applications.

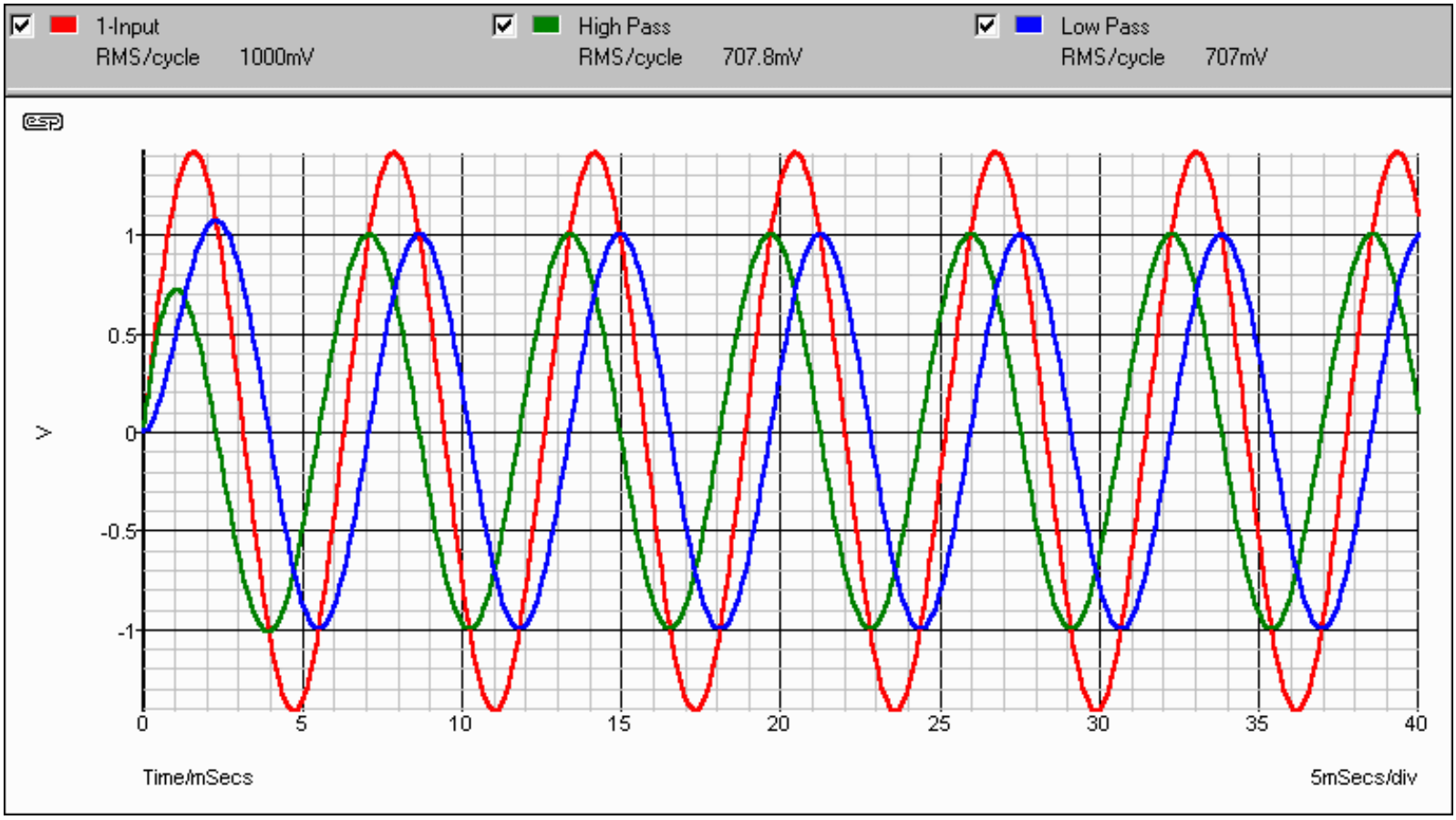
Failure in capacitors is a statistical phenomenon, accelerating at high voltages (voltages near the voltage rating).

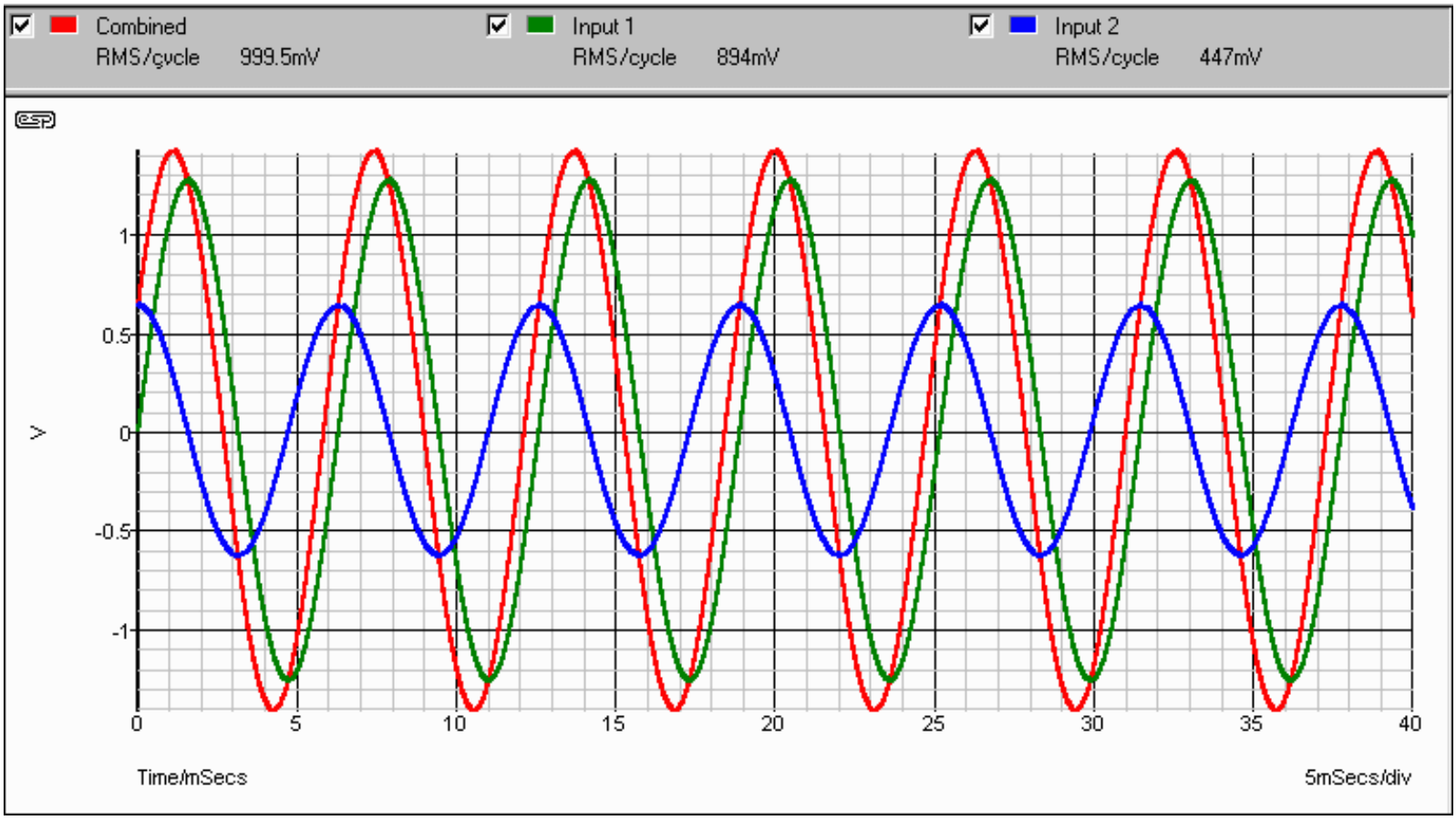
A couple of final notes:

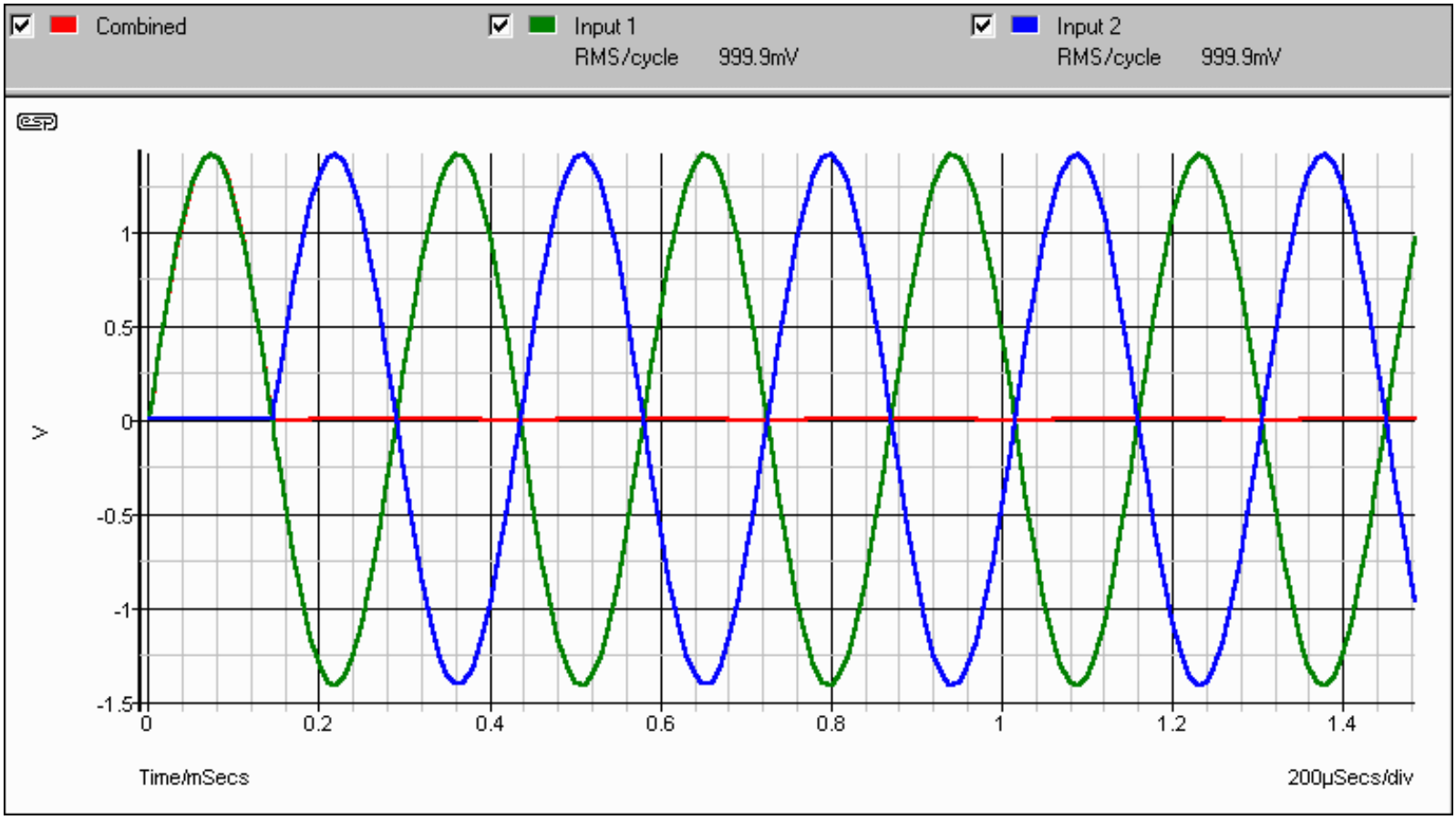
- Amongst similar sized capacitors, the physically larger one generally has features which might make it desirable (voltage rating, stability, higher tolerance, etc.). Except for electrolytics, which are way out on the dielectric constant scale, higher K values are achieved by sacrificing ideality.

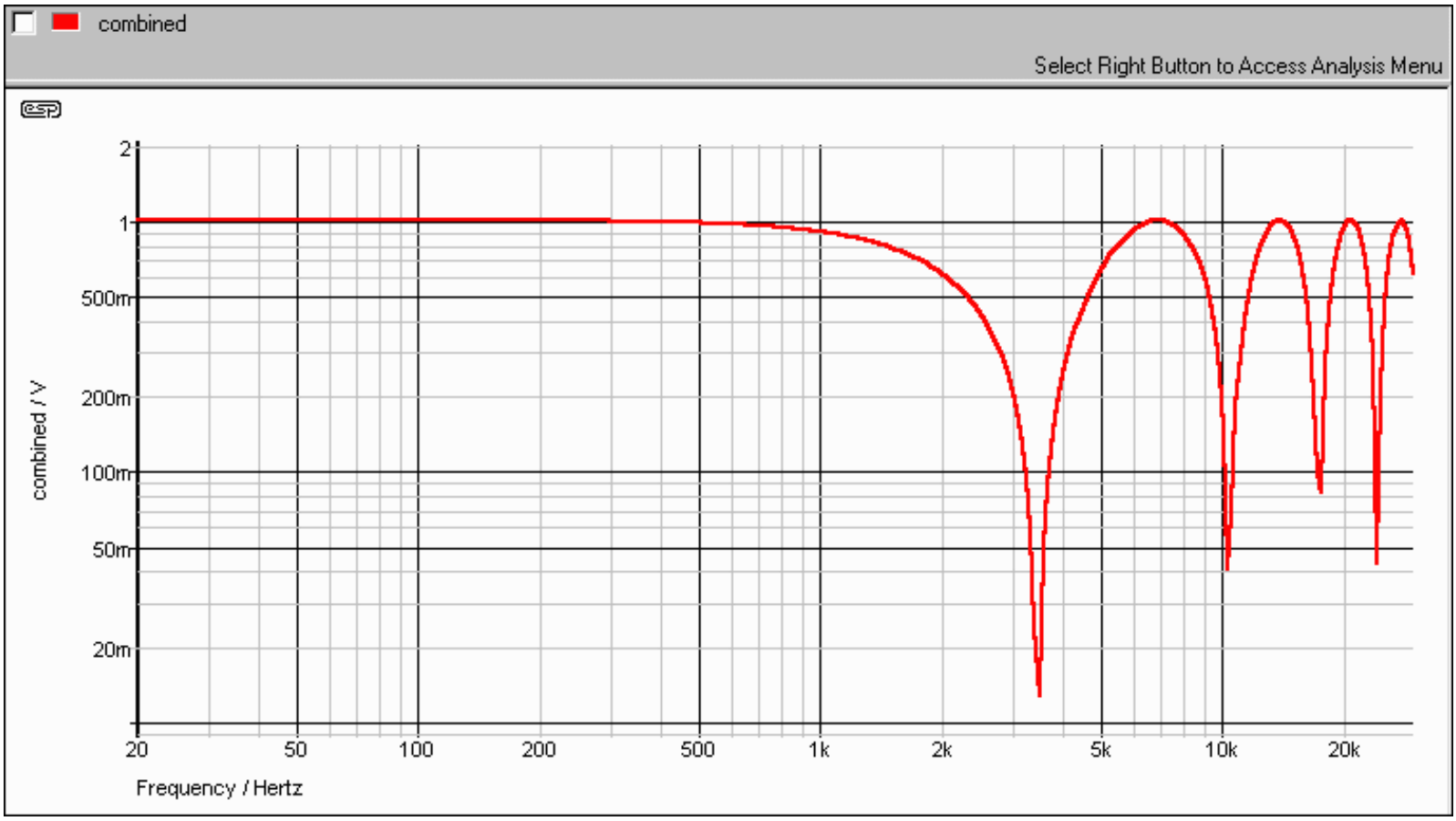
- Price is not such a poor indicator of quality; just don't pay for a quality you don't care about. Read critically!

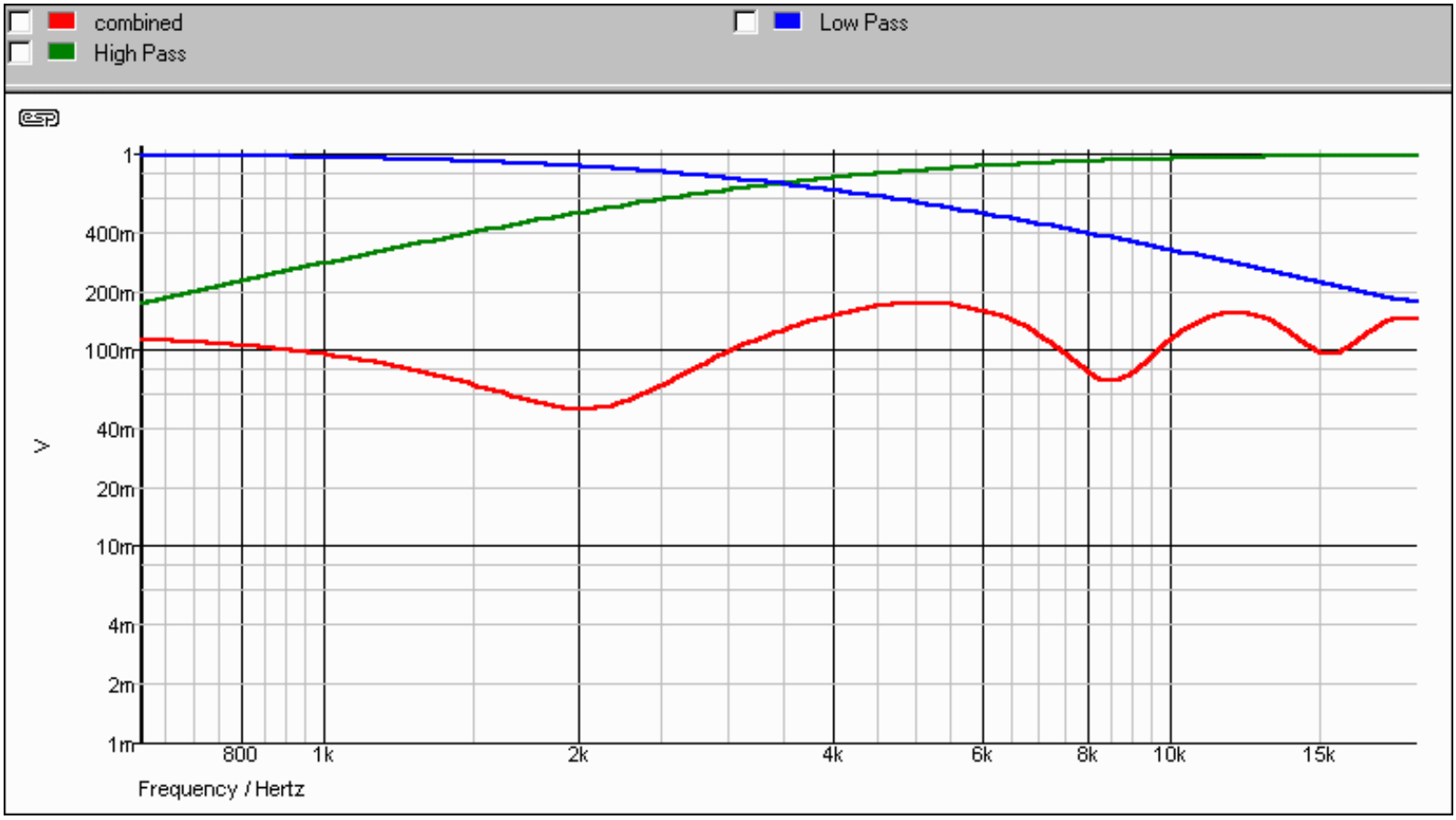

1-Input
High Pass
-3dB high pass 155.9Hz
Low Pass
-3dB low pass 168.7Hz
1
500m
200m
100m
50m
20m
V
10
20
40
60
80
100
200
400
600
800
1k
2k
4k
6k
8k
10k
Frequency / Hertz

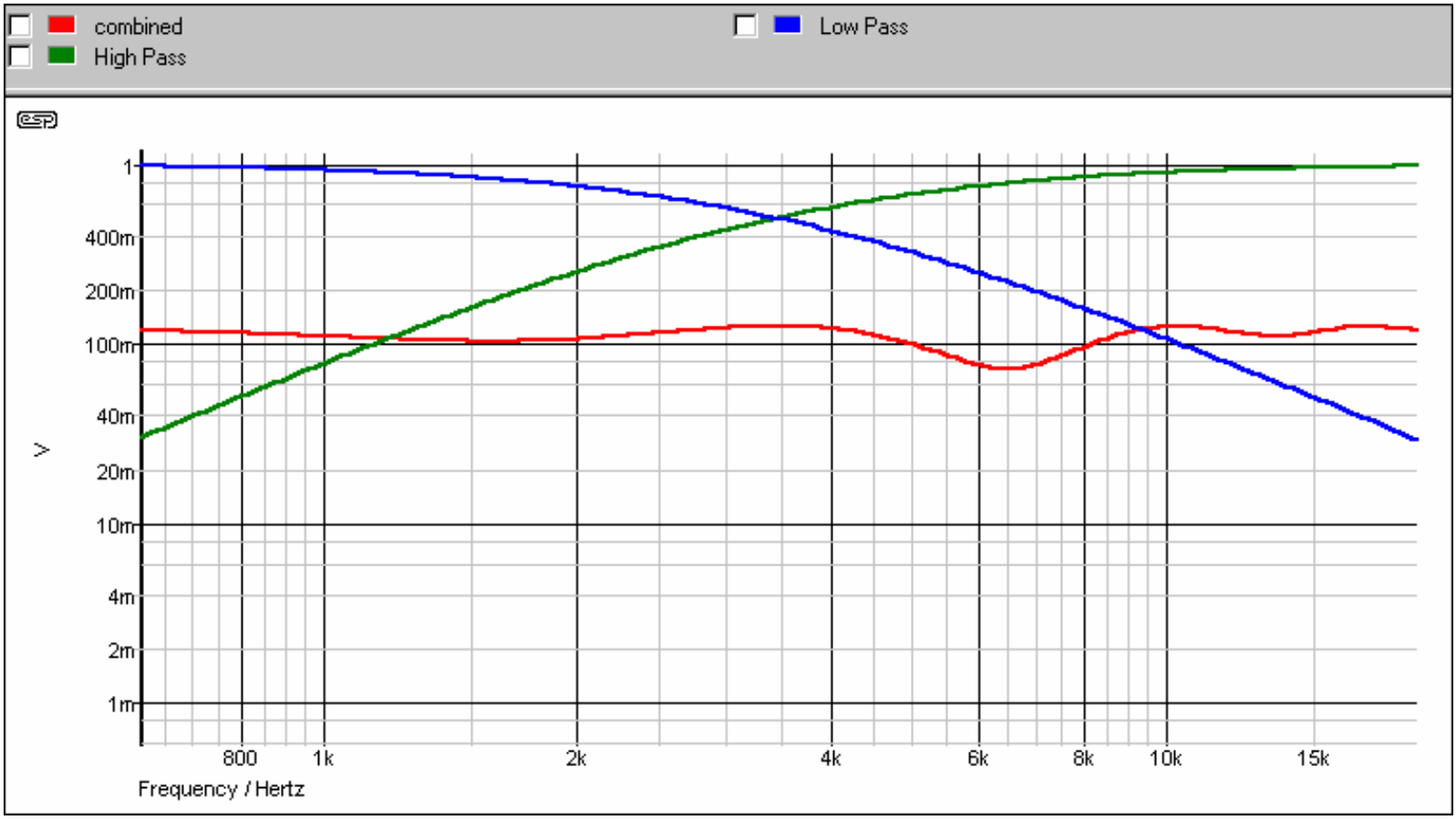

1-Input
RMS/cycle 1000mV
High Pass
RMS/cycle 707.8mV
Low Pass
RMS/cycle 707mV
1
0.5
0
-0.5
-1
V
0
5
10
15
20
25
30
35
40
Time/mSecs
5mSecs/div

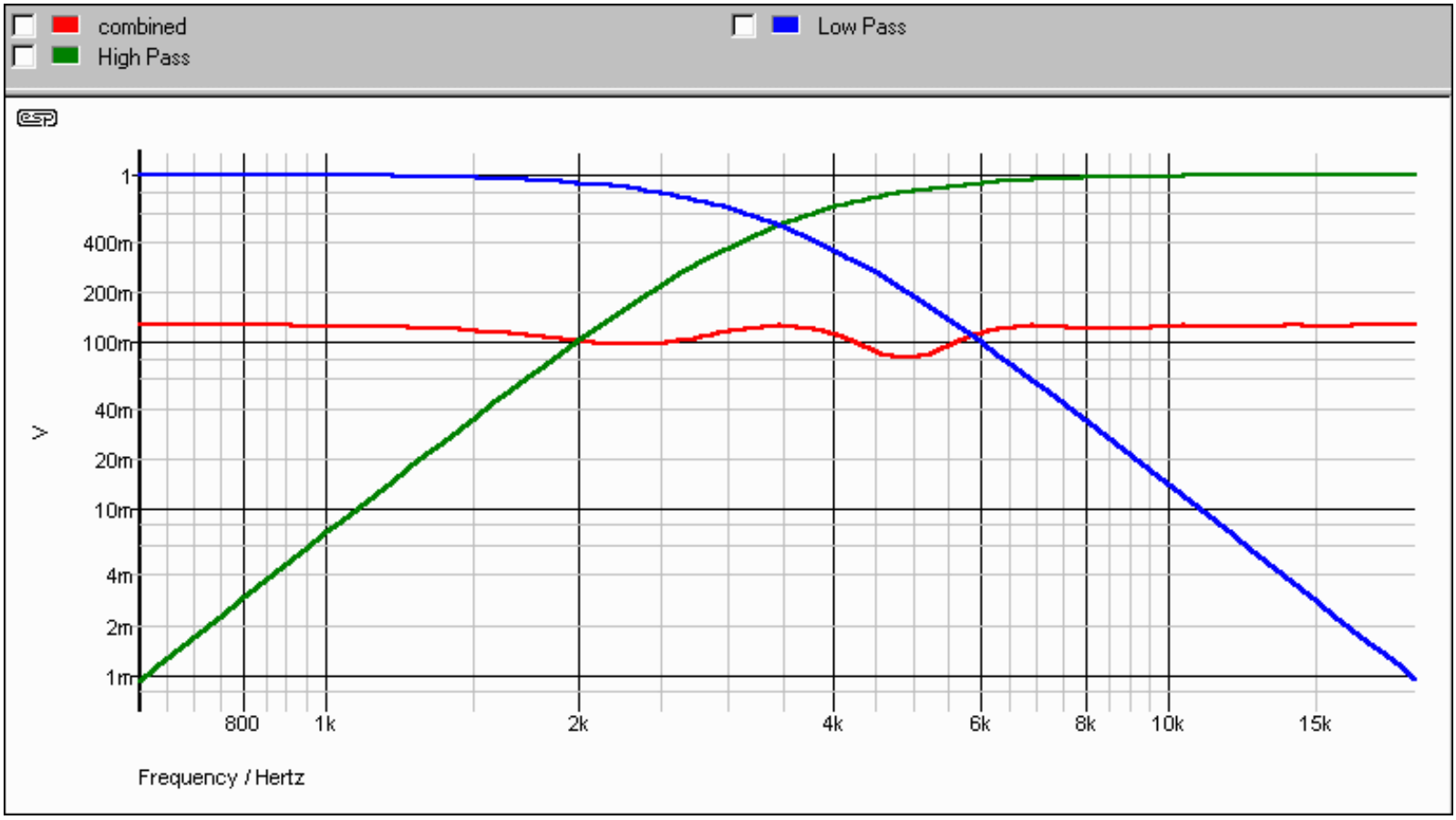

Combined
RMS/cycle 999.5mV
Input 1
RMS/cycle 894mV
Input 2
RMS/cycle 447mV
V
1
0.5
0
-0.5
-1
0
5
10
15
20
25
30
35
40
Time/mSecs
5mSecs/div

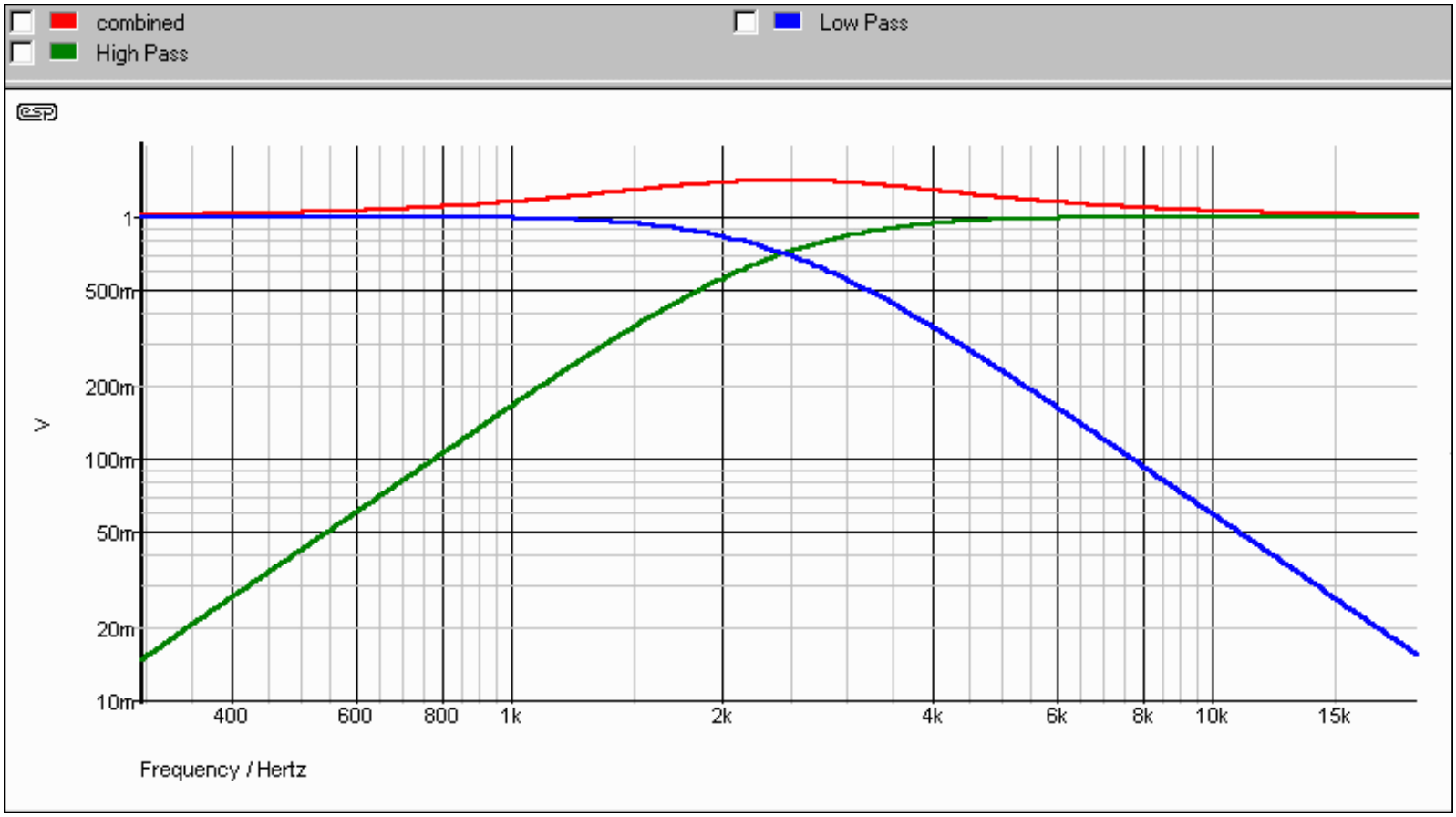

Combined
Input 1
RMS/cycle 999.9mV
Input 2
RMS/cycle 999.9mV
1
0.5
0
-0.5
-1
-1.5
V
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Time/mSecs
200µSecs/div

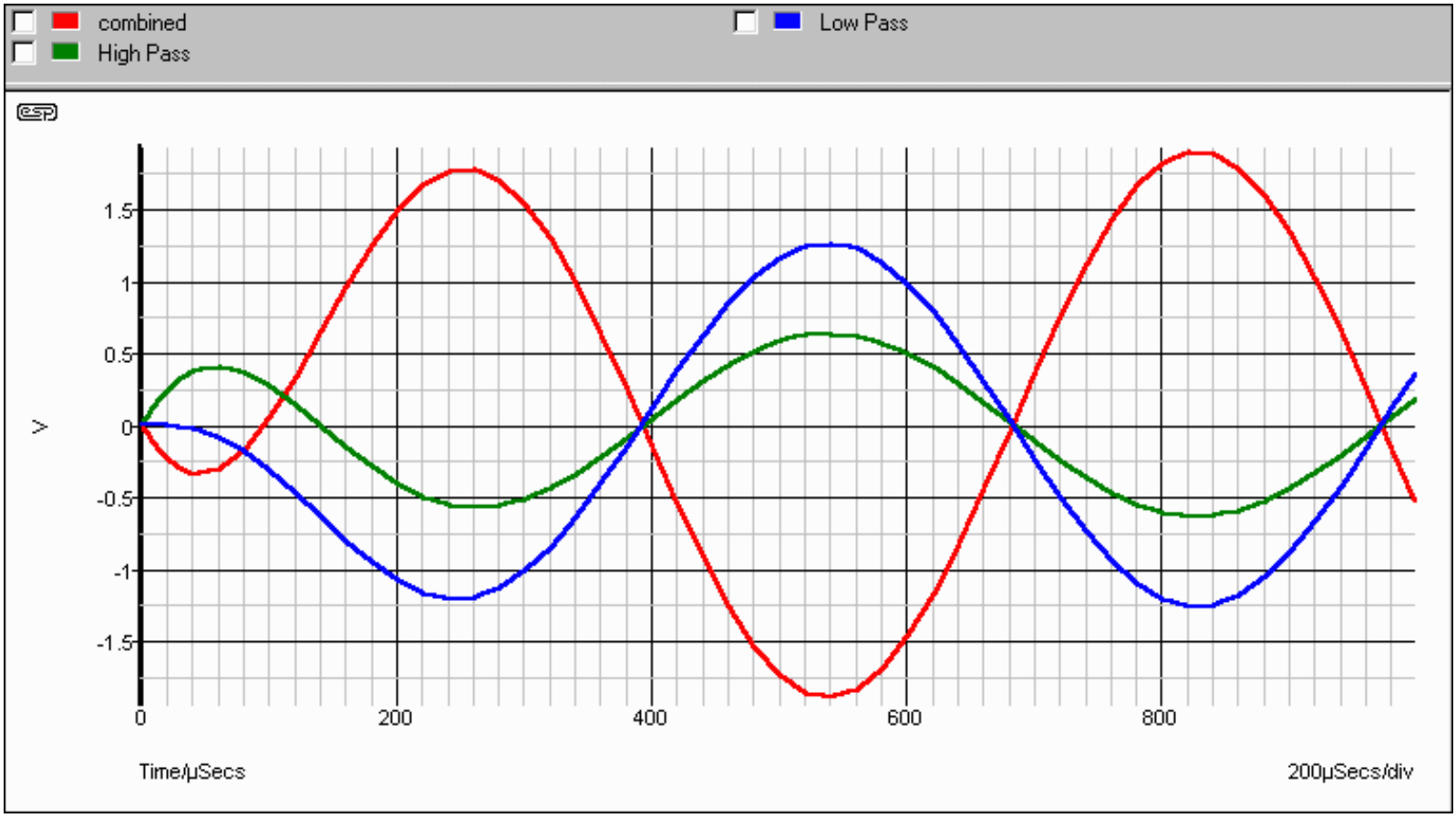

combined
Select Right Button to Access Analysis Menu
2
1
500m
200m
100m
50m
20m
combined / V
20
50
100
200
500
1k
2k
5k
10k
20k
Frequency / Hertz

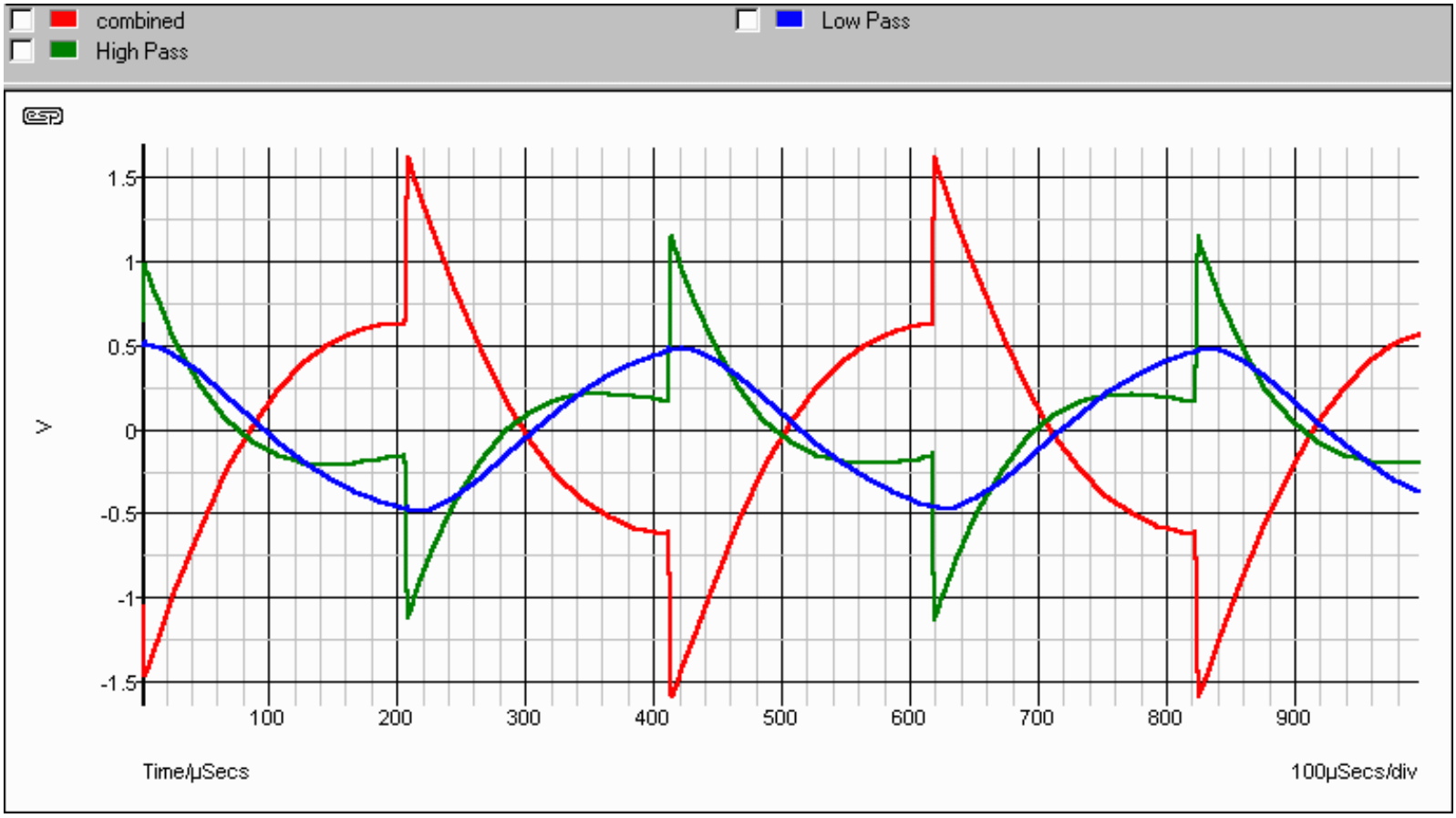

combined
High Pass
Low Pass
1
400m
200m
100m
40m
20m
10m
4m
2m
1m
V
800
1k
2k
4k
6k
8k
10k
15k
Frequency / Hertz

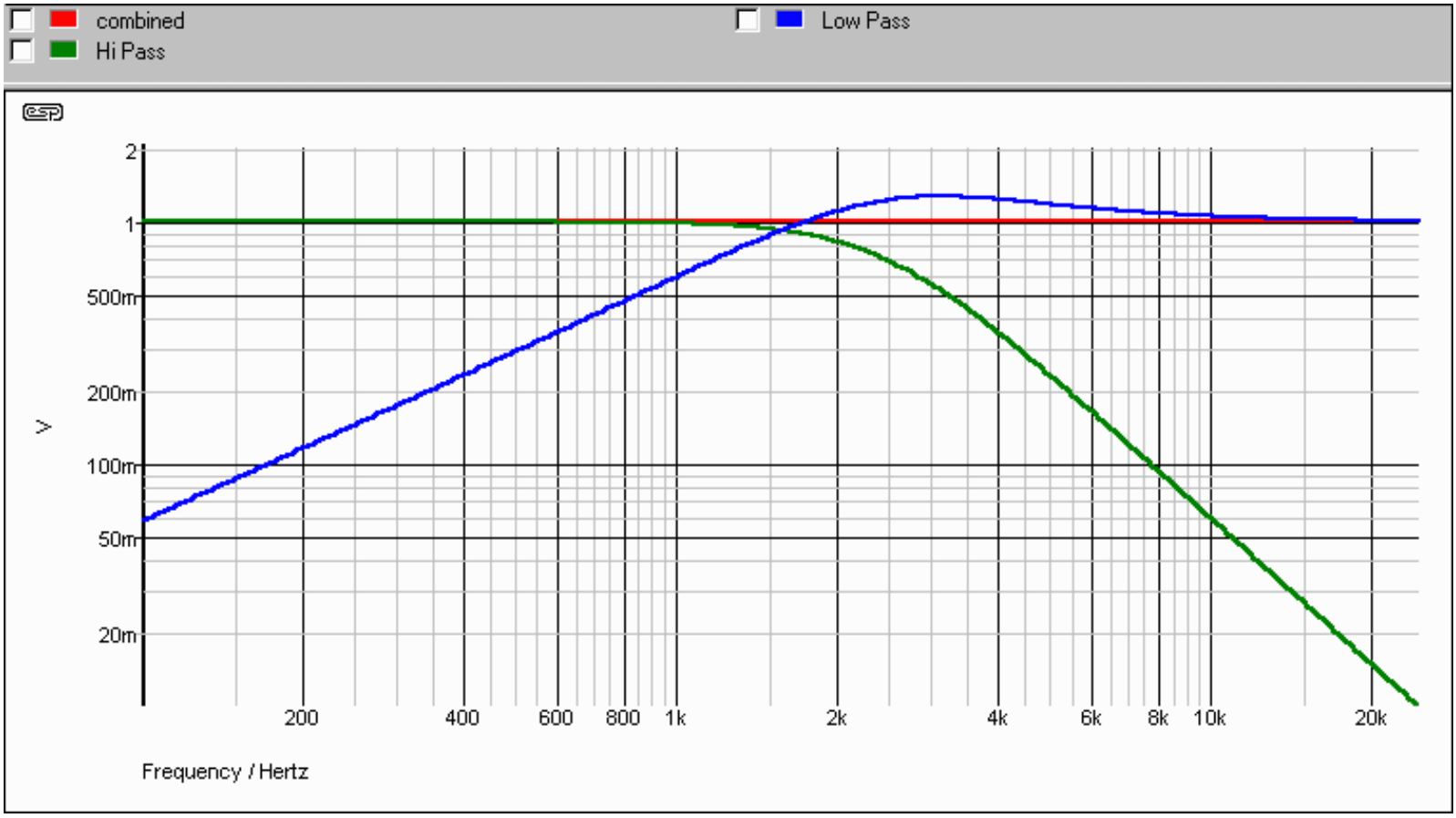

combined
High Pass
Low Pass
1
400m
200m
100m
40m
20m
10m
4m
2m
1m
V
800
1k
2k
4k
6k
8k
10k
15k
Frequency / Hertz

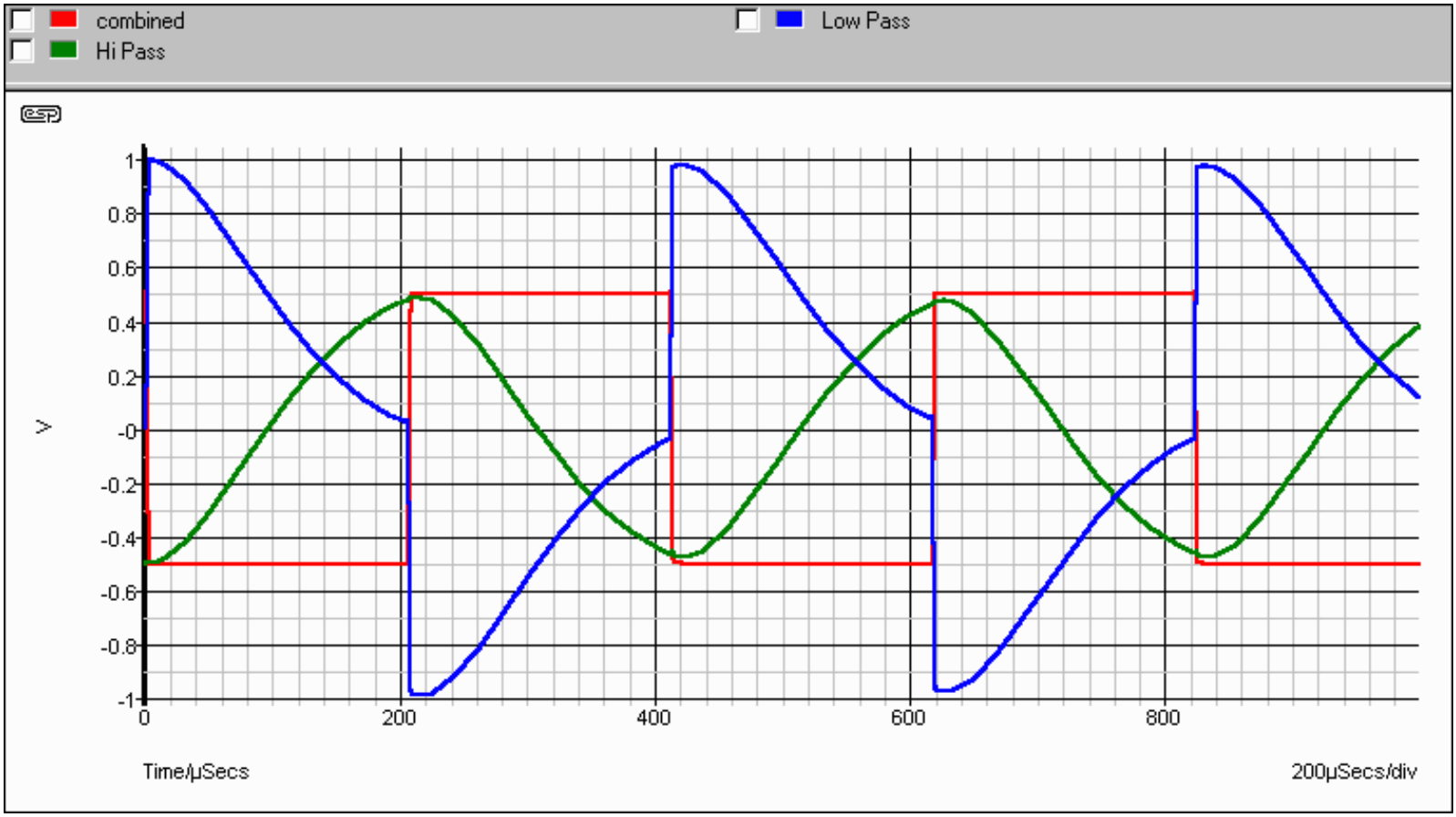

combined
High Pass
Low Pass
1
400m
200m
100m
40m
20m
10m
4m
2m
1m
V
800
1k
2k
4k
6k
8k
10k
15k
Frequency / Hertz

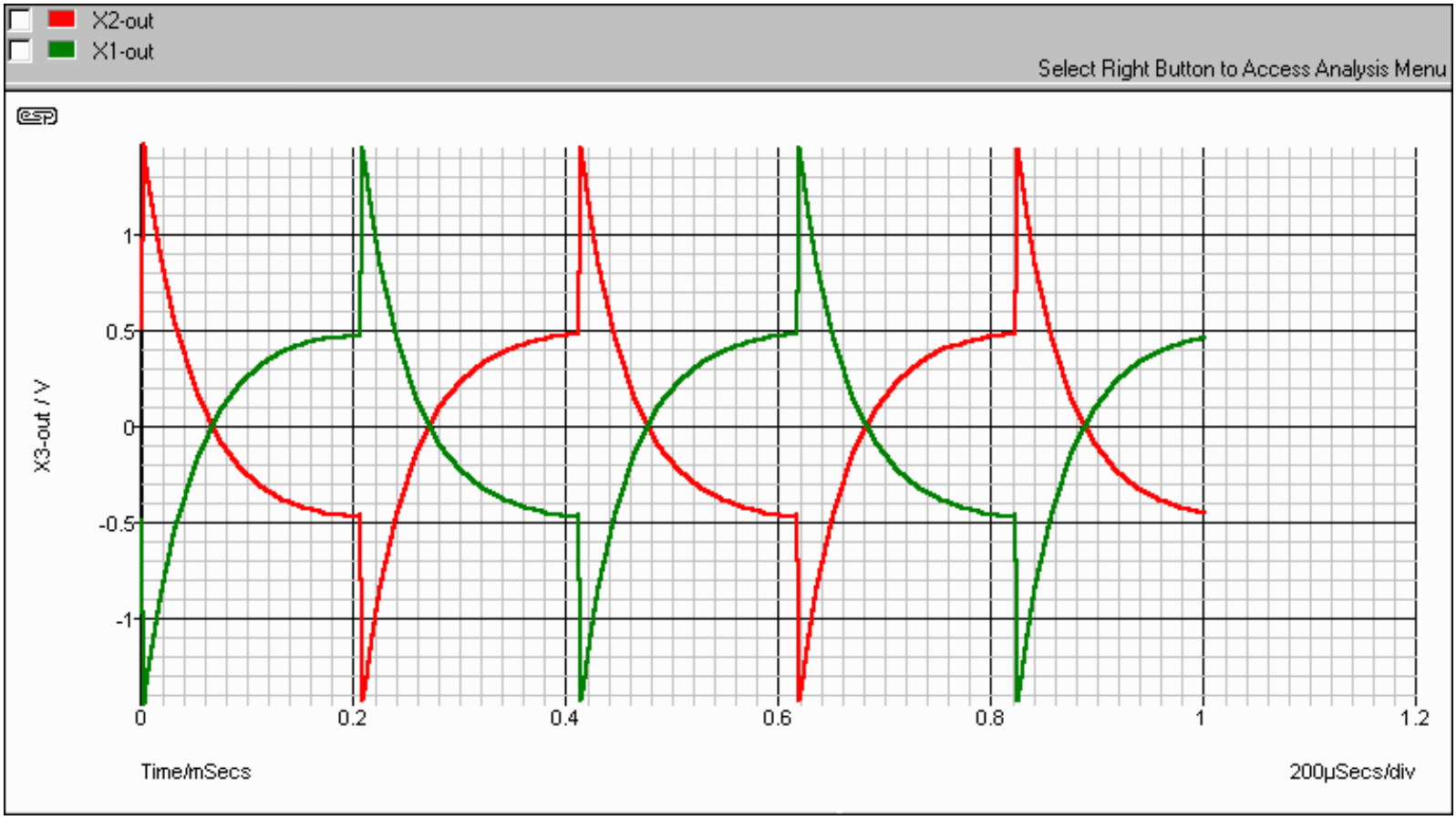

combined
High Pass
Low Pass
1
500m
200m
100m
50m
20m
10m
>
400
600
800
1k
2k
4k
6k
8k
10k
15k
Frequency / Hertz

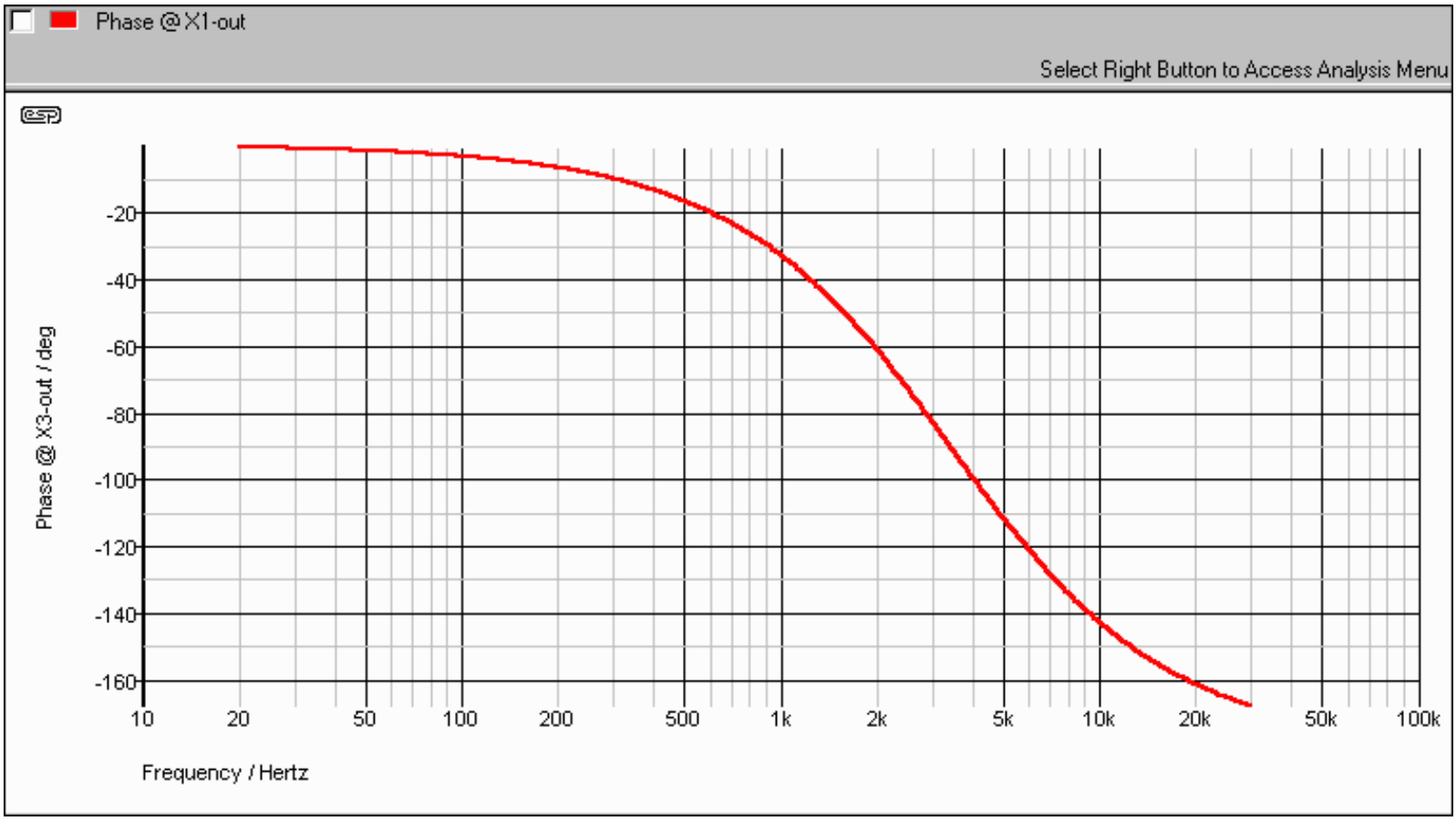

combined
High Pass
Low Pass
1.5
1
0.5
0
-0.5
-1
-1.5
V
0
200
400
600
800
Time/µSecs
200µSecs/div


combined
High Pass
Low Pass
1.5
1
0.5
0
-0.5
-1
-1.5
V
100
200
300
400
500
600
700
800
900
Time/µSecs
100µSecs/div


combined
Low Pass
Hi Pass
2
1
500m
200m
100m
50m
20m
V
200
400
600
800
1k
2k
4k
6k
8k
10k
20k
Frequency / Hertz


combined
Low Pass
Hi Pass
1
0.8
0.6
0.4
0.2
-0
-0.2
-0.4
-0.6
-0.8
-1
V
0
200
400
600
800
Time/µSecs
200µSecs/div


X2-out
X1-out
Select Right Button to Access Analysis Menu
X3-out / V
1
0.5
0
-0.5
-1
0
0.2
0.4
0.6
0.8
1
1.2
Time/mSecs
200µSecs/div


Phase @ X1-out
Select Right Button to Access Analysis Menu
-20
-40
-60
-80
-100
-120
-140
-160
Phase @ X3-out / deg
10
20
50
100
200
500
1k
2k
5k
10k
20k
50k
100k
Frequency / Hertz

ASA 130th Meeting - St. Louis, MO - 1995 Nov 27 .. Dec 01

1pEA4. Time Align(registered) loudspeaker crossovers.

Edward M. Long

E. M. Long Assoc., 4107 Oakmore Rd., Oakland, CA 94602

A Time Align(registered) loudspeaker system must produce an acoustical output at the listening position such that the fundamental and overtones of a complex, transient signal have the same relationships that they have at the electrical input to the system. To qualify for the use of the Time Align(registered) trademark, the natural time offset between the acoustical outputs of the loudspeaker drivers must be corrected. This time offset is caused by the natural low-pass filter characteristics of the drivers and the electrical crossover filters. The time offsets between the acoustical output of each of the drivers must be corrected and the electrical crossover filters must be designed to produce an acoustical output from adjacent drivers that is 6 dB down from its passband level. When the coherent acoustical outputs of adjacent drivers are combined, the result is a uniform acoustical output. The polarity of the acoustical outputs is also affected by the crossovers; this is also a major consideration in the design of a Time Align (registered) crossover. A Time Align(registered) loudspeaker system must be able to produce acoustical square waves and have a short impulse response. Time Align(registered) design techniques are shown by practical examples.

AUDIOHOLICS.com
Online A/V Magazine
Pursuing the Truth in Audio & Video...

Where audio and video equipment undergoes rigorous objective and subjective tests by our staff, ensuring that marketing slogans aren't the only guidelines for your home theater choices.

May 29 2006

Become an AH Member!

Heard something cool? **Send Us News!**

Search Audioholics

Get an A/V Education
What to Buy - Our Picks
Set-Up Your System
How to Shop
Research AV Topics
Learn About Specs & Formats
Fix Your Room Acoustics
Get the Latest Industry News
Read A/V Editorials
Read Press Releases
Try a DIY Project
Latest Trade Show News
Frequently Asked Questions
Forums: Talk With Others

Read Reviews
Receivers & Amplifiers
DVD & Media Players
HDTV, Projectors & Displays
Cables & Power Conditioners
Loudspeakers
Misc Equipment & Software
DVD Movie Reviews
CD/SACD/DVD-Audio Discs

Buy & Sell Online
Buying Guides - How to Shop
Recommended Online Stores
Find the Lowest Price Online
Sell your Gear - FREE!

About Audioholics
Our Mission Statement
Our Reference Systems
Read Staff Bios
Recommended Manufacturers
Contact the Staff
Advertise With Us
Get Cool Audioholics Stuff

Forums: Talk With Others | Professional Product Reviews | Consumer Product Reviews | Learn About Cables & Power | Set Up Your AV System | Our Picks on What to Buy | Find the Lowest Online Prices!

System Calibration & Room Acoustics

Human Hearing - Phase Distortion Audibility Part 2
Thursday, April 7 2005

Human Hearing - Phase Distortion Audibility Part 2

by: Mark Sanfilipo

Given all the foregoing research presented in various academic, scientific & industry journals, as well as other media, variously indicating the audibility of phase distortion, there remains no valid doubts regarding the existence of phase distortion.

See for yourself: pass an audio signal through, for example, a 4th order LR crossover network and view the output on an oscilloscope and the waveform will appear to be a grossly distorted copy of the original.

Only from SVS
The Sound Authority
Visit us at CES
Alexis Park #1654

Projector People
$2995 EXCLUSIVE
Mitsubishi HD4000U
CLICK!

Audioholics Supports

HOME THEATER ALLIANCE

Audioholics Editorials

Reaping what you Sow
Monday, May 29 2006

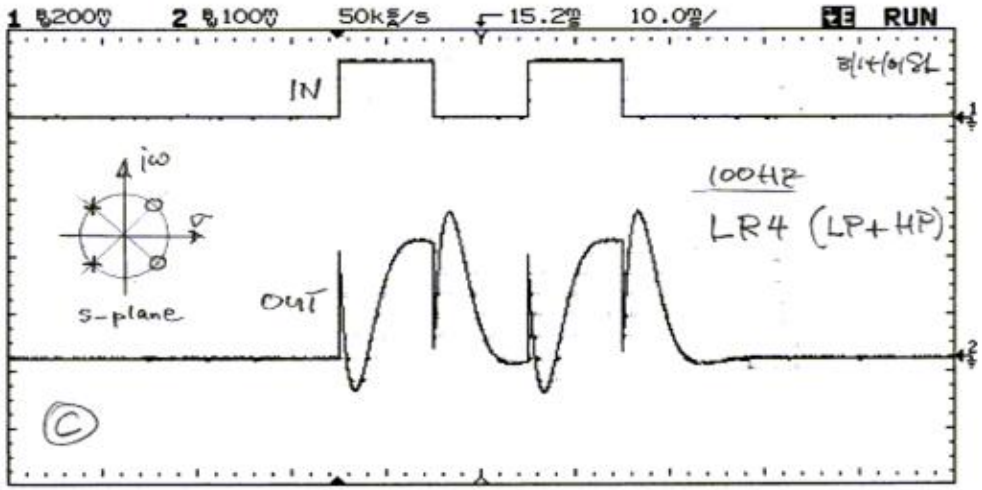
World's Smallest Projector - In Your Eye!
Wednesday, May 24 2006

Exploration of the Best Rock Saxophone Solos
Wednesday, May 24 2006

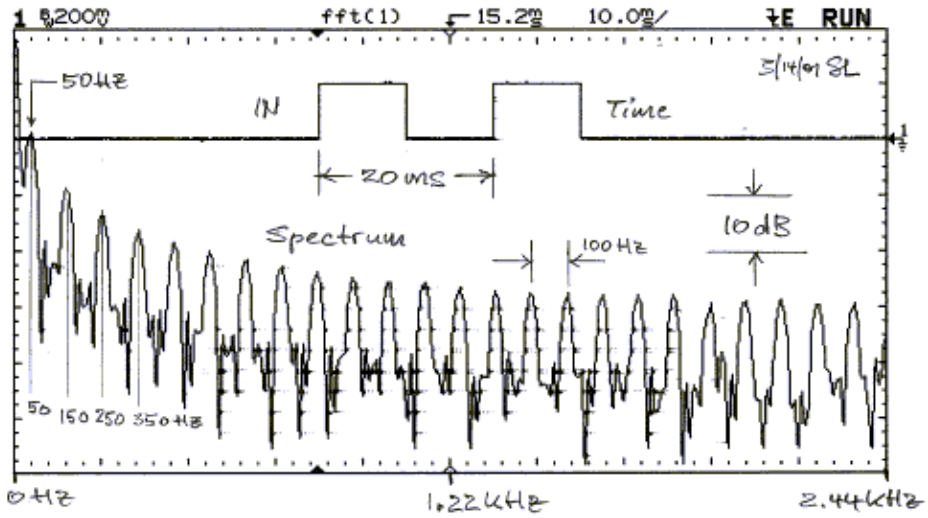
Pioneer's Inno Rocks the RIAA Boat
Wednesday, May 17 2006

Format Wars: Feel the Angst
Wednesday, May 17 2006

Flushing Out Myths In Consumer Audio
Monday, May 15 2006



Pass the same output signal through a spectrum analyzer and the amplitude of the various components will be correct, but their relative phase will have changed. Given the ready availability of software-based spectrum analyzer software and signal generators, investigation of this phenomenon is today almost a trivial exercise in signal analysis.



With these tools, it's certainly easy to visually assess phase distortion, but how easy is it to *aurally* assess this form of phase distortion? Let's take a look at a few examples of results & opinions arising from some commonly-cited research efforts.

In "On the Audibility of Midrange Phase Distortion in Audio Systems" by Lipshitz, et. al. we have:

"*1) Even quite small midrange phase nonlinearities can be audible on suitably chosen signals.*
2) Audibility is far greater on headphones than on loudspeakers.

IPTV - The Sky is Falling!
Friday, May 5 2006

Industry Press Releases

ELAN Home Systems Ships New A1240 Amplifier
Tuesday, May 16 2006

ReplayTV to Debut "PC Edition" at Digital Experience in New York
Tuesday, May 16 2006

Niles Introduces Subwoofers
Thursday, May 4 2006

VESAs DisplayPort to Compete with HDMI
Wednesday, May 3 2006

Mitsubishi XL5980U Anti-Theft Projector
Tuesday, May 2 2006

Belkin N1 802.11n Wireless Networking in June
Monday, May 1 2006

Pixel Magic Intros 1080p HD MediaBox
Saturday, April 29 2006

3) Simple acoustic signals generated anechoically display clear phase audibility on headphones.
4) On normal music or speech signals phase distortion appears not to be generally audible, although it was heard with 99% confidence on some recorded vocal material."

Lipshitz. et .al. used 1st & 2nd order unity-gain all-pass networks with f0s of ~ 100 Hz to 3kHz, switchable in increments. For the 2nd order networks Q ranged from .5 to 2 in increments. Some highlights of their experiments and results were:

	Item
1	Using 150 Hz square waves, pitch & timbre changes were notable for all Q and fo's. Though the effect was audible down to ~ 60 db SPL, it was most noticeable at higher dB SPL's.
2	Using 2 – 5 Hz repetition rate square waves audible phase effects were again noted, this time as a ringing of the networks at fo. This effect was most noticeable with high network Q settings and across the 113Hz to 529 Hz frequency range, with detection becoming more difficult above 1 kHz.
3	Using a raised-cosine tone burst, audible signal changes in the 160 Hz to the 353 Hz range fed through a 2nd order all-pass network, with Q's ≥1. When the same tone bursts were fed through Low-Q, 1st order networks, the audible signal changes were that occurred were of a much lower magnitude.
4	Using pre-recorded music (male & female singing) fed through a 2nd-order all-pass network, with a Q of .5, audible effects were noted with a 95% confidence level. Using a variety of unpitched sounds recorded anechoically, phase effects were again audible.

And from Dr Floyd Toole:

"*It turns out that, within very generous tolerances, humans are insensitive to phase shifts. Under carefully contrived circumstances, special signals auditioned in anechoic conditions, or through headphones, people have heard slight differences. However, even these limited results have failed to provide clear evidence of a 'preference' for a lack of phase shift. When auditioned in real rooms, these differences disappear...*"

Essentially, Dr Toole expressing the opinion he does here (based on decades of his own, often groundbreaking research), captures the essence of the conclusions reached by Lipshitz et. al.

In Daisuke Koya's "Aural Phase Distortion Detection":
"*Although not in large numbers, previous research in investigation of the audibility of phase distortion has proven that it is an audible phenomenon. Lipshitz et al. has shown that on suitably chosen signals, even small midrange phase distortion can be clearly audible. Mathes and Miller and Craig and Jeffres showed that a simple two-component tone, consisting of a fundamental and second harmonic, changed in timbre as the phase of the second harmonic was varied relative to the fundamental. The above experiment was replicated by Lipshitz et al., with summed 200 and 400 Hz frequencies, presented double blind via loudspeakers resulting in a 100% accuracy score.*

An experiment involving polarity inversion of both loudspeaker channels resulted in an audibility confidence rating in excess of 99% with the two-component tone, although the effect was very subtle on music and speech. Cabot et al. tested the audibility of phase shifts in two component octave complexes with fundamental and third-harmonic signals via headphones. The experiment demonstrated that phase shifts of harmonic complexes were detectable.

Another very simple experiment conducted by Lipshitz et al. was to demonstrate that the inner ear responds

asymmetrically. Reversing the polarity of only one channel of a pair of headphones markedly produces an audible and oppressive effect on both monaural and stereophonic material. This effect predominantly affects frequency components below 1 kHz.

Because reversal of polarity does not introduce dispersive or time-delay effects into the signal, but merely reverses compressions into rarefactions and vice versa, these audible effects are due only to the constant 180° phase shift that polarity reversal brings about.

Since interaural cross-correlations do not occur before the olivary complexes to which the acoustic nerve bundles connect, it must be concluded that what is changed is the acoustic nerve output from the cochlea due to polarity reversal. This change owes to two factors: cochlear response to the opposite polarity half of the waveform, and the waveform having a shifted time relationship relative to the signal heard by the other ear. This reaffirms the half-wave rectifying nature of the inner ear."

A frequent argument to justify why phase distortion is insignificant for material recorded and/or reproduced in a reverberant environment is that reflections cause gross, position-sensitive phase distortion themselves. Although this is true, it is also true that the first-arrival direct sound is not subject to these distortions, and directional and other analyses are determined during the first few milliseconds after its arrival, before the pre-dominant reverberation's arrival. Lipshitz et al. do not believe that the reverberation effects render phase linearity irrelevant, and there exists confirmatory evidence."

Let's take a closer look at Daisuke's test and results.

The test signals employed were:

1. 70 Hz sawtooth wave
2. 3.5 kHz sawtooth wave
3. 10 kHz sawtooth wave
4. Impulse
5. Jazz-vocal group
6. Percussion instruments.

A Kwalwasser - Dykema based test model was employed by the researchers. In this particular implementation, the test subjects were presented with two signals, one all - pass filtered, the other not. The test subjects (15 in all) then had to indicate whether or not a difference was heard. The 6 – signal test run was done using first headphones, then loudspeakers.

First, the headphone-series results:

Test Signal		**Average Correct Responses** (N = 15)
70 Hz Sawtooth Wave	$t$max = 4 msec	9

70 Hz Sawtooth Wave	*t*max = 8 msec	9.5
3.5 kHz Sawtooth Wave	*t*max = 4 msec	5.5
3.5 kHz Sawtooth Wave	*t*max = 8 msec	4
10 kHz Sawtooth Wave	*t*max = 4 msec	1.5
10 kHz Sawtooth Wave	*t*max = 8 msec	4.5
Impulse	*t*max = 4 msec	13
Impulse	*t*max = 8 msec	14
Jazz-Vocal	*t*max = 4 msec	6.5
Jazz-Vocal	*t*max = 8 msec	4
Percussion Instruments	*t*max = 4 msec	4
Percussion Instruments	*t*max = 8 msec	4

Here, τmax is peak group delay for the particular filter used, msec

And the results for the loudspeaker run:

Test Signal		**Average Correct Responses** (N = 15)
70 Hz Sawtooth Wave	*t*max = 4 msec	7
70 Hz Sawtooth Wave	*t*max = 8 msec	7

3.5 kHz Sawtooth Wave	*t*max = 4 msec	5
3.5 kHz Sawtooth Wave	*t*max = 8 msec	3.5
10 kHz Sawtooth Wave	*t*max = 4 msec	2
10 kHz Sawtooth Wave	*t*max = 8 msec	2
Impulse	*t*max = 4 msec	8.5
Impulse	*t*max = 8 msec	6.5
Jazz-Vocal	*t*max = 4 msec	4.5
Jazz-Vocal	*t*max = 8 msec	3.5
Percussion Instruments	*t*max = 4 msec	4
Percussion Instruments	*t*max = 8 msec	2.5

Comparing the results encapsulated in both tables it would appear we can reasonably draw the following conclusions:

1. The number of correct responses given for the loudspeaker-based test is lower for just about every test signal used when compared to those reported in the headphone run.

2. The results from the loudspeaker run appear more independent of the level of phase distortion present than that presented by the headphone run.

3. The audibility of phase distortion in steady state signals was frequency dependent for both the loudspeaker and headphone runs.

Research work done in a controlled, scientific manner is a good first step in sorting fact from fiction – this holds true whether the study's subject matter is subjective or objective in nature.

An essential second step is the correct application of statistical analysis to the results. Oftentimes, the raw data that are the study's results present an impression or picture that can easily lead to false or otherwise incomplete interpretation. From that, incorrect conclusions can easily be drawn. Quality research simply requires the application of statistical analysis in order for the results to be correctly interpreted.

Now have a look at the next table, which outlines the correct number of responses, r, needed to conclude that the performance in each trial was better than that obtained by simple chance.

N	***r***	**Type 1 Error** (*a*) actual value	**Type 2 Error** (*b*)			
			p = 0.6	*p* = 0.7	*p* = 0.75	*p* = 0.8
15	14	0.0005	0.9948	0.9647	0.9198	0.8329
	13	0.0037	0.9729	0.8732	0.7639	0.6020
	12	0.0176	0.9095	0.7031	0.5387	0.3518

11	0.0592	0.7827	0.4845	0.3135	0.1642
10	0.1509	0.5968	0.2784	0.1484	0.0611
9	0.3036	0.3902	0.1311	0.0566	0.0181
8	0.5000	0.2131	0.0500	0.0173	0.0042
7	0.6964	0.0950	0.0152	0.0042	0.0008
6	0.8491	0.0338	0.0037	0.0008	0.0001

For this experiment in the audibility of phase distortion, the researchers chose $r \geq 9$ as that point where r represents a statistically significant result.

Applying the $r \geq 9$ requirement to classify a result as statistically significant, we can now view the data presented in the two tables in a very different light.

When applying the $r \geq 9$ requirement, none of the 12 trials run in the loudspeaker series indicated a statistically significant result, though the Impulse test signal @ 4 m Sec comes very close at r = 8.5 . Only 4 of the 12 headphone trials indicated a statistically provable presence of the audibility of phase distortion.

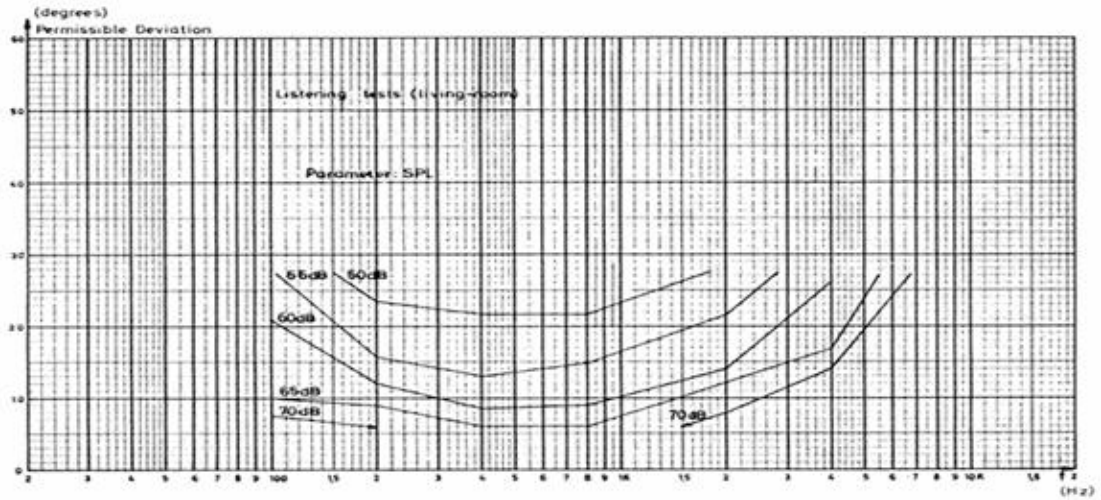
So what useful information can we draw from this particular study?

1. Phase distortion is audible, but only under very specific circumstances, using very specific, types of test signals.

2. There exists in this study no statistically significant evidence supporting the audibility of phase distortion in the musical samples provided, using the all-pass filter implementations chosen by the researchers.

3. Introducing the room acoustic variable in to the equation further lowers the already poor scores phase distortion audibility scores..

One other study I'd like to cite was that done by V. Hansen and E. R. Madsen. Originally published in the Journal of the Audio Engineering Society, in a 2-part series of articles.



They presented their test subjects a series of test tones, from the resulting data they were then able to determine permissible phase deviation levels for a variety of dB SPL levels. Of particular interest is how the presence of the positive slopes at the upper frequency portions of the curves support item 3 mentioned above in the first set of Daisuke study conclusions. These are given in Graphic 3.

So what conclusions regarding the audibility of phase distortion can we draw from the all of the above?

1. Given the data provided by the above cited references we can conclude that phase distortion is indeed audible, though generally speaking, only very subtly so and only under certain specific test conditions and perception circumstances.

2. The degree of subtly depends upon the nature of the test signal, the dB SPL level at which the signal is perceived, the acoustic environment in which the signal was recorded and/or played back as well as the Q & fo of any filter networks in the signal stream. Certain combinations of conditions can render it utterly inaudible.

3. Room acoustics further masks whatever cues that the hearing process may depend upon to detect the presence of phase distortion.

Read Part 1: Human Audibility to Amplitude Variation

Feedback from Dr. Floyd Toole (Added 4/08/05)

Thanks for bringing these articles to my attention. I guess the only comments I might have added relate to the differences in the perceptual tasks in the various studies leading to minimum audible amplitude shifts. Some were basically broadband or single tone, turn the volume up, kinds of tests. These 'difference limen, or just-noticeable difference' for loudness tests are fundamentally different from the ones that Sean Olive and I did, which was a just-noticeble timbre difference test - only a portion of the total sound spectrum was altered. Somewhere in the recesses of auditory signal processing they are probably related, but they are very different at a higher perceptual level. In any event, it is interesting to see anyone discussing the topic.

As was mentioned, our sensitivity to the timbral changes was very much dependent on the Q, or bandwidth of the

phenomenon - with much lower 'thresholds' being found for wider bandwidth spectral changes. Interesting, because the narrower the bandwidth, the more the resonant system rings. Except, perhaps, at very low frequencies, it seems that we do not exhibit a primary response to the time-domain (ringing) problems but instead to the spectral features. Our sensitivity to resonances was also dependent on program material - pink noise was the most revealing, big band/symphonic with reverb was next, and the least revealing was close-miked pan-potted pop and jazz. Reverberation alone could increase our sensitivity to a medium or low-Q resonance by about 10 dB - a huge effect. This latter fact explains why music is so much more satisfying in a reverberant space than outdoors - timbrally richer because we can hear more of the resonant subtleties. It also explains why the toughest test for loudspeaker accuracy is in a room with some reflections, and why headphones (which have no added reflections) have an inherent advantage, and can sound acceptable when measurements indicate that there really are resonant problems. Killing all early reflections with absorbers not only changes imaging, it also makes loudspeakers with poorly controlled directivity sound better. All interesting stuff for audio folk.

The piece on phase distortion was a refreshing change from the semi-science floating around our business. All things being equal, and if one has the option, of course get the phase correct - at least at the one point in space where it can be done!! However, this presents problems for two-eared listeners in multiple seats in reflective rooms (solve this one and a Nobel prize awaits). It is indeed fortunate that humans are so unresponsive to this effect because, if we could hear phase shift, we would go absolutely nuts in everyday life. Every time a reflected version of a sound adds to the direct sound, the phase shifts are enormous, and it happens in abundance in all rooms, even carrying on a conversation across a table. Do the stand up/ sit down test while speaking. The voice changes very subtly, but our hearing system compensates immediately and, on a scale of 10, the voice quality remains a 10. Yet the transfer function between the voice and the ears has greatly changed in both amplitude and phase. I cannot help but think of all the opera recordings and film voice overs that are done with librettos and scripts on large angled (sound reflecting) surfaces between the mouth and the mic. The signal is corrupted at the source! Thank your favorite diety for human adaptability.

Keep up the good work.

Floyd

Floyd E. Toole, PhD
Vice President Acoustical Engineering
Harman International Industries, Inc.

Bibliography

Koya, Daisuke: “Aural Phase Distortion Detection”, Masters dissertation, Master of Science in Music Engineering Technology, University of Miami, Coral Gables, Fla., May 2000

Hansen, V and Madsen, E. R.: “On Aural Phase Detection” (J. Audio Eng. Soc., vol. 22, pp. 10-14 (1974 Jan./Feb.))

Hansen, V and Madsen, E. R.: "On Aural Phase Detection: Part II," (J. Audio Eng. Soc., vol. 22, pp. 783-788 (1974 Dec.))

Lipshitz, S. P., Pocock, M., and Vanderkooy, J., "On the Audibility of Midrange Phase Distortion in Audio Systems," J. Audio Eng. Soc., vol. 30, pp. 580-595 (1982 Sep.).

Green, Steve: "A New Perspective on Decimation and Interpolation Filters,
Cirrus Logic, Inc., July 2004

Walker, R.: "Acoustic Modeling – approximations to the real world",
BBC R&D White Paper WHP 005, London, Eng., September 2001

Gerzon, Michael: "Why do equalizers sound different?", Studio Sound, July 1990

Cabot, Richard C., "Audible Effects vs. Objective Measurements In The
Electrical Signal Path", Audio Engineering Society,
8th International Conference

Linkwitz, S.: "Group delay and transient response"

Previous Page

Read more **Room Acoustics Articles**

Join the Home Theater Alliance Newsletter! Sign Up!

designed & hosted by CD MEDIA INCORPORATED

Copyright ©2005 Audioholics.com. All Rights Reserved. Reproduction rights, terms & conditions. Privacy Policy.

The Audio Pages

Elliott Sound Products **Project 73**

Hi-Fi PC Speaker System

Rod Elliott (ESP)

Updated 31 Aug 2004

All required PCBs are available for this project. Click the image for the price list.

Introduction

If you like listening to music while using your PC, or enjoy games with all their sound effects, you will generally be disappointed with most of the PC speaker systems available. Few systems use anything other than moulded plastic boxes, ratty little tweeters that don't really do very much at all, and commonly using only a tiny bipolar electro as the "crossover". Zero acoustic damping and a 5W amplifier completes the picture.

As for the sub-woofer, if it manages to wheeze its way down to 50Hz you are indeed fortunate, and with as much as 20W in a tiny ported plastic enclosure you can think yourself lucky to get that far. I know this from experience, and eventually got so sick of the grossly inferior sound that I just had to do something about it.

This is not so much a project in its own right as a collection of existing projects, and using some of the boards that I sell (I bet that came as a real surprise!). With a bit of work and not too many dollars, pounds, shekels or whatever, you can make a PC speaker system that will outperform many small hi-fi setups.

Description

There are two parts to the project, small satellite speakers that sit on the desk next to the monitor, and a sealed, equalised sub-woofer that also contains all the amplifiers, power supplies and other electronics. The sub extends happily to 30Hz, and makes quite a good account of itself considering its size. It also makes a fine foot rest :-)

Is this an audiophile project - no. It is head and shoulders above any PC speaker system I have ever heard, regardless of price, and also makes a great little hi-fi. The sound is very clean, and it has extended lows that put a lot of "proper" systems to shame.

The selection of speaker drivers is up to you, since I have no idea what you will be able to get from your local suppliers. Don't be too cheap - the better the drivers, the better the overall sound, and the happier you will be with the end result. At the other end of the scale, using the best drivers you can find is probably not worth the effort either, since there is already a limitation in sound quality imposed by your sound card.

Please note that all dimensions shown are in millimetres. For those who insist on those inch thingies, divide by 25.4. Litres can be converted to cubic feet by dividing by 28.

The parameters of the speaker drivers I used are shown below. These cost me a little under AU$150 from local Australian suppliers. The only speaker likely to be recognised is the Peerless tweeter.

Woofer	Re/sponse 200mm	**Midrange**	Unknown 100mm	**Tweeter**	Peerless 812978
Impedance	8 Ohms	Impedance	8 Ohms	Impedance	8 Ohms
Power	120 W	Power	30 W	Power	100 W **
Sensitivity	90 dB/W/m	Sensitivity	91 dB/W/m	Sensitivity	91 dB/W/m
Resonance	40 Hz	Resonance	91 Hz	Resonance	907 Hz
Total Q (Qts)	0.47	Total Q (Qts)	0.72	Total Q (Qts)	0.8944
Vas (litres)	32.8	Vas (litres)	2.11	Vas (litres)	n/a
Cone Area (m²)	0.0214	Cone Area (m²)	n/a	Cone Area (m²)	n/a

Parameters For Prototype System Drivers

** This is the nominal system power - the tweeter would self destruct with this power within its normal operating range
n/a - Not Available / Not Applicable

The above information is mainly for the sake of interest. The drivers you use will have to be those you can obtain, but the above figures give you a goal to aim for. If you can get drivers with similar specs, then you will have a system as good (or better) than mine.

Satellite Speakers

Firstly, you need to obtain some magnetically screened drivers. Ordinary drivers must not be used, as they will distort the colour and even the picture shape on your monitor. The ones I used are 100mm diameter (outside) and are rated at 30W RMS or 50W peak. They are both 91dB/W/m and are 8 ohms. The crossover is a simple passive 6dB network at 4kHz. This is not ideal, but is a reasonable compromise between complexity and performance. With an internal volume of about 3 litres, the boxes are sealed, well damped with fibreglass, and have a resonance of 110Hz.

It is important that the low-midrange driver and tweeter have the same efficiency. If not, make sure that the tweeter has the higher efficiency so it can be padded back to match the midrange. If not, you will have to pad the midrange driver, and this will soak up much more power (as heat), and reduces damping - neither is acceptable. A simple resistive pad can be used to reduce the tweeter level, or you can use a

"proper" wirewound attenuator pot so it can be adjusted.

The boxes are covered with carpet (the thin stuff for covering speaker boxes), which helps to eliminate diffraction, and adds extra damping to the boxes. The boxes themselves are made with 10mm medium density fibreboard (MDF) and because of their small size, are extremely rigid.

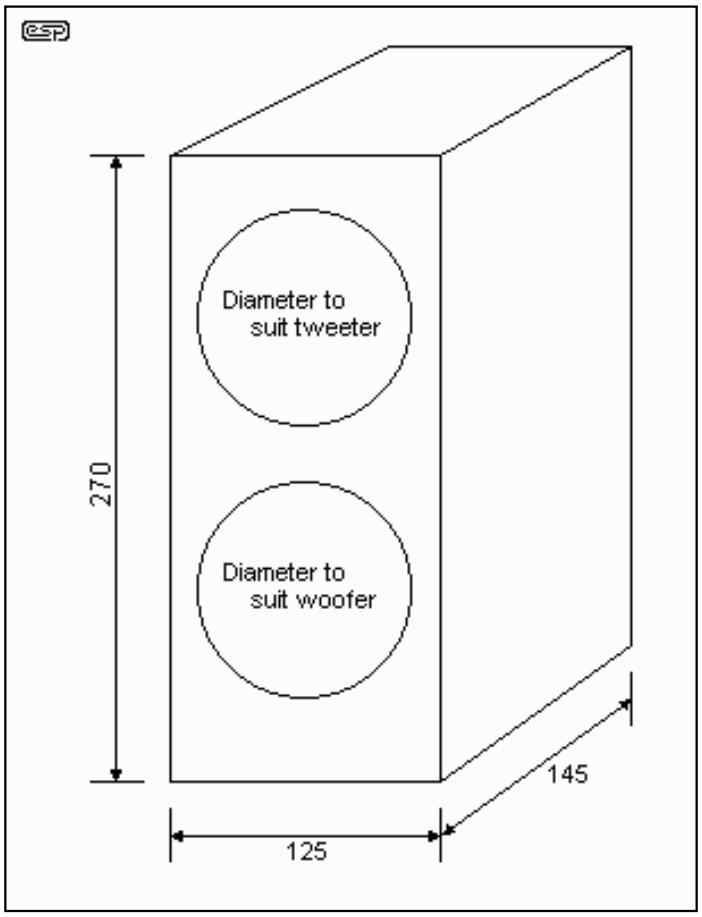


Figure 1 - Satellite Speaker Dimensions

The boxes measure 125(w) x 270(h) x 145(d) outside. I rounded off all the edges with a router, but this is not really essential if you don't have access to one. If you do, use a rounding bit with a 10mm or smaller radius, or you will remove too much of the material next to the joins.

The crossover consists of a single 470uH inductor and a 3.3uF polypropylene capacitor, wired as shown in Figure 2. This is a 6dB network, so it is important to make the crossover frequency high enough to ensure that there is sufficient attenuation at the tweeter's resonant frequency. I measured the woofer's impedance at 4kHz, and found that it was only marginally higher than at the midband frequency (about 400Hz), so decided against using an impedance correction network. Although I'm sure that this would improve matters, I suspect that it would be marginal, and not really worth the effort.

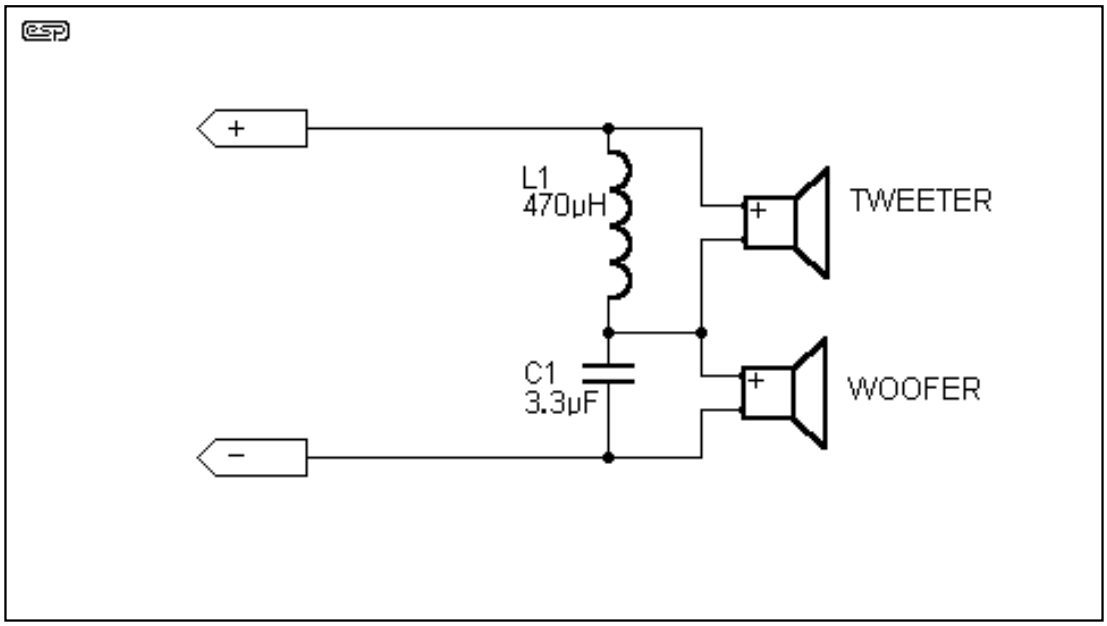


Figure 2 - Crossover Network

Since I have already established (some time ago, actually) that the series connection for a 1st order filter is to be preferred, I have changed the diagram to suit. While it is unlikely that the difference will be immediately audible in many cases, it is no more difficult to make the series version, and it does work better.

The connectors that I used were some I happened to have in my junk box, and are actually the 2-pin microphone connectors that are commonly used on CB radios and the like. They are small, airtight, and cheap (even if you have to buy them), but are a pain to mount, as they are intended for mounting in sheet metal. I recessed the mounting hole using a SpeedBor flat wood drill, and secured the final assembly with hot-melt glue. Feel free to use any connector and mounting that suits you.

Sub-Woofer

This is where all the action is, and I used the following project boards to make the final unit ...

- 1 x P3A dual 60W power amp - connected in bridge-tied-load (BTL) to obtain about 150W continuous
- 1 x P05 +/-15V power supply
- 1 x P09 stereo Linkwitz-Riley crossover, configured for a 200Hz xover frequency
- 1 x P19 dual 50W IC power amp - one amp for each satellite speaker
- 1 x P71 Linkwitz transform circuit

The entire system is powered from a 25-0-25 toroidal transformer. The one I should have used would be rated at 300VA, and I suggest that you don't use less than this or power will be severely restricted. My system is using a 160VA transformer, and this does reduce power considerably. Even so, total system power is around 200W, but can exceed this on transients. This is very loud (as in VERY loud), and may be considered overkill. Feel free to use a lower supply voltage, but the system as described can also be used as a small hi-fi and will give a good account of itself in this role. Reducing the power will limit your

options, and I will be replacing the transformer in my system as soon as I can.

The sub-woofer cabinet needs to be big enough to house the speaker, amplifiers, and the other electronics. For the sub driver, I used a 120W 200mm poly-cone, because it was reasonably close to what I knew would be needed for the job. Selection criteria for the sub are basically ...

- Low Vas - this means that a small box can be used, without raising the resonance too much
- Good efficiency - since the sub is a sealed box, equalisation is needed to get a good low frequency response. A low efficiency means that too much power is needed
- Low Fo - The lower the free air resonance, the better. The small sealed box will raise resonance, and if it's too high, excessive equalisation is needed (and an excessive power requirement results)
- High Xmax - the maximum excursion will be quite high, so the cone needs to be able to move freely
- 8 Ohms impedance - *do not use a 4 ohm driver in this project!*

The above may sound like a tall order, but such drivers exist at reasonable prices - I know, 'cos I bought one. As with the satellite speakers, there is no point getting the best driver you can - it needs to be able to do the job without distorting (too much). Remember that this is a PC speaker system, and makes no pretensions at being audiophile hi-fi.

My sub box has an internal volume of about 22 litres - not accounting for speaker displacement and the volume taken up by the electronics, which is quite small. I used 16mm MDF, which is again very sturdy and shows little or no signs of resonating because of the size. The box measures 480(w) x 280(h) x 230 (d), and again I rounded the edges - not because of diffraction, but I knew I would use the sub box as a footrest, so I wanted comfortable :-)

The box is made with a removable back, secured with lots of screws and sealed with foam tape to prevent air leaks. The back carries all the electronics, with the exception of the power transformer - the latter is mounted in the main case, as far from the electronics as I could get it. This is important (if somewhat inconvenient), because the system has high low frequency gain, and hum and buzz from the transformer is all too easily picked up.

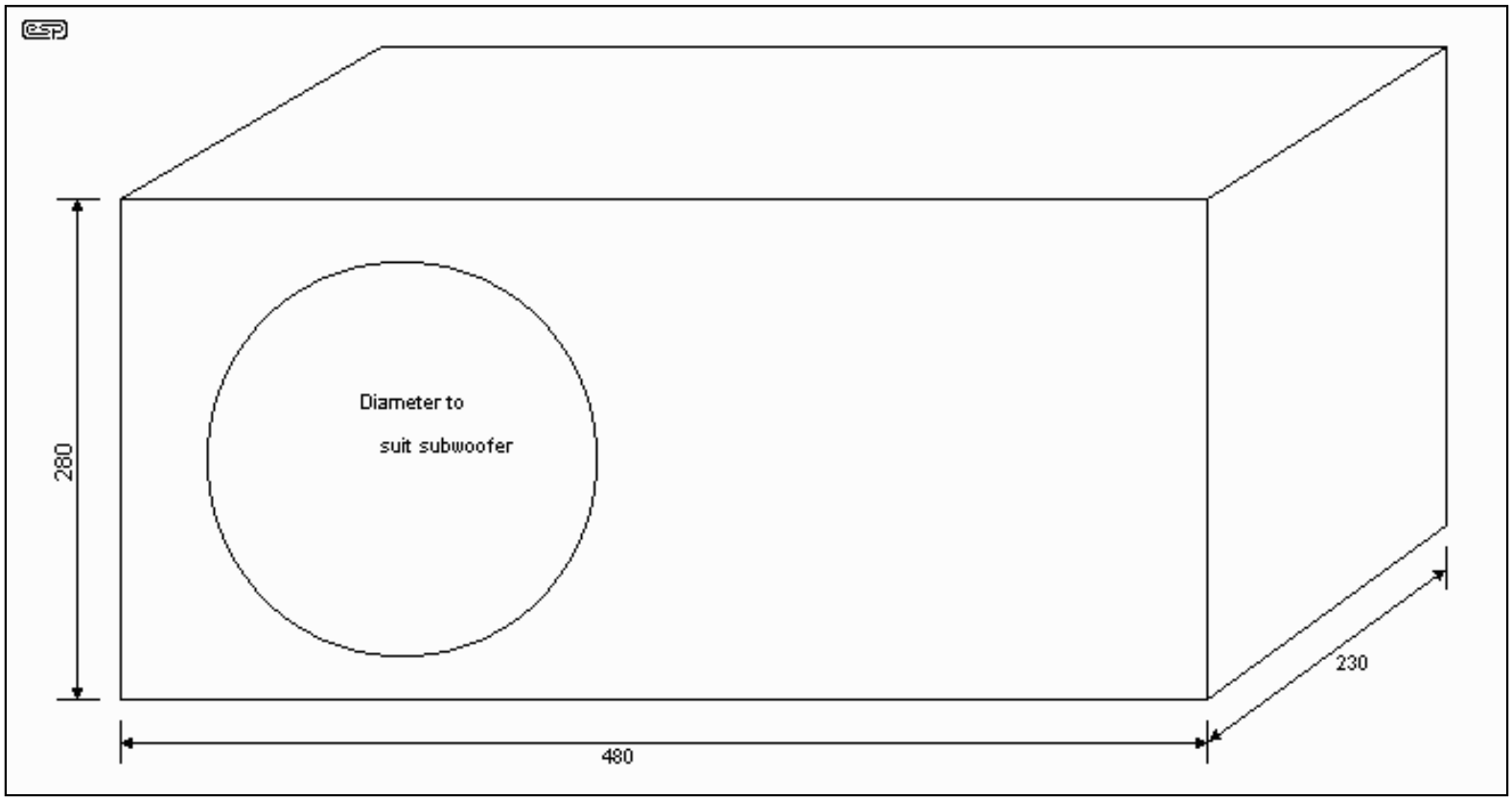


Figure 3 - Sub-Woofer Enclosure

The enclosure is not critical, but you will need to do something before you start hacking holes in the back (removable!) panel. Bolt the speaker in position. You need to do this anyway to make sure that everything fits, but in addition, include the fibreglass or other acoustic damping material, and solder a couple of thin wires to the speaker. Bring them out through a screw hole and screw the back into the box using half the screw holes. Now you can measure the speaker's resonant frequency in the box.

Attach an audio oscillator to the speaker leads. If your oscillator has a low output impedance (such as a sound card sinewave generator), then you will need a resistor in series - about 560 ohms will do fine. Carefully adjust the frequency around the expected resonance frequency (in the vicinity of 50Hz if you chose your woofer wisely). The frequency where the voltage across the speaker terminals is highest is the resonant frequency, and you will need this to double check the figures from the Linkwitz transform spreadsheet. This represents the unequalised -3dB frequency.

Using the spreadsheet, insert your speaker parameters in the "Box" page, and include the internal volume of your cabinet. The spreadsheet will calculate the values you need to properly equalise your box. In the main page of the spreadsheet, select 30Hz as the -3dB frequency, and select a Q of 0.8 for best results. The resistor and capacitor values obtained will be used in the Linkwitz transform circuit board. Have a look at the frequency response graph, and verify that the unequalised -3dB point is within a few hertz of the resonant frequency you measured. If it is not, you may need to "fiddle" the numbers a little until the graph and your loudspeaker measurement agree. Small errors are insignificant, as room acoustics will have a much greater effect than an error of a few hertz.

The Electronics

Finally, all the bits and pieces can be hooked up together. Figure 4 shows the final block diagram of the low level circuitry, and includes a level control so the bottom end can be matched to the satellite speakers. There is a deliberate 6dB extra gain in the low end to allow this, and it will normally be enough to compensate for different speaker driver efficiencies as well.

The low level electronics should be mounted on a piece (or pieces) of un-etched printed circuit board, or some aluminium sheet. The sheet must be connected to the input earth (ground) point, or it will be worse than useless. This provides some shielding, but make sure that whatever you use is securely attached to the rear panel so it can't rattle. If you wanted to, the low level stuff can all be mounted in a diecast aluminium box, but this is not really necessary unless your electrical environment is very noisy.

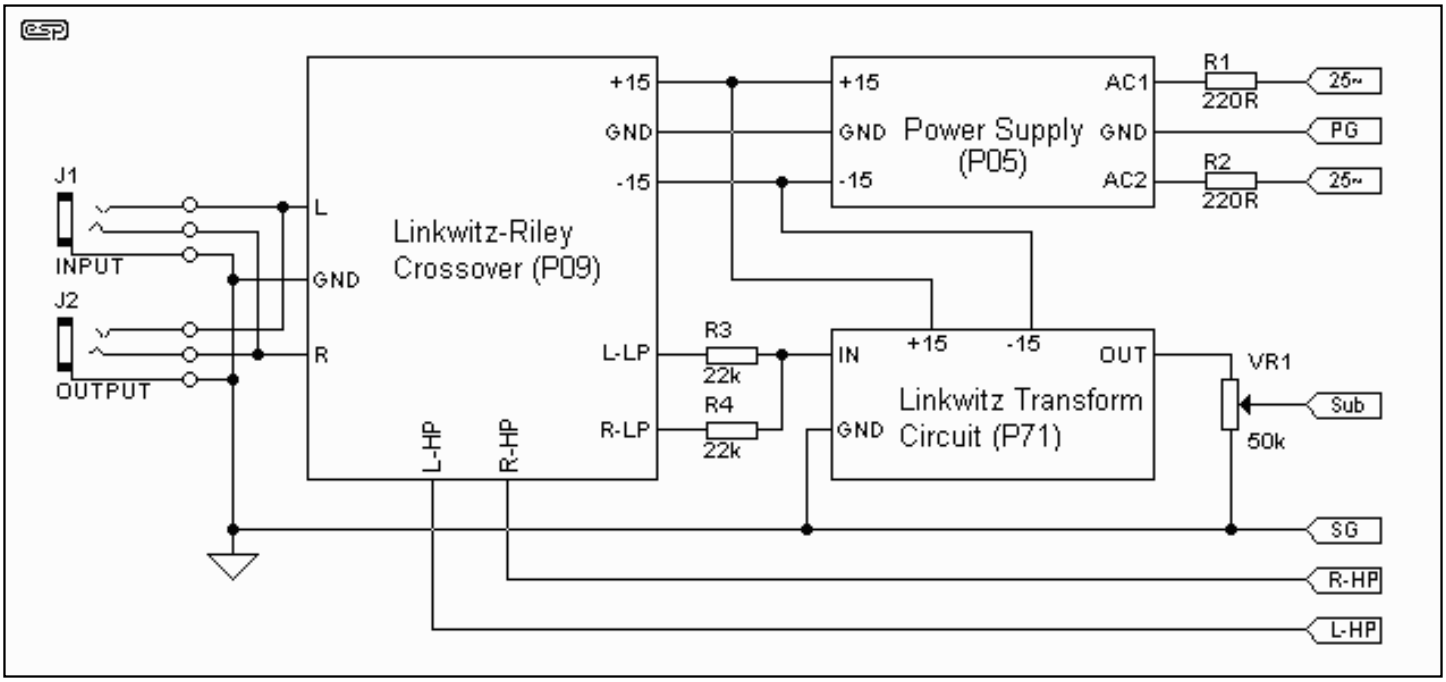


Figure 4 - The Low Level Stages

The input jacks are simply wired in parallel, so you can take an additional feed to other equipment. If you don't want to do this, leave out one of the jacks. The Linkwitz-Riley crossover is configured for a crossover frequency of 200Hz, which is nice and easy - all frequency setting resistors are 10k and caps are 100nF. None of the output buffers and trimpots (etc) are used, so leave out these sections of the circuit. The two low pass outputs are simply summed with the 22k resistors (R3 and R4). These resistors can be mounted on the L-R board, with the signal taken from the topmost trimpot hole. Full details will be supplied as part of the L-R crossover assembly instructions.

Similarly, 100 ohm resistors should be used at the high pass outputs, since shielded cable will be used to connect to the P19 power amp board. Without the resistors, the opamps can oscillate. These too can be easily mounted on the board.

The low pass combined output is connected to the input of the Linkwitz transform circuit, which will equalise the subwoofer to obtain a flat response down to 30Hz. The output of the transform circuit is connected to a pot as shown, and this needs to be accessible from outside the box so the system can be set up properly.

All the low level circuitry is powered from a P05 preamp supply. Note that I included 220 ohm resistors in series with the AC input - these must be 5W wirewound types. Since 35V is the absolute maximum DC input voltage for the 7815 and 7915 regulators, direct connection to the transformer would lead to a potential over-voltage, and the resistors tame that very nicely. The DC input to the regulators will be typically about +/-22V with this arrangement.

NOTE:- Do not operate the supply without a load from a 25-0-25 volt transformer..

Figure 5 shows the power amplifier connections. The P3A (60-100W) amp is wired up using the internal bridging resistor. An insulated wire link is needed between the output of the Right channel and the bridging resistor (R17) on the other. The input of the Left channel should be grounded with a wire link on the board. Note the 10 Ohm resistor and 100nF cap in the ground lead of the P3A power amp. This provides the high frequency ground, but the main earth line is back to the preamp (actually the crossover and equaliser). Likewise, a 1 ohm (1W) resistor is in series with the power ground to the preamp supply. These resistors were used to ensure continuity, but not allow any ground loops which will cause massive hum.

The crossover inputs are the ground reference for the whole amplifier, and the ground is physically connected to this point, where the input connectors are in contact with the input panel. This is shown in Figure 4.

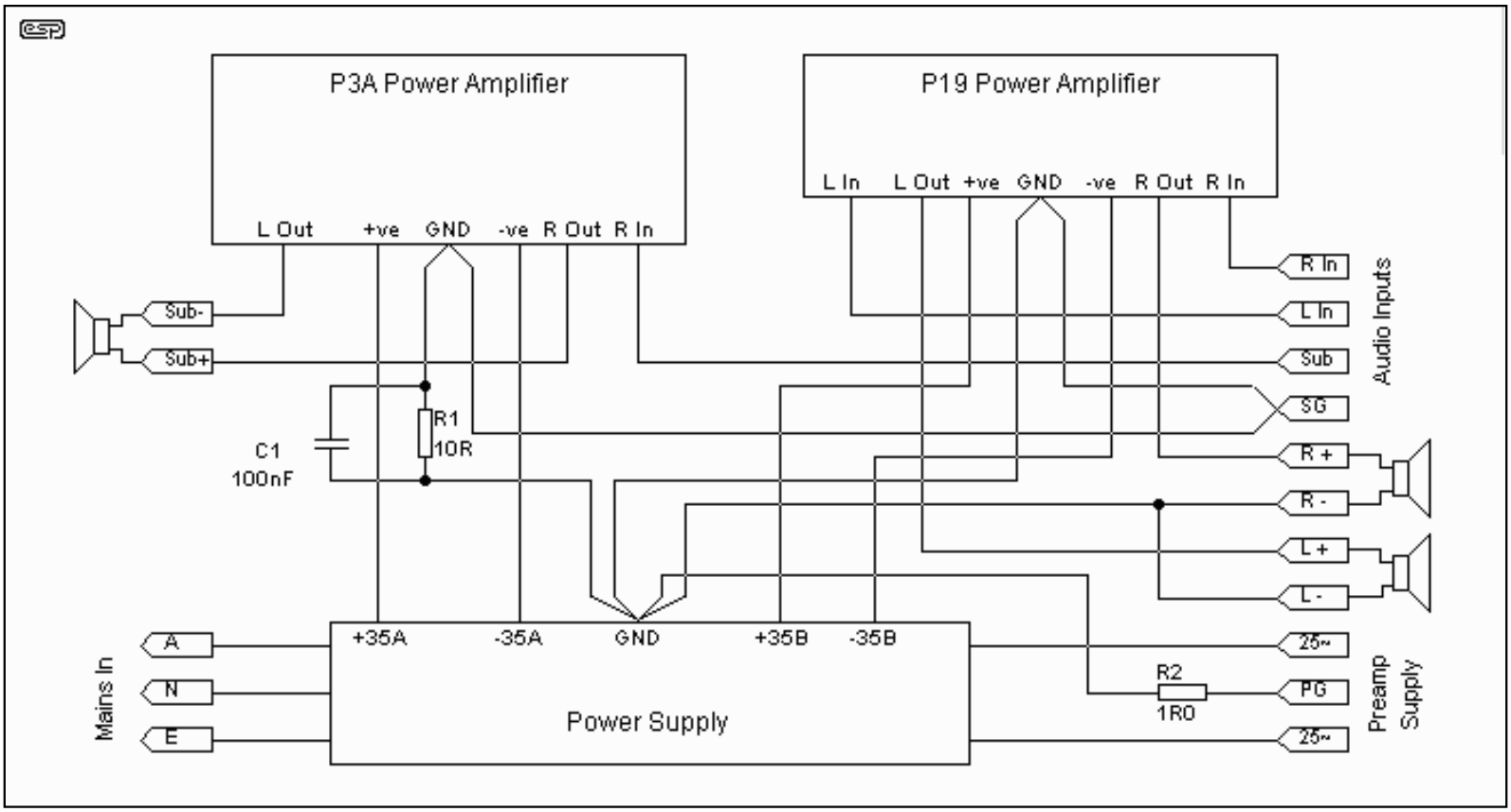


Figure 5 - Power Amplifier Connections

Before you complain, I know that the diagram is a jumble, but it actually does make sense :-) All audio leads should be shielded at the amplifier end only. Do not connect the shields at both ends, or you will create an earth loop, and hum will surely result.

The P19 50W IC power amp board provides the left and right channel feeds for the satellite speakers. These require a connector for the satellite boxes, and although fixed leads can be used, I don't recommend it. All leads should be detachable, as dangling leads are a safety hazard when you are carrying the box - be warned, it is quite heavy when everything is installed.

Heatsink selection is very important. Do not skimp on the heatsinks, as this is a high power system, and needs as much heatsink as you can afford. Use a separate heatsink for each amp module, and each should be no smaller than 1°C/W. Bigger heatsinks mean cooler running and longer life. The amplifier modules must be tested as described in the construction details before you apply power to the finished unit.

The earthing arrangement is very important, and may need some experimentation to get it right. As shown in these diagrams, it should be OK, but your layout will be different from mine, and this can have an influence on the final configuration. Note that the mains earth and signal earth must be joined, but only at one place. This is the classic "star" earthing, and must be used to avoid hum loops and other buzzes and noises. I made the star earth point the common connection between all four filter caps, and this has proved quite satisfactory. This is also the speaker return point for the satellite speakers - the sub-woofer driver is connected to the amp in bridge, so it has no earth return.

All shielded wiring should have the shield earthed at one end only. Do not rely on the shield to complete the earth connection, as noise will almost certainly be injected into the signal lead.

WARNING

Mains wiring should be carried out by qualified persons only. Do not attempt construction unless suitably qualified. Death or serious injury to yourself or others may result from a seemingly insignificant mistake.

Finally, Figure 6 shows the power supply in detail. As you can see, it is a dual supply, with only the transformer as a common point. This is not absolutely necessary, but does improve power output a little, and also prevents any possible interaction between the amps. I used an aluminium bracket fabricated from scrap sheet to secure the filter caps and act as a heatsink for the bridge rectifiers (I used 35A bridges - you don't have to, but they run cooler and are more reliable than smaller ones). The capacitors are mounted to the bracket using double-sided tape, hot-melt glue and a cable tie - I didn't want them to rattle, and you won't either.

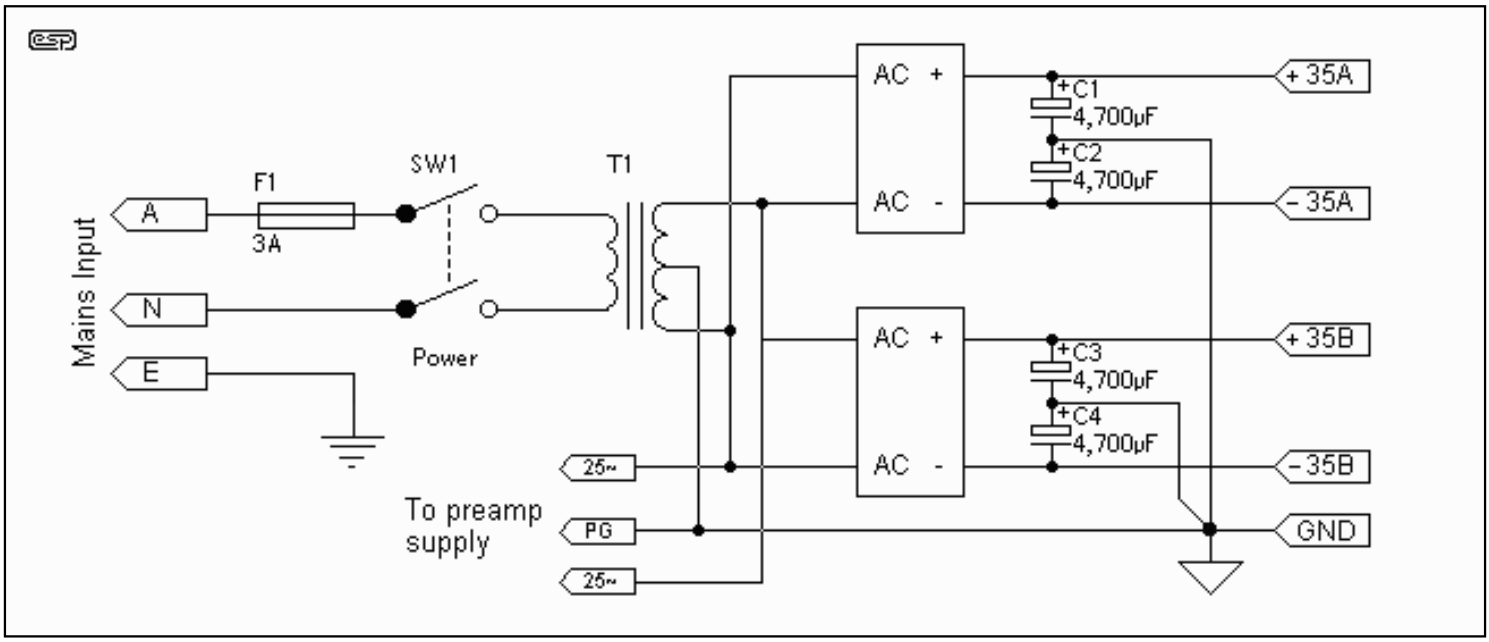


Figure 6 - Power Supply

The power supply mains wiring must be performed by a suitably qualified person, and all connections must be shrouded with heatshrink tubing to prevent accidental contact. The use of a fused IEC mains connector is highly recommended for the mains input. The designators "A", "N" and "E" mean Active (Hot), Neutral (Cold) and Earth (Ground) respectively. If you don't understand this - *don't even try to wire it!*

As previously discussed, I suggest a 25-0-25 volt transformer, preferably rated at 300VA. You will need to use an inline plug and socket (mains rated!) for the incoming AC, since the IEC mains input connector, fuse and power switch will be on the back panel, and the transformer is mounted in the case. The secondary leads should be connected using heavy duty cables, and a terminal block to join the transformer output to the bridge rectifiers and a separate light duty feed to the preamp supply.

Final Assembly

To give you a better idea of how everything fitted together in my system, Figure 7 shows the mechanical layout on the rear panel. Yours will probably be a variation on that shown, since the heatsinks must be accommodated, and yours will be different from mine. Make sure that signal and power leads are well separated, and keep speaker leads well clear of all signal leads, boards and inputs in particular - you don't need an RF oscillator !

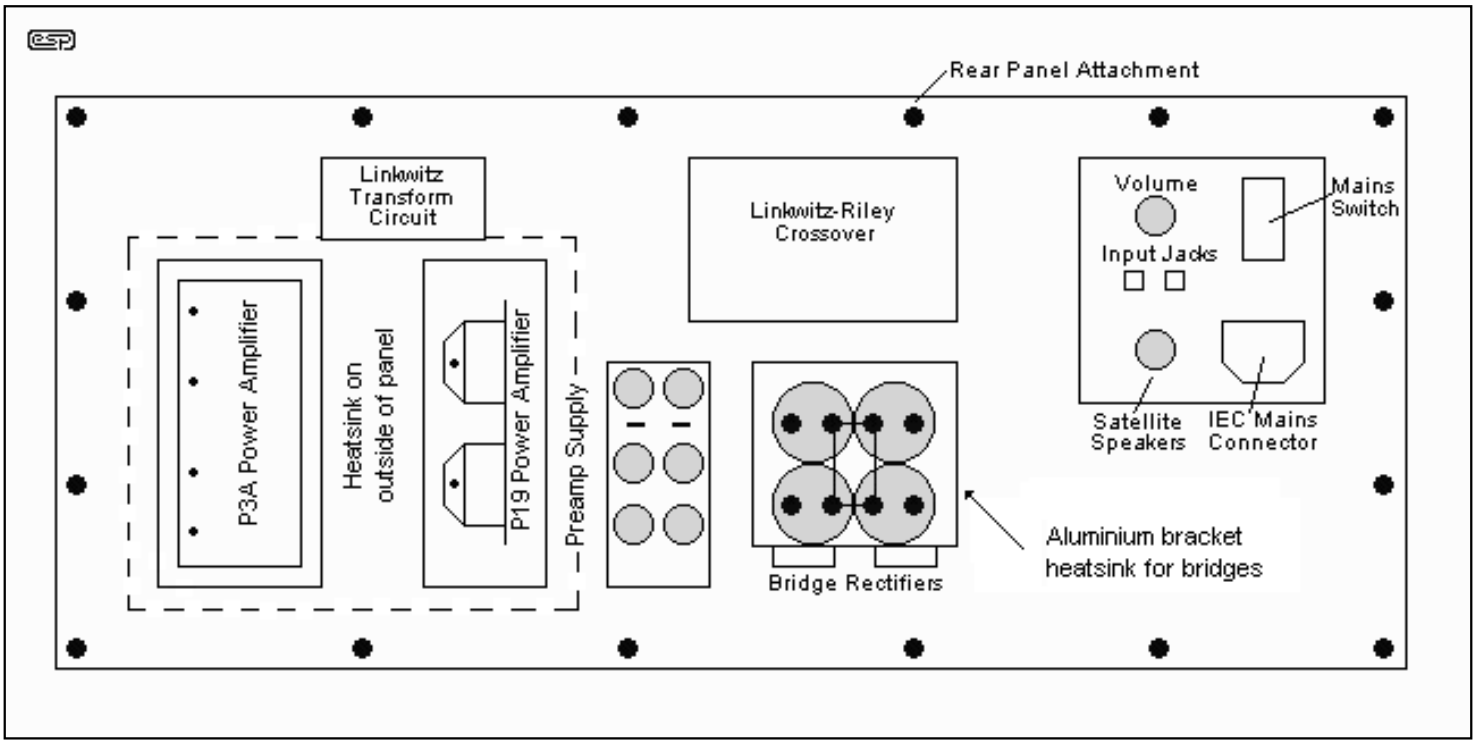


Figure 7 - Rear Panel Assembly Suggestion

The heatsink for the power amplifiers is shown dotted. This mounts on the outside of the rear panel, and must be securely fastened and sealed. The amplifiers are mounted directly to the heatsink through the holes shown (you may only need one hole, depending on your heatsink). The sheet aluminium bracket is used as a convenient mounting point for the power supply filter caps, and is also the heatsink for the bridge rectifiers.

Once everything is tested and operational, you can complete the final assembly. Make sure that nothing on the rear panel rattles before you screw it into position, and ensure that the box is completely sealed. Even a small hole will make noises at low frequencies, because of the large cone excursions and internal pressure. When powered on, the system should be silent after the small initial transient from the amps. Careful listening will reveal a slight hiss from the tweeter - this is quite normal.

Connect the system to your PC and hear your music and games as you have never heard them before. Be prepared to make more than one unit, as your friends will be green with envy after they hear yours!

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000-2004. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright © Rod Elliott 18 Jan 2001. Updated 17 Jun 2002 (added info about heatsinking), 31 Aug 2004 (changed passive xover to series).

NEW v 3! TrueRTA™ Real Time Analyzers from $39.95! LOUDSPEAKER Design Software only $79!

Selected Posts on Audio Topics

by John L. Murphy

Subj: Discussion of Group Delay in Loudspeakers

(posted around 1998 to Bass List)

Peter wrote:

> Does anyone know of demanding source material that readily demonstrates the need
> for linear phase/flat group delay?

ANY recorded music will reveal group delay errors if the errors are severe enough. Take the trivial example of a 24 hour delay between the highs and lows. The highs arrive normally but you have to wait until tomorrow to hear the lows. Would this be audible? I think so. Could we all hear it in an A/B test? YES.

The absolute audibility of group delay is not the question. If we make the delay severe enough anybody can hear it. The real question is what are the thresholds where group delay becomes "just audible". One of the most widely referenced reports of the audibility of group delay is:

> Blauert, J. and Laws, P "Group Delay Distortions in Electroacoustical Systems"
> Journal of the Acoustical Society of America
> Volume 63, Number 5, pp. 1478-1483 (May 1978)

Blauert and Laws report approximately the following thresholds for audibility:

Frequency	Threshold of Audibility
8 kHz	2 msec
4 kHz	1.5 msec

2 kHz	1 msec
1 kHz	2 msec
500 Hz	3.2 msec

The real work that needs to be done is to refine these threshold measurements and extend them further into the low frequency region (where the worst delay occurs in speakers).

In summary: Group delay is audible.

The relevant question is: "How much group delay is audible, at what frequency, with what program material."

Does anybody know of more refined measurements of these thresholds that have been reported in the literature?

Regards,

John

////////////////////////////////////
John L. Murphy
Physicist/Audio Engineer
True Audio
http://www.trueaudio.com

(posted 20Jul98 to Bass List)

Steve wrote:

>To repeat an earlier question that got lost, if I could hear group delay
>in a system, what would it sound like? Can you describe it?

The basic equivalent circuit model of a dynamic speaker system is that of an electrical highpass filter. Such filters have the characteristic of increasing group delay at low frequencies. Thus, in general, our speaker systems tend to delay the lows while allowing the highs to arrive without delay.

Let's consider the trivial case of EXTREME delay at low frequencies. For source material imagine that we use a recording of a single snare drum hit. We know what a snare drum normally sounds like. Now, imagine introducing a delay of 5 seconds on the lows. Play the drum recording through such a delay and you hear the high frequency "snap" of the drum followed 5 seconds later by the lower frequency

"thump" components. So instead of hearing one integrated "pop" sound we instead hear a "snap" followed much later (5 sec) by the "thump". We could all easily hear the delay in this exaggerated "thought experiment".

But real speakers don't add seconds of delay. They only add milliseconds or tens of milliseconds of delay. So the real question comes down to "how much delay is just audible". If we were to gradually reduce the delay in the above experiment to zero there would come a point where the delay was no longer audible as a separate distinct sound. This would probably happen above the Hass limit of 30ms or so. As the delay is reduced further there would come a point where the "delay distorted" or "time smeared" version was indistinguishable from the original recording. This delay would represent the threshold of audibility for this delay curve and source material.

Our mission as speaker designers is to keep our group delays below the threshold of audibility. In order to do this we need to first know what group delay our speakers are imparting on the music, and second, know what degrees of delay are audible.

On the subjective side I suspect that delayed bass or "late" bass is what we experience as "slow" bass. Alternately, I suspect the subjective quality of "fast bass" is likely to be associated with those systems where the bass arrives "on time" within the minimum detectable delay threshold.

For further study: Build a collection of allpass (delay only) filters and introduce stages of delay to your playback chain untill the effect is just audible in comparison with the undelayed program. Note the amount of delay that is just audible in your system. How does this compare to the delay contributed by your speaker system?

<Warning: Blatent Commercial>
Bassers, if you are speaker crazy enough to enjoy my audio ramblings then you can order my new book "Introduction to Loudspeaker Design" at either the True Audio web site or at Amazon.com. :-)

Regards,

John

////////////////////////////////////
John L. Murphy
Physicist/Audio Engineer
True Audio
http://www.trueaudio.com

(posted 2Jun99 to Audio Engineering List)

John wrote:

> GD is NOT the delay between input and output UNLESS the delay is constant
> or zero (that special constant). A constant delay results in a linear
> change in phase with frequency, but the "relative phase" of all the
> frequency components in the input (relative to each other) remains intact
> and unaltered at the output.

John is correct. group delay is a well defined variation on the concept of "time delay". It is NOT simply the difference in time between the input and output of a system. Rather it is the delay experienced by the signal "envelope" or "packet". Dick Heyser's works are well worth studying to explore these ideas further.

Here is another excellent reference on this subject which discusses phase shift, phase delay, group delay, time delay and defines a "differential time delay".

> "The Differential Time-Delay Distortion and Differential Phase-Shift Distortion as Measures of Phase Linearity" by Marshall Leach,
> JAES, Vol. 37, No.9, September 1989

Leach argues that what matters is the relative delay between the signal waveform and the envelope. He further goes on to argue that bandwidths on the order of 1 Hz (yes, one Hertz) are required in order to keep "phase shift distortion above 20 Hz" below 5%. But leach is discussing amplifiers and makes no comments on audibility of such "phase shift distortion". Generous low end bandwidth is easy to achieve in an amplifier. But with our speaker systems we do not have the luxury of dialing in an extra DECADE of low end bandwidth as a safety margin. Rather, we operate at the threshold of audibility with respect to magnitude and delay anomalies. We would like to know better where the audibility thresholds fall so that we can squeak in under them as we design speakers.

John also wrote:

> I have to backhandedly agree with Phil on the point that phase is the real
> overlooked factor. It would be nice to have zero phase shift from input to
> output (zero delay), but a constant delay is just as good, since as I noted
> above, a constant delay retains the integrity of the input wave form in the
> output.

Phil, Tom and now John have expressed preferences for "viewing" the "delay" problem in terms of phase. I just wanted to point out that (for a minimum phase system) frequency response, phase response, and group delay response are simply three different views of the same physical 'delay' phenomenon. Change one of these responses an the other will change accordingly (remember, it's a minimum phase system).

For example, given only the frequency response of a system the phase response and group delay response can be calculated. This implies that the frequency response curve contains all the information concerning both phase and group delay. So frequency, phase and group delay are really just different takes on the same information.

My current preference is to interpret this phenomenon in terms of "cycles of delay". This can be calculated from group delay by dividing the group delay by the period of the frequency at each point. I call this "normalizing" the group delay by period. If you do this the frequency dependence gets out of the way of seeing the phenomenon. For example, consider the Blauert and Laws data on audibility thresholds for group delay. Below is a table that shows the data in terms of both delay TIME in ms and normalized delay in cycles.

Frequency (Hz)	Threshold (ms)	Threshold in T (periods or cycles)
8k Hz	2 ms	16 T
4k Hz	1.5 ms	6 T
2k Hz	1 ms	2 T
1k Hz	2 ms	2 T
500 Hz	3.2 ms	1.6 T

Considering only the lower three data points for the moment, I see the Blauert and Laws data suggesting that group delay audibility thresholds appear to fall around 2 cycles. I find this "cycles of delay" view a much more intuitive representation of the data than either the phase or group delay views of the same data.

The essential question remains: what are human audibility thresholds for frequencies below 500 Hz.

Extrapolating the Blauert and Laws data to lower frequencies it could be that "one or two cycles" of delay continues to be the threshold as we move lower in frequency.

One comparison I did recently of a closed vs. a vented enclosure for ACI's SV-12 woofer revealed the following simulated results:

Frequency	Closed Box Delay (T)	Vented Box Delay (T)
100	.070 cycles	.074 cycles
50	.135 cycles	.149 cycles
20	.176 cycles	.338 cycles

Note that the vented system has more delay than the closed system but that both systems would appear to have delay well under my extrapolated B&L threshold of "one or two cycles". At 50 Hz the difference between these two systems is rather slight. But these two systems have very low F(3)'s (about 33 Hz for the closed, and about 23 Hz for the vented system). But a closed box speaker withe f(3) of 50 Hz and Q(tc) = .707 will still experience less than about .25 cycles (or 4.5ms) of delay at 50 Hz. This would SEEM to be well under the extrapolated B&L criteria of "one or two cycles". This would imply that the group delay of a 50 Hz closed box system would be inaudible. Leach's differential delay works

out to be about .475 cycles (or 9.5 ms) for this system at 50 Hz. Using Leach's differential delay makes the delay appear worse than if you consider group delay alone. (.475 cycles for differential delay vs. .25 cycles for group delay)

But what we still don't know is: "How much bass delay is AUDIBLE?"

My current opinion would go something like this: "One cycle or so of delay in the bass range is probably just audible under controlled listening conditions (phones) and with the most challenging program material (clicks and pops). Under less well controlled listening conditions and with less difficult program material the audibility thresholds are somewhat higher."

What do you think? Is the delay (whatever flavor) of a 50 Hz, Q = .707 closed box audible?

Regards,

John

////////////////////////////////////
John L. Murphy
Physicist/Audio Engineer
True Audio
http://www.trueaudio.com
Check out my new book "Introduction to Loudspeaker Design" at Amazon.com

---

(posted 3Jun99 to Audio Engineering List)

I previously wrote:

> Phil, Tom and now John have expressed preferences for "viewing" the "delay"
> problem in terms of phase. I just wanted to point out that (for a minimum
> phase system) frequency response, phase response, and group delay response
> are simply three different views of the same physical 'delay' phenomenon.
> Change one of these responses an the other will change accordingly
> (remember, it's a minimum phase system).
>
> For example, given only the frequency response of a system the phase
> response and group delay response can be calculated. This implies that the
> frequency response curve contains all the information concerning both phase
> and group delay. So Freq, Phase and Group Delay are really just different takes on
> the same information.

Tom responded (and Philip agreed):

>The statement "given only the frequency response of a system the phase
>response and group delay response can be calculated." is too often not true..

But I was careful to explicitly say that I was talking about "minimum phase systems". So I stand by my original statement as it applies to such systems. For a minimum phase system the frequency and phase responses are Hilbert Transforms of one another and can therefore be calculated from one another. However, I didn't and wouldn't claim that actual speakers are always "minimum phase systems".

To quote Dick Heyser
(from "Loudspeaker Phase Characteristics and Time Delay Distortion: Part 1", JAES, Jan 1969):

"If a network is minimum phase, there exists a unique relationship between
amplitude and phase which allows a complete determination of phase from amplitude"

So there should be no question that phase response can be calculated from frequency response for a minimum phase system.

Now, the real question becomes: 'to what extent are REAL speakers minimum phase?'

If I recall correctly, Dick Heyser expressed the opinion that loudspeaker drivers are 'largely' minimum phase systems. Otherwise why would Dick have spent so much time and energy discussing the Hilbert Transform and its use in acoustic analysis. But I couldn't find a statement to this effect in my brief search of his papers.

But I did see an AES preprint in which the author carefully measured the "excess phase" of a pair of tweeters.

AES Preprint 2118, A Micro-Computer Program for Computing the Phase Response of Dynamic Loudspeaker Systems"

The author measured the response of Dynaudio D-52 and D-21 dome tweeters and concluded: "It is seen that these two drivers are, very accurately, minimum phase devices."

I believe that deviation from minimum phase behavior is an exception for loudspeaker drivers under small signal conditions, and not the rule.

Tom went on to say:

> Zero degrees phase over the entire band (with flat amplitude response) is the
> ONLY condition which a loudspeaker can faithfully reproduce (acoustically) a
> complex non repetitive signal or asymmetric signal, just as in electronics..

This is an example of the "phase view" leading to an erroneous conclusion.

Example: If an audio signal experiences a broadband delay such that all frequencies are delayed equally, the waveform is reproduced EXACTLY in spite of the fact that the delay introduced a "zillion degrees"

of phase shift. The REAL requirement for faithful reproduction of a waveform is that the system have both flat frequency response and flat time delay. That is, all frequency components must arrive with the correct amplitude and time alignment. Flat phase response is just one case where the time alignment criteria is met. But what "delay" are we talking about? Phase delay? Group delay? I think the relevant delay here is Leach's "differential delay".

Does anybody have any measured data on the "excess phase" response of some of our popular drivers?

I am enjoying this "delay" discussion and appreciate that we may have differing opinions. I do hope I have convinced you that minimum phase systems really do have this characteristic that frequency, phase and group delay are just three different views of the same response phenomena.

In any event, do you think these "delay" effects are audible in our speakers?

Regards,

John

////////////////////////////////////
John L. Murphy
Physicist/Audio Engineer
True Audio
http://www.trueaudio.com
Check out my new book "Introduction to Loudspeaker Design" at Amazon.com

NEW v 3! TrueRTA™ Real Time Analyzers from $39.95! LOUDSPEAKER Design Software only $79!

TRUE AUDIO®

387 Duncan Lane
Andersonville TN
37705

Phone/Fax: 865-494-3388

True Audio Home Page | Catalog | Tech Topics | Audio Links | Book Store
WinSpeakerz Info | WinSpk Support | WinSpk Drivers | Try the WinSpeakerz Demo
MacSpeakerz Info | MacSpk Support | MacSpk Drivers | Try the MacSpeakerz Demo
TrueRTA Info | TrueRTA Support | TrueRTA Downloads
Introduction to Loudspeaker Design | About the Author | Box Chart | Privacy Page | Security Policy

Your comments are welcome at webmaster@trueaudio.com

Copyright©1990-2006 True Audio, Andersonville, TN USA, All rights reserved.

The Subwoofer DIY Page
Measurements

last updated: 15 April 2006

- My home system
- Audio DIY Projects
- Audio DIY Links

- my home page

There are many different ways of measuring the Thiele/Small parameters of a driver. The methods outlined below provide a way of for the beginning DIYer to measure these parameters with the minimum of equipment.

Measuring Re, Fs, Qes, Qms and Qts

To measure these parameters using the method outlined below, you'll need to have the following items:

- One (1) amplifier
- One (1) audiofrequency generator
- One (1) digital multimeter
- One (1) 5 watt resistor (approximately 4 or 8 ohms)
- Two (2) pairs of alligator clip leads

Preferably, the multimeter should be capable of measuring frequency, as well as voltage, resistance and current. The amplifier should be capable of reproducing 20 Hz to 200 Hz with no variation in output and it should be insensitive to the loads above 4 ohms. The frequency generator should also produce a signal whose voltage doesn't vary as the frequency is adjusted.

Method:

- Measure the resistance, Re, across the driver terminals.
- Measure the resistance, Rs, across the resistor.
- Connect the audiofrequency generator to the input terminals for the amplifier
- Connect the multimeter to the speaker terminals of the amplifier
- Set the audiofrequency generator to approximately 100 Hz
- Set the output of the amplifier to Vs, where Vs~0.5V to 1.0V. You may have to experiment with different voltages, depending on the accuracy of your reading for Im.
- Calculate Is, where Is = Vs/(Re+Rs)
- Connect the following circuit (using the alligator clip leads where necessary)
    - Attach one leg of the resistor to the positive speaker terminal on the amplifier
    - Attach the other leg of the resistor to the positive terminal on the driver
    - Attach the negative terminal of the driver to the negative terminal on the amplifier
    - Attach the multimeter's leads to each side of the resistor
- Adjust the frequency until the voltage across the resistor reaches a minimum level.

- Measure the frequency, Fs
- Measure the voltage across the resistor, Vm
- Calculate the current, Im = Vm/Rs, flowing through the circuit
- Calculate the driver's impedance at resonance, Rm = (Vs-Vm)/Im
- Calculate the -3db curent, Ir = (Im*Is)^0.5
- Calculate r0=Is/Im
- Calculate the -3dB voltage, Vr = Ir*Rs
- Measure the frequencies, Fl and Fh, for which the voltage across the resistor is equal to Vr
- Check that (Fl*Fh)^0.5 = Fs
- If the above check out, then Qes, Qms and Qts can be calculated as follows:

Qms = Fs*(r0^0.5)/(Fh-Fl)
Qes = (Qms/(r0-1))*(Re/(Rs+Re))
Qts = Qms*Qes/(Qms+Qes)

You can use the following spreadsheet to perform the calculations automatically for you: TSPARAM.ZIP

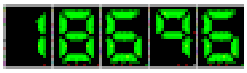
Measuring Vas (equivalent air compliance)
To measure Vas, use a good solid enclosure of known volume that's about a cube of the nominal speaker size. Determine the total volume, including the speaker cut-out and that trapped by the cone with the speaker mounted on the outside of the box for easy access. Measure the resonant frequency in this situation as well as the free air space resonant frequency.

Vas = Vb((Fb/Fs)^2 - 1)

where

Vb is the volume trapped by the speaker and box
Fb is the resonance frequency of speaker and box

my thanks to John Straede of the ***BASS and DIY Loudspeakers List*** *for providing this method of measuring the Vas of a driver*



The Audio Pages

Elliott Sound Products **Project 84**

Eight Band Sub-Woofer Graphic Equaliser

Rod Elliott (ESP) [1]

Updated 02 Feb 2002

PCBs and a set of pots for this project are available. Click the PCB image for details

Introduction

Your sub is installed and set up as best you can, but you can't quite get it to sound right. Some frequencies are too prominent, while others seem subdued. If this sounds familiar, then this equaliser is what you need to fix it. It is not a panacea, and will not cure an impossible room, but the majority of lumps and bumps in the subwoofer response will respond very well to an equaliser as described here.

The unit is an 8 band variation on the expandable equaliser described in the Project Pages, and is dedicated to its task. Boards can be stacked to get more bands if desired, but the arrangement shown will be quite sufficient for most installations.

The equaliser is a constant Q design, so unlike most "ordinary" equalisers, it does not have a very low Q at low settings of boost and cut. This is a major problem with the standard (graphic) equaliser circuit, and is completely avoided by the constant Q version. Using the Multiple Feedback Bandpass design, these filters can be designed for any (reasonable) frequency and Q desired. As a 1/3 Octave equaliser, the filter Q should be 4.3, but I have deliberately lowered this to 4 for this design to allow a little overlap.

While there will always be "that" room which defies all attempts to make anything sound halfway decent, this EQ will dispose of the majority of problems likely to be encountered.

Description

The equaliser is 1/3 Octave band, with centre frequencies at 20, 25, 32, 40, 50, 63, 80 and 100 Hz. The circuit itself uses an opamp as an input buffer (U1A), ensuring a low impedance drive to the following inverting buffer. All filters are driven by an inverted signal from U2B, and the maximum amount of boost or cut is determined by the value of R18.

U1B is a summing amp, and it takes its input from the combination of the input, and the output signal from the CUT bus - this comes from the pots used as the level control for each frequency band. The combined signal is summed again by U2A, this time with the signal from the BOOST bus added. The signal drive to

all filters is performed by U2B, the gain of which determines the maximum boost and cut allowed. As shown, The circuit will provide about +/-14dB, and the response is completely flat with all pots centred. Reduce the range by reducing the value of R18 - a value of 10k gives 6dB of boost and cut.

The actual operation of the circuit relies primarily on the amplitude and phase of the selected frequency, and it is beyond the scope of this article to cover it in great detail. The inverted signal drive is compensation for the fact that a standard multiple feedback filter is inverting at the resonant frequency - I shall leave it to the reader to work out exactly what happens (assuming you care, of course :-) For full details of the circuit topology, see Reference below.

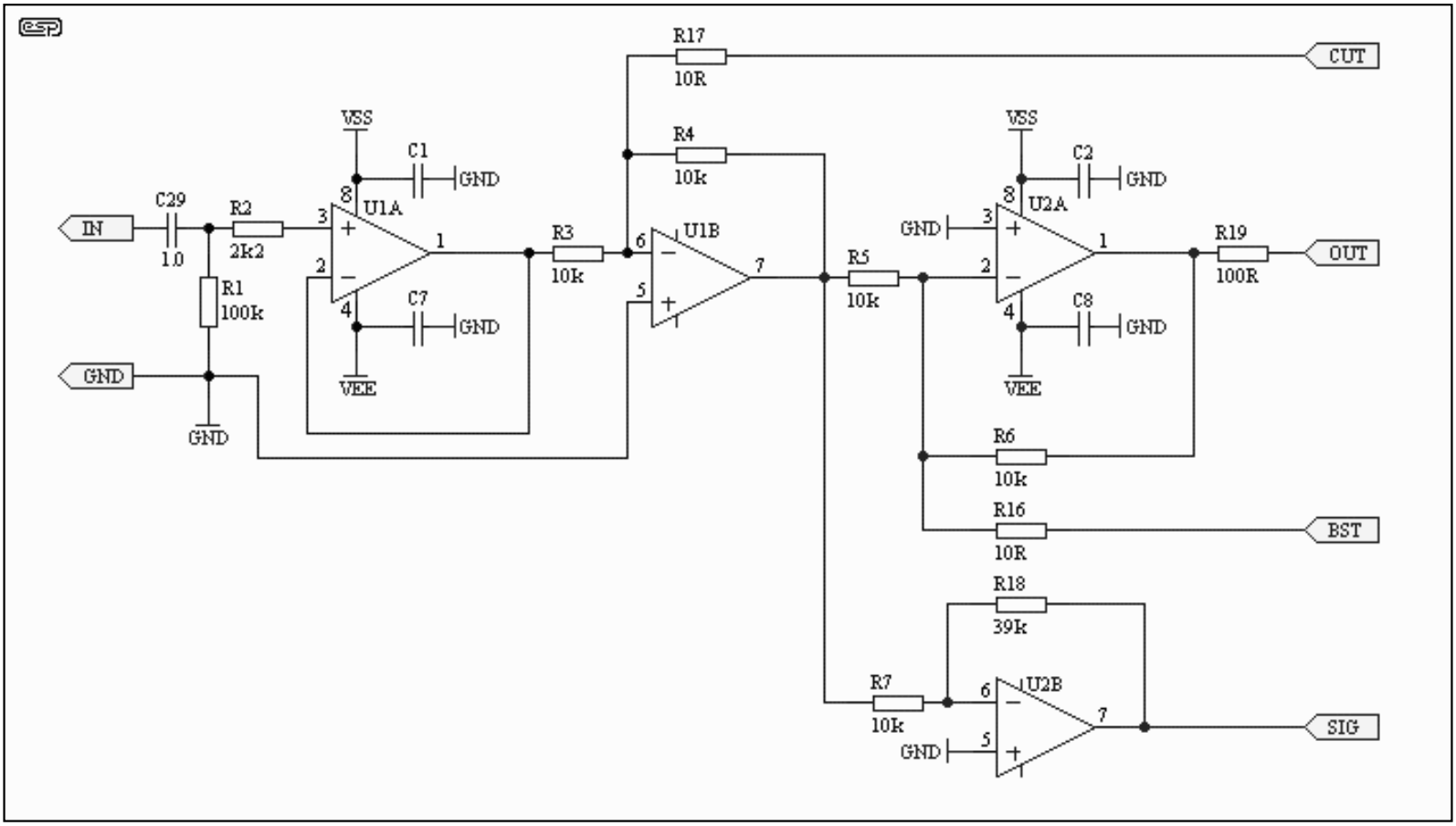
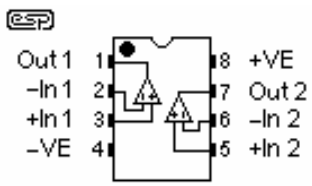


Figure 1 - Input and Output Circuitry

All unmarked capacitors (C1, C2, C7 and C8) are opamp supply bypass caps, and should be 100nF ceramic types for best performance. This is critical if high speed opamps are used, but still important if using the recommended TL072 opamps. There is little or nothing to be gained in using "audiophile" grade opamps for a subwoofer, since the TL072 has more than sufficient bandwidth for the job. Naturally if it makes you feel better, then OPA2134s or similar work beautifully.



The figure on the left (top view) shows the standard pinouts used for the vast majority or basically *all* dual opamps. If the PCB is used for this project (highly recommended, by the way), then only opamps with this pin configuration may be used. This is not a limitation :-) Correct insertion is (as always) essential, or the opamps will die ***!***

The filters are repeated - two are shown in Figure 2, and this functional block is then repeated 4 times to get the eight bands used. Again, unmarked caps C3 and C9 (and those for the additional 4 filter sections)

should be 100nF ceramic bypass types.

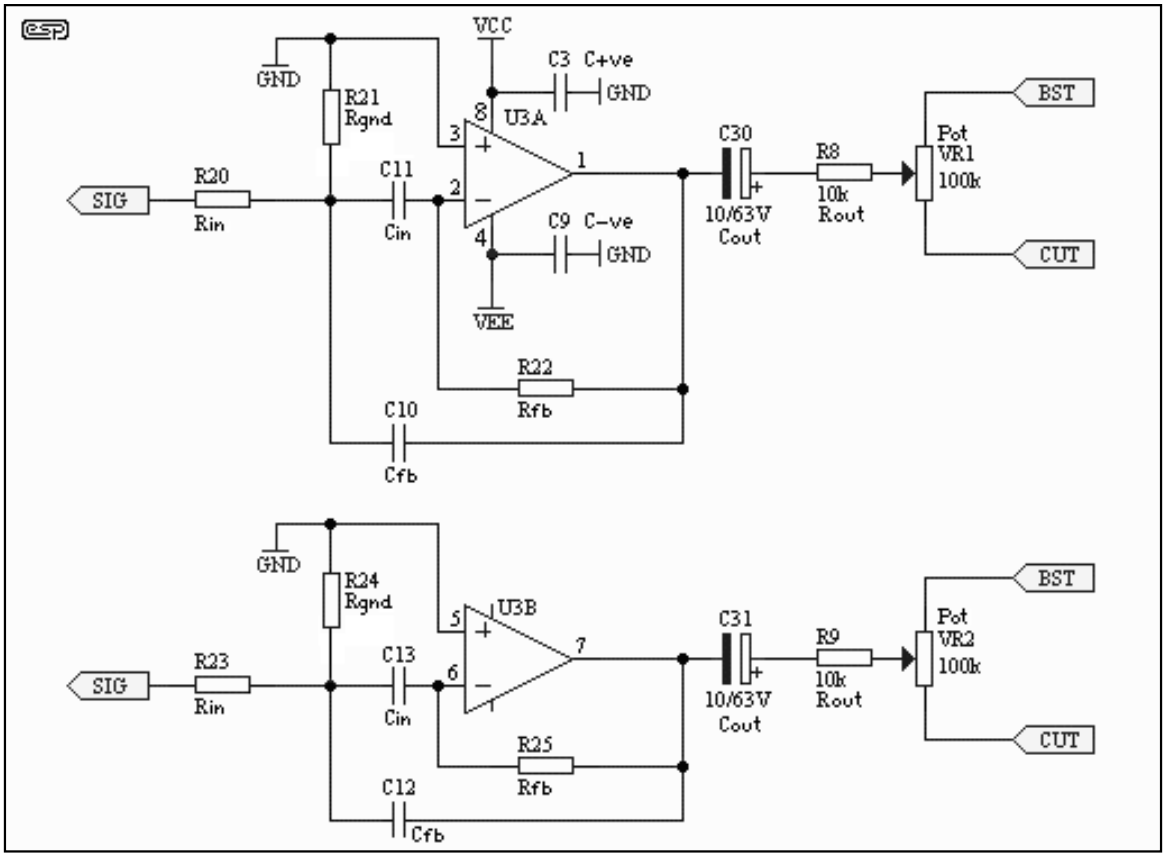


Figure 2 - Multiple Feedback Bandpass Filters

The table below shows the designations for all the filter sections. The output caps (10uF 63V as shown) may also be bipolar electrolytics - film caps would be nice, but are simply too large to fit on a decent sized board. The difference in performance is unlikely to be audible with any system, since the caps are very much bigger than they need to be at even the lowest frequency. The -3dB frequency for all output networks in this section is 1.6Hz worst case - well below anything we can hear.

	Band 1	**Band 2**	**Band 3**	**Band 4**	**Band 5**	**Band 6**	**Band 7**	**Band 8**
Pot	VR1	VR2	VR3	VR4	VR5	VR6	VR7	VR8
IC	U3A	U3B	U4A	U4B	U5A	U5B	U6A	U6B
Rin	R20	R23	R26	R29	R32	R35	R38	R41
Rgnd	R21	R24	R27	R30	R33	R36	R39	R42
Rfb	R22	R25	R28	R31	R34	R37	R40	R43
Rout	R8	R9	R10	R11	R12	R13	R14	R15
Cin	C11	C13	C16	C18	C21	C23	C26	C28
Cfb	C10	C12	C15	C17	C20	C22	C25	C27
C+ve	C3	-	C4	-	C5	-	C6	-
C-ve	C9	-	C14	-	C19	-	C24	-
Cout	C30	C31	C32	C33	C34	C35	C36	C37

The frequency selection components are shown in the following table - these are quite accurate, and will

generally be suitable for all applications. I have included the 125Hz band for those who may want to move the range up slightly - only eight of the frequencies are used for the unit.

Freq Band	Rin	Rgnd	Rfb	Cin, Cfb
20	330k	10k	680k	100nF
25	270k	8k2	510k	100nF
31	270k	8k2	510k	82nF
40	330k	10k	680k	47nF
50	330k	10k	680k	39nF
63	270k	8k2	510k	39nF
80	82k	2k7	160k	100nF
100	82k	2k7	160k	82nF
125	150k	4k7	330k	33nF

Arrrgh! The values are all over the place - this was done to avoid having to use caps in parallel (there is no room on the board), and I have tried to maintain at least passably sensible values. Unfortunately, maintaining the Q and frequency for such closely spaced filters is not easy, and the table above is the result. Feel free to use the MFB Filter calculator program to see if you can do any better - it is available from the Download page.

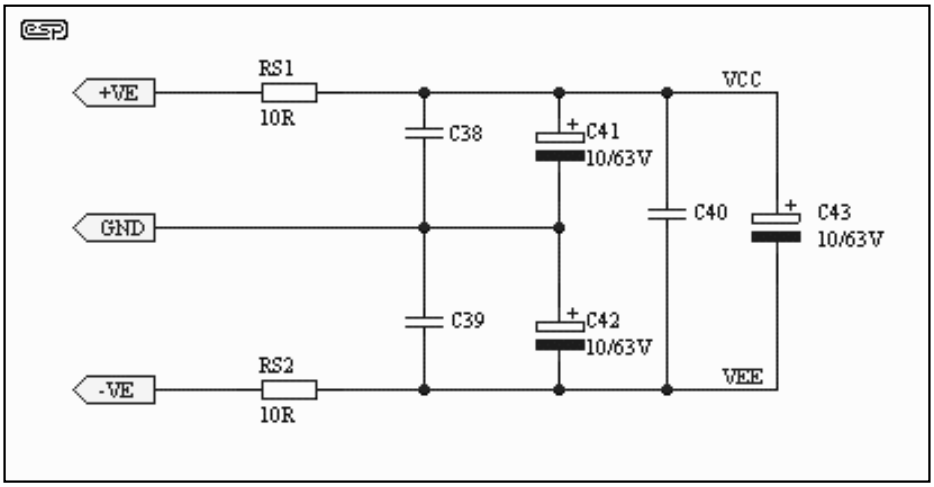


Figure 3 - Supply Conditioning Circuit

Finally, the DC input conditioning section. The 10 ohm series resistors can be omitted, but they do no harm, and help to isolate the supply from other circuitry that shares the same supply module (the P05 Power Supply is recommended for this unit). Small inductors can also be used if interference is a major problem, but these are unlikely to be needed in 99% of cases.

Although C41 and C42 are shown as 10uF 63V caps, I suggest that they be 100uF 25V - this is provided for on the PCB, which is now available.

Note

If you use the PCB for this project, it is designed to use miniature (9mm square) rotary pots, so the term "graphic" equaliser is something of a misnomer. If desired, slide pots may be used, but will have to be

wired to the board. This is not as arduous as may first be thought, since there are only 10 wires needed. If required, the 9mm pots can be obtained from ESP.

Using The Equaliser

Connect the equaliser into the signal path (usually between the source and the subwoofer equalisation and/or power amplifier). Make sure that all pots are centred for an initial flat response. Verify that the sub sounds the same as before, then preferably with a test CD (known music will work too, but is not as predictable), run a frequency sweep (or burst signals) and adjust the equaliser for the smoothest response in the low frequencies. You should be in your normal listening position for this - the sound quality will be different in different parts of the room, and this is part of the problem in the first place.

Make adjustments sparingly - over use of an equaliser is a guaranteed way to ruin the sound, so make adjustments in small increments, one band at a time. It may take a while before you are completely happy, but careful listening and perseverance are the key to getting ot right. Generally, you are more likely to need a reasonable amount of cut than boost, and although possible, it is not really practical to make the circuit asymmetrical.

Once set, the EQ settings will not need to be changed, so the unit should be placed where it is not readily accessible - you ***know*** what will happen if others know that it's there, and what it does.

The settings will need to be changed if the subwoofer is moved, or if furniture is moved, added or removed from the listening area. Large soft furnishings will make the biggest difference, while small (open) shelves will usually make little or no change at all. Book cases are highly unpredictable animals, and only careful evaluation will determine if the settings remain accurate if a bookcase is added or removed.

Reference

The design presented here is based on a paper (Constant Q Graphic Equalisers), written by Dennis A. Bohn of Rane Corporation. The original work can be downloaded from the Rane site ... http://rane.com/pdf/constanq.pdf

---

Projects Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2001. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference while constructing the project. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright © Rod Elliott 03 Jan 2002./ Updated 02 Feb 02 - added some extra info, and PCB details.

The Audio Pages

Elliott Sound Products **Why Do It Yourself?**

Copyright © 2005 - Rod Elliott (ESP)
Page Created 17 Mar 2005

Main Index

Why DIY?

Contrary to popular belief, the main reason for DIY is not (or should not be) about saving money. While this is possible in many cases (and especially against 'top of the line' commercial products), there are other, far better reasons to do it yourself.

The main one is knowledge, new skills, and the enormous feeling of satisfaction that comes from building your own equipment. This is worth far more than money. For younger people, the skills learned will be invaluable as you progress through life, and once started, you should continue to strive for making it yourself wherever possible.

Each and every new skill you learn enables the learning processes to be 'exercised', making it easier to learn other new things that come your way.

Alvin Toffler (the author of Future Shock) wrote:- *"The illiterate of the 21st century will not be those who cannot read and write, but those who cannot learn, unlearn, and relearn."*

This is pretty much an absolute these days, and we hear stories every day about perfectly good people who simply cannot get a new job after having been 'retrenched' (or whatever stupid term the 'human resources' people come up with next). As an aside, I object to being considered a 'resource' for the corporate cretins to use, abuse and dispose of as they see fit.

The skills you learn building an electronics project (especially audio) extend far beyond soldering a few components into a printed circuit board. You must source the components, working your way through a minefield of technical data to figure out if the part you think is right is *actually* right. Understanding the components is a key requirement for understanding electronics.

You will probably need to brush up on your maths - all analogue electronics requires mathematics if you want to understand what is going on. The greater your understanding, the more you have learned in the process. These are not trivial skills, but thankfully, they usually sneak up on you. Before you realise it, you have been working with formulae that a few years ago you would have sneered at, thinking that such things are only for boffins or those really weird guys you recall from school.

Then there is the case to house everything. You will need to learn how to perform basic metalworking skills. Drilling, tapping threads, filing and finishing a case are all tasks that need to be done to complete your masterpiece. These are all skills that may just come in very handy later on.

Should you be making loudspeakers, then you will learn about acoustics. You will also learn woodworking skills, veneering, and using tools that you may never have even known existed had you not ventured into one of the most absorbing and satisfying hobbies around.

Ok, that's fine for the younger generation(s), but what about us 'oldies'? We get all the same benefits, but in some cases, it is even possible to (almost) make up for a lifetime spent in an unrewarding job. As we get older, the new skills are less likely to be used for anything *but* the hobby, but that does not diminish the value of those skills one iota.

However, it's not all about learning, it's also about **doing**. Few people these days have a job where at the end of the day they can look at something they built. Indeed, in a great many cases, one comes home at the end of the day, knowing that one was busy all day with barely time for lunch, yet would be hard pressed to be able to say exactly what was achieved. What would have happened if what you did today wasn't done? Chances are, nothing would have happened at all - whatever it was you did simply wasn't done (if you follow the rather perverse logic in that last statement).

Where is the satisfaction in that? There isn't any - it's a job, you get paid, so are able to pay your bills, buy food and live to do the same thing tomorrow.

When you build something, there is a sense of pride, of *achievement* - there is

something to show for it, something tangible. No, it won't make up for a job you hate (or merely dislike), but at least you have created something. Having done it once, it becomes important to do it again, to be more ambitious, to push your boundaries.

Today, a small preamp. Tomorrow, a complete state of the art 5.1 sound system that you made from raw materials, lovingly finished, and now provides enjoyment that no store-bought system ever will.

The Financial Side

(Advertising slogan ... 'Buy NOW and save!' Translation ... buy now and **spend**)

It has to be considered that no hobby is financially 'viable' as such. People who build model planes or railway layouts or knit jumpers don't do it to save (or make) money, they do it for enjoyment, for the love of creating (or re-creating) something.

For some, it is imagined that by using the DIY approach, they will save money. No, you won't (well, you *might*, but if you do, that's a bonus, not the reason). In general, you will spend money, and if you were to add in the tools that you buy to DIY, plus the book(s) that you figured you needed, plus the costs of the occasional mistake that destroyed an amplifier's output stage, plus a bit of this and a bit of that ... it all adds up.

No. You won't save money. After the first few projects are working and have become part of the furniture, then you realise the real benefits. The gain is just a bit too low - open up the case, change a resistor, and voila! Easy. Cheap. Try getting a major manufacturer to do that with a commercial product - even worse, take it back to the shop and ask them to increase the gain by 6dB ...

- They won't know what you are talking about
- They don't care what you are talking about
- Strangely, they invite you to go away and urinate

Having built a pair of speakers and a preamp, you read an article on the Net that claims that biamping is almost magic! Well, you do have another amp lying about, but can you get everything else you need to try this for a sensible price from a shop? No. Can you make an electronic crossover yourself? Yes. This time, it *will* be cheaper than you can buy one for, because you already built your own preamp, you know what's in it, you can add the board to include a crossover. Your preamp and speakers will be out of service for a couple of days at most, and if you decide for whatever reason that you don't like it,

you can change back. The total cost in real terms is peanuts.

More importantly though, the whole process is one of learning, experimentation and experience. These are all priceless - you can't buy them. By doing it yourself you can only improve yourself. If things go wrong, this is even better (believe it or not).

There is absolutely no doubt that few things are as discouraging as a DIY project that doesn't work. Despite that, there are few things more encouraging than (eventually) finding the problem and fixing it. Don't expect it to be easy, because it almost certainly won't be. Servicing and fault finding are special skills, and are almost impossible to teach (except perhaps for a particular product where common faults are known).

To track down a fault that exists in something that has never worked is particularly difficult, but it can be done - many people do just that on a daily basis. For the DIY enthusiast, it will be harder, because most are amateurs without electronics training. Experience is one of the best forms of education - you rarely forget things you learned the hard way.

There are countless debates on the Web and elsewhere about 'esoteric' components. Some may claim that silver wire (for example) will magically transform your listening experience ... it probably won't, but that's another issue entirely. Others claim that this capacitor or that resistor is so markedly superior that nothing else should ever be used. Again, maybe, maybe not. Only with the DIY approach can these claims be tested unless you have money to burn. Having equipment modified is expensive, and there's no actual guarantee that its technical specifications will be improved - indeed, in some cases the reverse is true. You are forced to believe that it sounds 'better' regardless of technical specs. Yet again, maybe, maybe not.

If you can make the modifications yourself, then the cost is minimal. You only pay for the components, and install them yourself. If there is no improvement (or worse, performance is degraded), then it is easy and cheap to revert to the original circuit.

As a DIY person, you will also be able to make (rather than try to buy) an AB switch box so that you can make direct (blind) comparisons, and find out for yourself if there is any difference between 'ordinary' and 'magic' components. More knowledge to you either way. Should you correctly identify the magic component 70% of the time, then you *know* it really does make a difference. Likewise if you hear no difference - you *know* this because you did the test! Without this first hand knowledge, you are the mercy of the snake-oil vendors and their often very convincing sales banter, or those who say that nothing makes a difference and all amps sound the same.

While I can tell you that neither side is right for the most part, it is only with your own curiosity and test processes that you will ever know the truth. Some things may make a difference in your particular case, but only experimentation will reveal what works and what does not.

Should you simply want to use the best components you can obtain, again, this is your choice. Most of the parts you buy will be no worse (and often much better) than those used in mass-produced commercial equipment anyway, you control the standard of workmanship, and if it fails, you will be able to repair it yourself. These are all major benefits, and to get the same benefits from any commercial product, it will cost you a great deal of money. If looked at from that perspective, then you actually will save by adopting the DIY approach.

Contra-Indications

There are some important factors that you will miss out on if you follow the DIY approach, but in fairness to the hobby, they are only important to some people. Should you be the type who is impressed by the front panel, brand names, image, and fancy advertising, then DIY is **not** for you. You will get none of these things, and the appearance of the finished article will rarely be as fancy as the commercial offering.

Never mind that fact that many commercial products use a plastic front panel that may be dressed up to *look* like solid metal, or the likelihood that the internals are built on phenolic PCB (the cheapest available material). The chassis will be of thin pressed steel or maybe plastic as well as the front panel, and the top cover will almost invariably be thin sheet steel, with a spray coating.

But ... they are dressed up to look great, as long as you never remove the cover. The obviously cheap components in most consumer goods are probably not much worse than the ones you can buy, but the standard of workmanship often leaves a great deal to be desired.

None of this matters if you are only interested in the image. None of it matters if you update your gear regularly whether it still works or not. It definitely doesn't matter if you get the equipment for the right price and it does everything you ever want or need.

Of course, with time, patience and a willingness to pay for specialised work, you can build something that is vastly more impressive than the commercial offering - but if it

lacks the image you are looking for, then your friends will fail to be impressed. This is regardless of performance, which in many cases is secondary to image.

So, if any of the above applies to you and/or your circle of friends, then don't bother. No-one but you will appreciate the effort you put into it, and without the image it might as well be salvaged from the local dump, or <insert local charity here>.

The image from DIY is the one that ***you*** create, and when all is taken in context, you have something of which you can be rightfully proud. If others don't like it, that's their problem, not yours. If it happens to outperform the system they paid $thousands for, then they probably *won't* be impressed, but usually not for any of the reasons they may claim.

Conclusion

People choose DIY for the fun of creation, to learn, or to get something that can't be bought because it is too specialised - even in a seemingly minor respect. Sometimes, all three will be involved at the conscious level, but all three will be usually be involved at the subconscious level.

When you make something (even from a kit), you have the opportunity to customise it so that it does exactly what you want, not what someone else's marketing department told you you want. You will always learn from the experience of building it, even when it seems like a mindless chore stuffing components into a PCB and soldering them in. When it's finished, installed in your system, and doing exactly what you want, then the fun and pride of having made it will always be there - even long after the event.

Do you get any of these things when you buy a product? In a word, no. It is simply a commodity, something that countless others have, exactly the same as yours. If it doesn't do exactly what you want, then you have to live with it - even make excuses to yourself in extreme cases (where you'd like to strangle the salesthing given the chance).

DIY is not for everyone. Some people are forced into it because they *can't* get exactly what they want, and others do it because they think they'll save money. These are not good motives for DIY, although once they get into it, the motives will hopefully change.

The number one reason for DIY is simple - fun. Audio is a hobby for most people, and hobbies are meant to be fun - recreation at its best. In the same way that listening to your system is a recreational activity, so too is building your system yourself. As with all

hobbies, there are new skills to learn, a complete jargon to master (that part is admittedly not so much fun), and something to show for it when it is completed.

Having acquired various tools (and talents) along the way, you may find that you can use them for other DIY activities - especially woodworking tools. Again, don't expect to save money. Many goods are available that are made in China (or perhaps India or some other developing country) for far less than you could build them for. Some are real bargains - well made, and will last well in normal use. Others are terrible - cheap materials, flimsy and with a marginal finish that won't last until next Thursday.

The old saying that 'you get what you pay for' no longer holds relevance - some bargains are real, others are very obviously false. Some highly priced goods are no different from the bargains, many having been made in the same factory (some may even be *identical* to a bargain version).

Again, the DIY approach is more about satisfaction and creation than anything else. If you do happen to save money in the process, then so much the better.

Finally, to get an idea of the reasons people get into DIY for the home itself (and yes, it is relevant), have a look at the UK site Social Issues Research Centre. There are differences of course - a house and a hi-fi system tend to be rather different by their very nature, but the reasons for DIY in any form are often very similar.

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 17 Mar 2005

Elliott Sound Products **Disclaimer**

Your Rights and Obligations - DISCLAIMER - You Must Read This Section

(the fine print)

By entering this site and using any circuit or description or other material presented, then you (meaning the reader of this material being yourself, or any other person that you may subsequently pass any material to) explicitly accept the following. Should you choose not to read this information, this shall be deemed implicit acceptance of these terms and conditions in their entirety.

It is believed by Rod Elliott and Elliott Sound Products and/or Rod Elliott's heirs and assigns (hereinafter referred to as ESP) that the designs and articles presented are either original, in the public domain or may be considered common knowledge or based thereon, and do not violate the Copyright of any other party. This cannot be guaranteed however, and ESP accepts no liability for loss or damages that may be imposed upon you by any Court of Law, statutory or other body, should it transpire that a similar circuit has been simultaneously or previously designed, published or patented by another party. Nor shall ESP be held responsible for the failure of any project described to operate as described, loss or destruction of components or any other loss whether direct or consequential including loss of life or personal injury even if ESP has been advised previously of any error. It is entirely the reader's responsibility to determine the suitability of any design for the intended purpose.

The projects are presented "as is" and are believed to be without error, however this cannot be guaranteed, and it is reasonable to assume that mistakes or other errors will occur from time to time. If errors are found, please contact ESP and describe the error (and its consequences) so that corrections may be made.

No project or other activity referred to directly or by reference in this web site or any other as may be used or linked by ESP should be attempted before checking for updates or additional material that may make a substantial difference to the operation, safety or other aspect of a design, schematic or idea.

All text, diagrams, circuits and any other material presented in these project and articles pages are Copyright © ESP / Rod Elliott unless otherwise stated. These circuits and descriptions are available for personal use only, and may not be used commercially without the author's written consent nor may the projects or articles presented be copied or directly linked to other sites or re-published in whole or in part in any form whatsoever whether electronic or otherwise without the author's written consent. Violators will be vigorously pursued by any and all means at my disposal.

Likewise, ESP will not knowingly use or misuse or otherwise falsely represent the intellectual property or original ideas of others, nor re-publish or link to circuits on other sites without the consent of the author of that site, other than as a reference to my published work. All references to the work of others is acknowledged in the article or project page where such references are made. Any omissions should be referred to ESP for correction.

The material on this Internet site is made available for the purpose of providing information to the general public and is not to be taken as professional advice.

While every care has gone into the preparation of the information on this website, there are some matters of which you must be aware ...

- While we believe that the information within this site is accurate, such accuracy or reliability cannot be guaranteed. This information may change at any time
- ESP may not be held responsible for any errors or omissions in the information provided
- We recommend that you review the information in any article or project and verify it before making use of any such material
- We reserve the right to change prices without notice

Any decision to implement a strategy or recommendation contained in this web site is made solely by you. No claim may be made against ESP or any of ESP's heirs, assigns or agents for any loss or damage (arising through negligence or in any other way) which you may suffer by relying on the information on this web site.

Before relying on the material, users should obtain appropriate professional advice relevant to their particular circumstances to evaluate its accuracy, currency, completeness and relevance for their purposes.

Links provided to other Internet sites are provided for the user's convenience and do not constitute endorsement of the information at those sites. ESP accepts no responsibility for material contained in any site that is linked to this or any other ESP website.

Whenever you see this symbol, beware. Under no circumstances should any reader construct any mains operated equipment unless **absolutely** sure of his /her abilities in this area. ESP takes no responsibility for any loss, injury or death howsoever caused resulting from, whether directly or indirectly, the reader's inability to understand and appreciate the hazards of household mains or other voltages as may be applicable to any published material. All mains wiring should be performed by suitably qualified persons only. The use or otherwise of the symbol does not imply that a design is safe or otherwise, and the failure of ESP to apply the symbol does not influence the liability of ESP or the responsibility of care by you.

Main Index

Copyright Notice. All projects and articles referred to herein, including but not limited to all text and diagrams, are the intellectual property of Rod Elliott unless otherwise stated, and are Copyright © 1999-2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical or any other means, is strictly prohibited under International Copyright laws. The author /editor (ESP) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference and/or while constructing the project. Commercial use in whole or in part is prohibited without express written authorisation from Rod Elliott and the owner of the copyright in the case of articles submitted by readers or reproduced with the permission of the original Copyright holder.

Update Information: 11 Mar - Created as separate page./ 27 Apr 02 - added and amended text

Elliott Sound Products — Site Map

Copyright © 2003-2006 Rod Elliott

Page Last Updated - 11 May 2006

Indexes

Main Entry Page
Main Index
Articles Index
Projects Index

Articles

Active Vs. Passive Crossovers
A-B Blind Comparator
ABX Test Set
AM Radio - Wideband/ Hi-Fi
Amplifier Basics - Intro
Amplifier Basics - Part 1
Amplifier Basics - Part 2
Amplifier Basics - Part 3
Amplifier Basics - Part 4
Amplifier Basics - Part 5
Amplifier Basics - Part 6
Amplifier Bridging
Amplifier Design
Amplifier Efficiency
Amplifier Troubleshooting
Amplifier Protection

Projects

Projects Index
Power Amps Index
Headphone Amps Index
Preamps Index
Crossovers & Effects Index
Power Supplies Index
Musical Instrument Index
Mixers & Meters Index
Digital Audio Index
Audio Testing Index
Microphones & Preamps Index
Miscellaneous Index
Lighting Index
Numeric Project List

01 Better Volume Control
02 Simple Preamp
02 60W Power Amp
04 Power Amp Supply
05 Preamp Power Supply
05A Updated Preamp Power Supply
06 Phono Preamp
07 Discrete Opamp
08 3rd Order Electronic Crossover

Amplifier Sound

Baffle Step EQ
Balanced Circuits
Beginners Guide to Components
Bi-Amplification (Part I)
Bi-Amplification (Part II)
Bulletin Boards

Cables - Part 1
Cables - Part 2
Cables - Part 3
Cables - Part 4
Cables - Part 5
Cables - Part 6
Cable Impedance
Cable Information
Capacitors in Depth
CD Vs. SACD Vs. DVDA
Class-A Amplifiers
Class-A Amplifiers - Part 2
Class-D Amplifier Theory & Design
Compliance Scaling Loudspeakers
Compression in Audio
Counterfeit Transistors
Current Sources, Sinks & Mirrors

Derived (Subtractive) Crossover Networks
Designing with Opamps - Part 1
Designing with Opamps - Part 2
Designing with Opamps - Part 3
Distortion & Negative Feedback
Doppler distortion in Loudspeakers

Earthing (Grounding) Techniques
Electronics Tools

09 Linkwitz-Riley Crossover

10 Class-A Amplifier
11 Pink Noise Generator
12 Current Feedback Power Amp
12A El-Cheapo Power Amp
13 Low Noise Preamp
14 Power Amp Bridging
15 Capacitance Multiplier
16 Audio Millivoltmeter
17 A-Weighting Filter
18 Surround Decoder
19 50W IC Power Amp

20 Simple Bridge Adaptor
21 Stereo Width Controllers
22 Audio Oscillator
23 Clipping Indicator
24 Headphone Amplifier
25 Phono Preamps For All
26 Digital Delay
27 Guitar Amp (Original)
27 Guitar Amp (Pt 1)
27 Guitar Amp (Pt 2)
28 Quasi-Parametric EQ
29 Tremolo Unit

30 Mixing Desk (Pt 1)
30 Mixing Desk (Pt 2)
30 Mixing Desk (Pt 3)
30 Mixing Desk (Pt 4)
30 Mixing Desk (Pt 5) 31 Transistor Tester
32 Car Audio Preamp
33 Speaker Protection and Mute
34 Spring Reverb Unit
35 Direct Injection Box

Glossary of Terms

Heatsinks
Heatsinks - DIY
HEXFET Audio Amplifiers
High Speed Opamps

Impedance
Impedance Effects
Instrument Amplifiers
Introduction to Potentiometers

JLL Hood Class-A Amp
Linkwitz Transform

Meters, Multipliers & Shunts
Microphones
Miscellaneous Components

Noise in Audio Circuits
Neville Thiele Method (NTM™) Crossovers

Opamp Alternative Circuits

Passive Crossovers
Phase Angle Vs. Transistor Dissipation
Phase Correction - Myth or Magic?
Phase, Time & Distortion
Power Amplifier Clipping Behaviour
Power Ratings
Power Supply Design

Review of BE Valve Amplifier

Satellites & Subs (QB5 Alignment)
Safe Operating Area (SOA)

36 DoZ Class-A Amplifier
37 DoZ Preamp
38 Auto Power-On
39 Soft Start Power Switch
3A 60W Hi-Fi Power Amp
3B 25W Class-A Power Amp

40 Load Sensing Auto Switch
41 Opamp Design & Test Board
42 Thermo-Fan Controller
43 Simple DC Split Supply
44 Dual Lab Supply
45 Bass Compressor
46 Thermal Shutdown
47 VOX AC30 Simulator (withdrawn)
48 Sub Equaliser
49 Guitar Vibrato

50 Mic Circuit Tester
51 Balanced Line Drivers
52 Distortion Analyser
53 Power Limiter
54 Low Power FM Transmitter
55 PPM and VU Meters
56 Variable Amp Impedance
57 Sound Impairment Monitor
57 Speaker Measurement Set
59 Self Oscillating Amp

60 LED VU Meter
61 (Withdrawn)
62 Lighting System (Pt 1)
62 Lighting System (Pt 2)
62 Lighting System (Pt 3)
62 Lighting System (Pt 4)
62 Lighting System (Pt 5)
62 Lighting System (Pt 6)

(The) Subwoofer Conundrum
Series vs. Parallel Crossovers
Sound Impairment Monitor
State of Manufacture

Thiele-Small Parameters
Transistor Matching
Transistor listing
Transformers - Part 1
Transformers - Part 2
Transformers - Part 3
Triamped Speaker System - Part 1
Triamped Speaker System - Part 2

Voltage Dividers & Attenuators

What Makes Tweeters Blow
What Is Hi-Fi
Why Do It Yourself?
Wiring A Power Supply

Application Notes

Application Notes Index
Precision Rectifiers
Analogue Metering Amplifiers
High Power Led SMPS
Car Dome Light Extender
Zero Crossing Detectors
Ultra-Simple SMPS
Transistor Assisted Zener
Using Zener Diodes

Purchases

63 MFB Filter
64 Instrument Graphic EQ
65 Strobe Light
66 Balanced Mic Preamp
67 Audio Peak Limiter
68 Sub Power Amp
69 Low power +/-12V supply

70 DoZ Headphone Amp
71 Linkwitz Transform Circuit
72 20W IC Stereo Amp
73 PC Sound System
74 RF Probe
75 Constant Q Graphic EQ
76 Opamp Based Power Amp
77 High Current 13,8V Supply
78 2nd Order Electronic Crossover
79 Current Sense Power Switch

80 Reverse RIAA Equaliser
81 12dB L-R Crossover
82 Loudspeaker Test Box
83 MOSFET Follower Amplifier
84 1/3 Octave Sub Equaliser
85 S/PDIF DAC
86 Mini Audio Oscillator
87 Balanced Transmitters & Receivers
88 Audio Preamp
89 Car Switchmode Supply

90 Dimmer Polarity Reversal
91 78RPM Phono EQ
92 Guitar and Bass Sustain
93 Recording & Measuring Mic
94 Universal Preamp Mixer
95 Low Power -VE Supply
96 48V Phantom Supply

Ordering Info
Price List
Order Form
Encrypted Order Program
Downloadable price file for above
Fully Built Modules
Class-D amplifier
Guitar Amp Module
Digital Flash Module
VP-103 Valve Preamp
PCB Catalogue (HTML)
Paymate Details
PayPal Details
Dead Letters

Gallery

L-R Crossover
L-R Crossover (2)
Reader Feedback - Part 1
Reader Feedback - Part 2

Downloads

Download Index
Encrypted Order Form (ZIP)
Semiconductor Data (EXE)
Heatsink Calculator (ZIP)
MFB Filter Calculator (EXE)
Help Kill Spam (ZIP)
LED VU Design (ZIP)
Loudspeaker Parameters (ZIP)
Crossover Design (ZIP)
Linkwitz Transform Design (ZIP)

97 Hi-Fi Preamp
98 Auto Charger for Battery Hi-Fi
99 36dB/ Octave Subsonic Filter

100 Headphone Adaptor for Amplifiers
101 200W MOSFET Power Amplifier
102 Simple Pre-Regulator
103 Subwoofer Phase Control
104 Preamp/ Crossover Mute Circuit
105 Build an Electrostatic Loudspeaker (ESL)
105-2 Electrostatic Loudspeaker - Part 2
105-3 Electrostatic Loudspeaker - Part 3
106 h_{FE} Tester for Transistors
107 Phase / Polarity Reversal Switching
108 Switchmode PSU Protection
109 Headphone Amp

110 IR Remote Control
111 PIC Speaker Protection
112 Dummy Head Microphone
113 PIC Hi-Fi Headphone Amp
114 Class-D (PWM) Power Amp
115 GainClone Amplifier (Part 1)
115 GainClone Amplifier (Part 2)
116 Class-D Subwoofer Amp
117 1.5kW Power Amplifier

Project Suggestions

Reminder Program (EXE)
Stores Program (ZIP)
PC Timer Program (EXE)
Crossover Calculator (EXE)
Heatsink Design (XLS)
Transformer analysis program (ZIP)
Transformer analysis spreadsheet (ZIP)
Transformer design program (ZIP)
World Time Clock (EXE)

Humour

Humour Index
Humour - Part 1
Humour - Part 2
Murphy's Laws
A Space Oddity
If Your O/S Were an Airline
Waltzing Matilda Explained
Monty Python
Dynamic Range Vs, Ambient Noise
Satcure Products

Miscellaneous

Home Page
Main Index
Main Links Page
Miscellaneous Links
Mad As Hell - Part 1
Mad As Hell - Part 2
Mad As Hell - Part 3
Mad As Hell - Part 4

ESP Ethics Policy
About ESP Site
Contact the Author
Submit Your Material
Contracting Details
About the Author
Custom Projects for Education
Frequently Asked Questions
Spam, Scams & Security
Scams - how they work
Your on-line security
Spam, Spam, Spam, Spam
Disclaimer
Search Page
Contribute Project/ Article

All documents referred to herein are Copyright, and may not be reproduced, copied or republished by any means whatsoever, either in whole or in part, without the express written permission of Elliott Sound Products (ESP). Copyright is maintained even by the omission or misrepresentation of any document, drawing, image or other material as may be presented on this site, or any other site as may be maintained or operated by ESP.

ESP reserves the right to make changes, updates, modifications or additions to any document, drawing or image at any time, and copyright shall be automatically extended to the changed, updated, modified or added material. Errors and omissions may occur within articles or other material, and no claim may be made against ESP for any damages howsoever caused.

Note that some links on this page may refer to documents in preparation, and these may not be available at any given time.

Your search for " did not produce any results.

Here are some suggestions to improve your search:

- Select "Match ANY word"
- Make sure you spelled the words correctly
- Search " with Google

Search again: Match ANY word Search

Perlfect Search 3.31 (c) Perlfect Solutions 1998-2003

Latest search index update: 2005-05-31 21:02

ESP Projects and Articles

Dedicated to ESP Projects and Discussions

The time now is Tue May 30, 2006 2:27 am

ESP Projects and Articles Forum Index

View unanswered posts

Forum	Topics	Posts	Last Post
Announcements			
Forum & Site Information Miscellaneous announcements, general forum info and new stuff on the ESP Website. Moderators Scorpion, richie00boy	4	4	Fri Jun 17, 2005 11:12 pm espadmin
Special News Items Information about fake components, or warnings about dishonest vendors. Please make sure that your information is factual (no hearsay) and does not leave you open to libel laws. Moderators Scorpion, richie00boy	6	52	Thu Mar 16, 2006 3:48 am jaycee
Projects			
Power amplifiers P3A, P68 or any of the other ESP power amplifier designs Moderators Scorpion, richie00boy	179	1846	Tue May 30, 2006 12:41 am Reactance
Preamps Preamplifiers - phono preamps, mic preamps, all kinds of preamps Moderators Scorpion, richie00boy	71	559	Mon May 29, 2006 11:40 pm richie00boy
Power Supplies All kinds of power supplies, transformers, regulators, etc. Moderators Scorpion, richie00boy	108	931	Mon May 29, 2006 5:56 pm Paul_J
ESP Projects Comment or share ideas about miscellaneous ESP designs here Moderators Scorpion, richie00boy	107	691	Tue May 30, 2006 1:03 am mikee55
General			

General Information Swap ideas and comments about electronics Moderators Scorpion, richie00boy	120	858	Mon May 29, 2006 6:03 pm rohitbd
Digital DACs, ADCs, CD-Roms and other digital equpiment Moderators Scorpion, richie00boy	6	82	Thu Apr 06, 2006 3:39 am richie00boy
Loudspeakers Building, measuring and testing speakers and systems Moderators Scorpion, richie00boy	71	603	Mon May 29, 2006 11:47 pm rohitbd
Crossovers Active and Passive Crossover Discussions Moderators Scorpion, richie00boy	28	146	Sun May 28, 2006 3:03 am richie00boy
Application Notes Discussions on any of the ESP App Notes Moderators Scorpion, richie00boy	1	13	Wed Jul 06, 2005 11:51 pm vkroenen
Class-D Curious about Class-D amps? Post questions or comments here Moderators Scorpion, richie00boy	7	58	Tue Mar 28, 2006 7:58 pm peranders
Miscellaneous Stuff that just doesn't seem to fit anywhere else Moderators Scorpion, richie00boy	57	485	Sun May 28, 2006 7:33 am AaronV
Gallery			
Preamps Links or photos of preamp projects - with brief description Moderators Scorpion, richie00boy	6	58	Mon May 29, 2006 6:05 am Electro-Dan
Power Amps Links or photos of power amp projects - with brief description Moderators Scorpion, richie00boy	23	350	Fri May 05, 2006 5:29 am roren
Loudspeakers Links or photos of speaker projects - with brief description Moderators Scorpion, richie00boy	6	90	Sat May 13, 2006 4:57 am roren
Miscellaneous Links or photos of other projects - with brief description Moderators Scorpion, richie00boy	12	53	Tue May 23, 2006 5:25 am Electro-Dan
Buy, Sell, Swap			

	Buy, Sell or Swap Buy, sell or exchange parts, components or complete systems. Commercial posts must obtain approval first, or your post will be deleted. Moderators Scorpion, richie00boy	14	50	Fri Apr 21, 2006 2:13 am ll81
Off Topic				
	Off Topic Want to discuss something that has nothing to do with ESP projects (or even electronics)? This is the place! Moderators Scorpion, richie00boy	57	496	Mon May 29, 2006 8:57 am Nerval_Heavy_Ind.

Mark all forums read

All times are GMT + 10 Hours

Who is Online

Our users have posted a total of **7425** articles
We have **428** registered users
The newest registered user is **mikee55**

In total there is **1** user online :: 0 Registered, 0 Hidden and 1 Guest [Administrator] [Moderator]
Most users ever online was **43** on Tue May 02, 2006 12:58 pm
Registered Users: None

This data is based on users active over the past five minutes

Log in

Username: Password: Log me on automatically each visit Log in

New posts No new posts Forum is locked

Powered by phpBB © 2001, 2005 phpBB Group

Elliott Sound Products **Price List**

Price List Vers 1.21 - 02 Feb 2006

Updated 26 May 2006

This is the latest pricelist. The prices shown are current as of the date shown above and as always, will remain stable for as long as possible. Using a credit card will usually get you the best exchange rate possible at any point in time.

Please see the purchase information for complete purchase details, order forms, warranty information, etc.

ESP implemented a loyalty programme, effective as of 1st July, 2005. Click here for more information about the programme, redeeming discount vouchers, and the terms and conditions for earning 'loyalty points'.

For added peace of mind, your package may be sent as registered mail (within Australia only), or insured (International only) The codes and prices are shown below. Be sure to check the details below first, and if you are still unsure, send me an e-mail. International post can be very frustrating because of the different regulations world-wide.

Order Code	Description	AUD (Aust)	AUD (RoW *)	USD (RoW *)	
REG	Registered Mail to a value of AU$100 (Australia Only)	AU$ 5.50	n/a	n/a	
INS1	Insurance to AU$100 (Overseas Only)	n/a	AU$15.00	US$12.50	
INS2	Insurance to AU$200 (Overseas Only)	n/a	AU$17.00	US$14.20	
INS3	Insurance to AU$300 (Overseas Only)	n/a	AU$19.00	US$16.00	
	Pack & Postage	AU$ 5.50	AU$12.00	US$10.00	

Where registered mail (Australia) or insurance (RoW) is requested, the fee for either is in addition to pack and postage charges.

* RoW - Rest of World

Please note that insurance is not available to all countries. The ones I know that will not accept insurance are ...

> **Canada**(see below), Indonesia, Israel, Malta, Mexico, Philippines, Solomon Islands, Vietnam, Botswana

There may be others, but I have not had to deal with them at this stage. In addition ...

- Germany will accept insurance on letters only - packages may not be insured. Most articles are sent as 'large letter', but some items (notably bulky things like modules, microphone housings) must be sent as packages.
- The US only accepts insured packages - large letters (the majority of my outgoing mail) cannot be insured. Where insurance is requested the goods will be sent as a package (***which costs me a lot more!***).
- Canada has recently been declared an insurance free zone by Australia Post, so Canadians can no longer obtain insurance (don't ask - I have no idea why). The best I can offer is 'delivery confirmation' at AU$5.00 in addition to normal postage charges.
- Switzerland insists you must collect insured items from the Post Office
- Sri Lanka has a AU$60 maximum limit for insurance

If unsure, See the Australia Post web site for more info, or contact ESP for confirmation.

Note Carefully: If you choose not to take out insurance sufficient to cover the full cost of the order, my (official) responsibility stops upon lodgement of the package at the Post Office. Should your package go missing (very rare, but it has happened), then the loss is yours. I will always endeavour to track what happened on your behalf and recover your goods if possible, and in the majority of cases a mutually agreeable solution can be found that minimises both your loss and mine. Please note that if an insurance claim must be made, there is extensive paperwork, and an even more extensive waiting period as Australia Post argues with the destination postal service. This can typically take two *months* or more! (Yes, I know this is stupid beyond belief, but I didn't make the rules.)

Although most people do not have a high opinion of the postal services, it is my experience that international mail is treated by Australia Post and the postal services of all major countries with the highest integrity.

> **Notes:**

- Available only with PCB purchase. *
- RoW - Rest of World (Anywhere outside Australia).
- US$ Prices are for PayPal Purchases *only!*

Unless noted otherwise, all PCBs are sold as a bare board, and *no* components are included.

Order Code	Description - Fully Built Modules	AUD (Aust)	AUD (RoW)	USD (RoW)	Details
BP4078 [1]	400W Class D power amp module (available now new) 1 piece price ...	AU $206.80 Ea	AU $188.00 Ea	US $157.00 Ea	More ...
M27A	100W Guitar Amp Module (Built & Tested)	AU $145.50 Ea	AU $135.00 Ea	US $113.00 Ea	More ...
DF100M	Digital Flash Trigger Module	AU$66.00 Ea	AU$60.00 Ea	US$50.00 Ea	More ...
Order Code	**Description - PCBs**	**AUD (Aust)**	**AUD (RoW)**	**USD (RoW)**	**Details**
P05	Preamp Power Supply (Note - Rev-A board now shipping)	AU$19.80 Ea	AU$18.00 Ea	US$15.00 Ea	More ...
P06	High Quality Phono Preamp	AU$17.60 Ea	AU$16.00 Ea	US$14.00 Ea	More ...
P09	Linkwitz-Riley Electronic Crossover	AU$26.40 Ea	AU$24.00 Ea	US$20.00 Ea	More ...
P19	Single Chip Stereo 50W Amplifier	AU$24.20 Ea	AU$22.00 Ea	US$19.00 Ea	More ...
P27A	100W Guitar Power Amplifier (Mk II)	AU$28.60 Ea	AU$26.00 Ea	US$22.00 Ea	More ...
P27B	Guitar Preamplifier	AU$19.80 Ea	AU$18.00 Ea	US$15.00 Ea	More ...
P33	Loudspeaker Protection and Muting	AU$16.50 Ea	AU$15.00 Ea	US$12.50 Ea	More ...
P36	Death of Zen Class-A Amplifier (2 PCBs - Rev-A board now shipping)	AU$28.60 Pr	AU$26.00 Pr	US$22.00 Pr	More ...
P37	DoZ Preamp and Power Supply (2 PCBs)	AU$28.60 Pr	AU$26.00 Pr	US$22.00 Pr	More ...
P39	Power Transformer Soft-Start Circuit (2 PCBs)	AU$22.00 Pr	AU$22.00 Pr	US$17.00 Pr	More ...
Order Code	**Description - PCBs (Continued)**	**AUD (Aust)**	**AUD (RoW)**	**USD (RoW)**	**Details**

P3A	60W / 8 Ohm Stereo Power Amplifier	AU$28.60 Ea	AU$26.00 Ea	US$22.00 Ea	More ...
P48	Subwoofer Equaliser & Controller	AU$16.50 Ea	AU$15.00 Ea	US$12.50 Ea	More ...
P66	Dual Low Noise Microphone Preamplifier	AU$22.00 Ea	AU$20.00 Ea	US$17.00 Ea	More ...
P68	300W Subwoofer Amplifier	AU$35.20 Ea	AU$32.00 Ea	US$27.00 Ea	More ...
P68-2	Dual Board 300W Subwoofer Amplifier	AU$57.20 Pr	AU$52.00 Pr	US$44.00 Pr	More ...
P70	Use Order Code P36				
P71	Linkwitz Transform Circuit	AU$19.80 Ea	AU$18.00 Ea	US$15.00 Ea	More ...
P72	Stereo 20W/ Channel IC Amplifier	AU$24.20 Ea	AU$22.00 Ea	US$18.50 Ea	More ...
P84	8 Band Constant Q Graphic EQ for Subwoofers	AU$34.10 Ea	AU$31.00 Ea	US$26.00 Ea	More ...
P84P	8 Band Constant Q Graphic EQ Pot Set (8 Pots) *	AU$34.10 Set	AU$31.00 Set	US$26.00 Set	More ...
Order Code	**Description - PCBs (Continued)**	**AUD (Aust)**	**AUD (RoW)**	**USD (RoW)**	**Details**
P87A	Dual Balanced Line Level Receiver	AU$16.50 Ea	AU$15.00 Ea	US$12.50 Ea	More ...
P87B	Dual Balanced Line Level Transmitter	AU$16.50 Ea	AU$15.00 Ea	US$12.50 Ea	More ...
P88	High Quality Audio Preamplifier	AU$19.80 Ea	AU$18.00 Ea	US$15.00 Ea	More ...
P93	Microphone Capsule Amplifier	AU$15.40 Ea	AU$14.00 Ea	US$12.00 Ea	More ...
P94	Universal Preamp & Mixer	AU$19.80 Ea	AU$18.00 Ea	US$15.00 Ea	More ...
P96A	48V Phantom Feed Supply	AU$16.50 Ea	AU$15.00 Ea	US$12.50 Ea	More ...
P96B	Phantom Feed Microphone Interface	AU$16.50 Ea	AU$15.00 Ea	US$12.50 Ea	More ...
P97	Complete Hi-Fi Preamplifier	AU$26.40 Ea	AU$24.00 Ea	US$20.00 Ea	More ...
P99	Stereo Subsonic Filter	AU$16.50 Ea	AU$15.00 Ea	US$12.50 Ea	More ...

P101	200W MOSFET Power Amplifier	AU$22.00 Ea	AU$20.00 Ea	US$17.00 Ea	More ...
P110AS	Infrared Remote Control Transmitter (Small Switches)	AU$27.50 Ea	AU$25.00 Ea	US$21.00 Ea	More ...
P110AL	Infrared Remote Control Transmitter (Large Switches)	AU$28.60 Ea	AU$26.00 Ea	US$22.00 Ea	More ...
P110B	Infrared Remote Control Receiver	AU$27.50 Ea	AU$25.00 Ea	US$21.00 Ea	More ...
Order Code	**Description - Miscellaneous**	**AUD (Aust)**	**AUD (RoW)**	**USD (RoW)**	**Details**
P+P	Pack and Post Charges (not including registration / insurance)	AU$ 5.50 / pkg	AU $12.00 / pkg	US $10.00 / pkg	
ES250	Rotary Switch Extension Shaft	AU$27.50 Ea	AU$25.00 Ea	US$21.00 Ea	More ...
M01	Measurement Microphone Hardware Kit	AU$85.00 Ea	AU$78.00 Ea	US$65.00 Ea	More ...
PCD	Documentation CD-ROM [2]	AU$ 7.70 Ea	AU$ 7.00 Ea	US$ 6.00 Ea	More ...
PINS	1mm Dia PCB Pins (Pack 50) [2]	AU$ 4.40 Ea	AU$ 4.00 Ea	US$ 3.50 Ea	More ...
KAP1	1 Metre x 25mm x 25um Kapton Tape [2]	AU$ 5.50 Ea	AU$ 5.00 Ea	US$ 4.50 Ea	More ...

Notes:

1. The price for the BP4078 has been reduced due to economy of scale at ColdAmp, making the BP4078 even more competitive
2. Available only with PCB purchase.
3. RoW = Rest of World (Anywhere outside Australia).
4. US$ Prices are for PayPal Purchases *only!*

Projects Index

Main Index

Copyright Notice. The information in this article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2002-2006. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only. Commercial use is prohibited without express written authorisation from Rod Elliott.

The Audio Pages

Elliott Sound Products **Purchase Project PCBs**

Copyright © 2000-2006 - Rod Elliott (ESP)
Last Updated 26 May 2006

- **Main Index**
- **Projects Index**
- **Access Secure Area** (User Name and Password are required)
- **Check the ESP Dead Letter Office** Moved since ordering? Look here now.

Fully Built Equipment

- **ESP VP103 Valve Preamplifier**; (updated 18 May 2001)
- **Built and Tested Modules** - Ready to use modules, fully built and tested

---

Note that Javascript *must* be enabled on your browser, or most of the links on this page will not work.
If your browser does not support Javascript, Right-Click the link, and select "Open in New Window".

The boards listed here represent very good value for money - all boards supplied have very comprehensive documentation available from ESP's secure section of the Web site. A CD-ROM is available for a small extra charge. The CD is provided free for orders over AU$75.00 - note that this specifically *excludes* postage and handling charges, registered mail (in Australia) or insured mail (rest of world).

Your product purchase supports The Audio Pages, and is the ***only*** way that this site can continue. Without your purchase, the articles and projects will cease to exist, as the cost of maintaining the site would be prohibitive without any way to pay for the site and its

maintenance.

My thanks to all those who have bought (or are planning to buy) project PCBs or other products. Without you, The Audio Pages would be no more. Your continued support is greatly appreciated.

Please read this page thoroughly, so that you understand the payment methods and warranty provisions, and any country specific information. To simplify the page, I have separated the major sections into Javascript "popups", making navigation a lot easier because of the large amount of information provided here.

Information and Board Index

General Information

- **Order Forms**
- ESP Loyalty Programme new
- Purchaser Agreement
- How To Purchase
- International Customers
- Currency Conversion
- Australian Customers
- Shipping and Documentation
- Order Codes
- Miscellaneous Information
- Warranty
- Returns
- Package Loss Policy
- Air Mail Postage Rates, Insurance, etc.

Modules (... or see Fully Built Modules above)

- BP4078 Class-D Amplifier Module
- M27A Guitar Power Amplifier
- DF100 Digital Flash Trigger

Printed Circuit Boards

- Complete Listing of all Products ... Javascript not needed

- Price List for all Products ... Javascript not needed
- Project 05 - Preamp Power Supply
- Project 06 - Hi-Fi Phono Preamp
- Project 09 - Linkwitz-Riley Crossover
- Project 19 - 50W IC Power Amp
- Project 27A - 100W Power Amplifier MkII
- Project 27B - 100W Guitar Preamplifier
- Project 33 - Loudspeaker Protection and Muting
- Project 36 - Death of Zen Class-A
- Project 37 - DoZ Preamp + Power Supply
- Project 39 - Soft-Start circuit for power transformers
- Project 3A - 60-100W Hi-Fi Power Amp
- Project 48 - Active Subwoofer Equaliser
- Project 66 - Low Noise Balanced Mic Preamp
- Project 68 - 300W Subwoofer Amplifier
- Project 70 - DoZ Headphone Amplifier
- Project 71 - Linkwitz Transform Circuit
- Project 72 - Stereo 20W IC Amplifier
- Project 73 - Hi-Fi PC Speaker Board Set - Save 10%
- Project 84 - 8 Band constant Q Subwoofer Equaliser
- Project 87A - Dual Balanced Receiver
- Project 87B - Dual Balanced Transmitter
- Project 88 - High Quality Audio Preamplifier
- Project 93 - Microphone Amplifier
- Project 94 - Universal Preamp & Mixer
- Project 96A - P48V Phantom Feed Power Supply
- Project 96B - P48V Microphone Interface
- Project 97 - Complete Hi-Fi Preamplifier
- Project 99 - 36dB/octave Subsonic Filter
- Project 101 - 200W MOSFET Power Amplifier
- Project 110 - Infrared Remote Control (Transmitter & Receiver)

Incredibly Useful Stuff (that isn't boards :-)

- Rotary Switch and Pot Extension Shaft
- Measurement Microphone Housing

- 25um Kapton Tape for Transistor Insulators
- 1mm Diameter PCB Pins
- Documentation CD

IMPORTANT: If you don't see the board you want here, then it is not available. This page describes all boards currently for sale, and if it is not listed, I don't have it. Also, do not ask for artwork, as refusal is **guaranteed**. In some cases I have been asked because "it is difficult to get boards shipped to [wherever]". While this may well be true, I *will not* supply my original artwork to anyone for any reason.

Boards are supplied as a blank PCB only - no components or other parts are included. Please do not ask about kits, as these are not available, nor are heatsinks, transformers or other components not shown in the pricelist.

ESP Loyalty Programme

ESP has just implemented a loyalty programme. This is effective as of 1st July, 2005, and several discount vouchers have already been delivered. Click the link above for more information about the programme, redeeming discount vouchers, and the terms and conditions for earning 'loyalty points'.

On-Line Construction Info & Terms and Conditions

When boards or other products are ordered, you will be sent a URL, Username and Password to access the board construction articles. The latest information is available on these pages, and you will be able to see the information for the boards you have purchased as well as the other constructional articles. Please click on the header to read the full details.

How to Purchase

There are several options for purchase, and all are quite straightforward. Please click on the section header for the full details. ESP accepts Visa, Mastercard, Bankcard (Australia only), PayPal, Paymate and direct bank deposits.

On-line purchase is not available at this time. When sending your card details, make sure that your details are entered on the order form or the Encrypted Form

Welcome Here

PayPal Info **Paymate Info**

International Customers

Use the GST free board prices listed - you will be billed in Australian Dollars, and the final price that appears on your card statement is determined by the exchange rate prevailing when the card details are processed. Click on the header for more information, currency conversion details, etc..

Australian Customers
For local customers, please note that the GST included price must be used for all orders. A personal or bank cheque is acceptable - but goods will not be shipped until the cheque is cleared. The preferred payment methods are Visa, Mastercard, Bankcard, Australian Post Office money order or Paymate.

Australian residents using PayPal must add GST to the $US price - multiply the price shown by 1.1

Please note that free postage for Australian orders no longer applies - see pricelist form for details

Shipping And Documentation

Boards are normally shipped in a padded postal bag via normal mail, and this is at the rate shown in the pricelist. Where possible, orders are sent as 'large letter' in a padded bag. Customs forms are not required by Australia Post for the padded bags, and are not used unless the package has to be sent as a 'parcel' (anything that is bigger than a large letter). Click on the header for more information.

Order Codes
The order code for each PCB or other product is shown in the price list. Where there is

more than one board for a project, they will have the suffix A, B, (etc.) - for example P27a and P27b (guitar amplifier boards). All boards are made with fibreglass PCB material. Some boards are sold in pairs, please note that paired boards will not be sold individually, nor in odd numbers.

Miscellaneous Info.
All board sizes are in millimetres and are approximate to the nearest millimetre. Board thickness is approx. 1.6mm unless otherwise stated.

Some boards have relatively small tracks and closely spaced components. These require a fine tipped soldering bit (preferably temperature controlled), and careful soldering techniques. If you are new to soldering components into a PCB, please exercise caution during construction, as it is possible to damage the board with excessive or prolonged heat.

Warranty

All goods are checked before shipping, but it is possible that a defective board or disk may be supplied. The boards are warranted to be free from defects that would make them unusable for the intended use. Such boards will be replaced free of charge or a refund given upon receipt of the defective board(s). Click on the header for more information.

Returns

ESP regrets that if you change your mind, returns are only available under some conditions. Full details are available by clicking on the header above.

Package Loss

Package losses are extremely rare, but for shipping to the some countries, insurance is highly recommended. Please see the pricelist for the most up-to-date information I have available. More information is available by clicking on the header.

Order Forms

Load the HTML form, save a copy of the text form, or download the order encryption program. Please fill in all sections. Print a copy of order form to mail or fax, or e-mail the

encrypted order.

Printed order forms can be sent by post or FAX to a secure machine. Please do not send cover sheets or other un-necessary sheets - all needed information is on the order form. The facsimile number is ...

In Australia	**02 9945 1385**
International	**+612 9945 1385**

US Customers please note that +612 is the country and area code, and is not a US domestic area code.

Simple Text Form **HTML Order Form** **Encrypted Order Program** *

Current Pricelist Version 1.21 - 02 Feb 2006

* See general notes for more information.

Projects Index

Main Index

Contact ESP

Total Visitors since Jan 2001 - 5322214

Copyright Notice. This document, including but not limited to all text and diagrams and the order form, is the intellectual property of Rod Elliott, and is Copyright © 2000 - 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright (c) 25 Jun 2000./ Last update 27 Jul 2005 - added loyalty programme

Elliott Sound Products **Application Notes**

Page Last Updated - 03 July 2005

These application notes are presented as a means of making useful circuits and sub-circuits available, along with some information about how they work, test results (where applicable), etc. Some of these are adapted from the projects page, as their application has the potential to be broader than indicated in the project itself. More will be added as time permits.

Few of these (already published or yet to be published) are original, although some have been adapted and changed to the extent that the original may barely be recognisable. As always, if you have a good circuit idea feel free to submit it (along with any reference material).

ESP Main Index Page

ESP Application Notes

No.	Description	Date	Flags
AN-001	Precision rectifiers. Half and full wave types for signal processing, instrumentation, etc.	02 Jun 05	
AN-002	Analogue metering amplifiers	15 Jun 05	
AN-003	Simple switchmode Supply for Luxeon Star LEDs	18 Jun 05	
AN-004	Car dome light extender - make the dome light stay on for a while after the car door is closed	18 Jun 05	
AN-005	Zero crossing detectors and comparators, Unsung heros of modern electronics design	20 Jun 05	

AN-006	Ultra Simple 5V Switchmode Regulator - voltage regulator version of AN-003	30 Jun 05	new
AN-007	High Power Zener Diode, boosting normal zeners to allow high power usage	30 Jun 05	new
AN-008	How to Use Zener Diodes, the things the data sheets do not always tell you	30 Jun 05	new
AN-009	Versatile DC motor speed controller (and where to get high power geared motors)	03 Jul 05	new
AN-009/2	DC motor speed controller (Part 2)	04 Jul 05	new

Manufacturer Application Notes *

No.	Description	Date	Flags
AN-166	Basic Feedback Theory (Philips Semiconductors)	Dec 1988	
AN-1000	Mounting Guidelines for the SUPER-220 - Transistor Mounting Techniques (IR)	Unknown	
AN-72	A Seven-Nanosecond Comparator for Single Supply Operation (Linear Technology)	May 1998	

Each application note may have one or more "flags". These indicate the status of the app. note, and are as follows ...

new	The design (or update) is less than 2 months (or thereabouts) old.
	Mains wiring is involved, and is potentially dangerous - heed all warnings ! Note that other app. notes may also need a power supply which also requires mains wiring.
Date	The page was added or updated on the date shown

* Manufacturer Application Notes will most commonly be links to external sites, although a small number may be included on the ESP site. These remain the intellectual property of the original author/manufacturer in all cases.

Updates

ESP reserves the right to change or update application notes, projects and articles

without notice, so it is important to be aware that a change may have been made. You should always watch for updates of previously published items. Do not build any of the circuits presented here without checking for updates first. A "new" symbol indicates an addition or update within the last two months or so.

Note Carefully

Please see the ESP disclaimer for important information about this site and the contents thereof.

WARNING

Where applicable, mains wiring should be carried out by suitably qualified persons only. Under no circumstances should any reader construct any mains operated equipment unless *absolutely* sure of his/her abilities in this area. The author takes no responsibility for any injury or death resulting from, whether directly or indirectly, the reader's inability to appreciate the hazards of household mains voltages, lack of knowledge of correct wiring practices, etc. Please read the disclaimer now if you have not done so already.

Please note that these application notes are not supported, so do *not* ask for assistance or explanations. All circuits are checked for accuracy, but it is not guaranteed that they will work for your application. In the same way that you might find application notes published by semiconductor (or other) manufacturers, you don't e-mail them for help and the same applies here.

Submissions are welcome, but unlike magazines where they may offer prizes or cash for submissions, all I can offer is wide distribution of your idea. Please ensure that any submission is accompanied by full disclosure of references - do not try to claim the work of others as your own.

Main Index

Total Visitors since Jan 2001 - 5322215

Copyright Notice. Application Notes referred to herein, including but not limited to all text and diagrams, are the intellectual property of Rod Elliott and the person or company whose work was referenced in the description, and are Copyright © 2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author / editor (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use in whole or in part is prohibited without express written authorisation from Rod Elliott and the owner of the copyright in the case of submitted articles.

Elliott Sound Products
PO Box 233,
Thornleigh NSW 2120 Australia
www.sound.au.com
rod@sound.au.com

Elliott Sound Products **Custom Design and Consulting**

ESP Services

Telecommunications

I will perform contract work for telecommunications systems, especially IVR (Interactive Voice Response) systems, and PABX (PBX) and Call Centre integration of same. This will be by direct negotiation in all cases, and as initial contact please send me an e-mail. The details (including confidentiality agreements and scale of charges) are negotiable, and will not be posted here.

The range of services provided is likewise subject to discussion and negotiation - please use the contact page to send me an e-mail with a brief description of your needs. All contact details (phone, mobile, etc.) will be provided by return e-mail.

Audio & Analogue

Should you have specific requirements for a customised design or something completely new, ESP may be able to help. I have developed many products over the years, both for individual customers and various employers, and during those periods when I was operating my own business (Elliott Sound Products).

Please Note - The availability of my time at the moment is not at all good, as I have been very busy. The chances of a complete design for a product are poor - by all means ask, but don't expect that I will be able to help. Projects for individuals or minor modifications to an existing published design are not possible at this time.

Some examples will show the range and diversity (customer names and specific details will not be disclosed under any circumstances).

- ***Audio circuits.*** (Naturally) some samples are to be seen in my Projects Pages.

These say far more than I can here, and as noted, most are original designs. This is actually a small sample, as there are still some very original ideas that I am unwilling to release to the public domain.

- ***Valve (Tube) Power Amplifier.*** This was one of my most satisfying designs, producing a product that achieved critical acclaim and was deemed by most to be one of (if not the) best sounding amp they (the critics) had ever heard.
- ***Animation Camera controller.*** This was a one-off development to repair an animation camera that was deemed a write off by everyone else. It included all the necessary motor controls for a standard 6 axis (zoom, top peg bar, bottom peg bar, east-west, north-south and rotate) camera, and included movement counters, limit switching and shutter controller.
- ***Printed Circuit Board Design.*** I have been doing this since the days of sticky pads and tapes, and have designed boards for all my own designs (including all the products described here).
- ***Telecommunications System Power Fail Transfer Unit.*** This was designed for my employer at the time, since a new version of the telecommunications system we sold did not have this feature. The unit was so successful that many years later when the system was replaced, it took the installers two hours to work out why the system was still taking calls - even though all the leads to the PABX were disconnected!
- ***Telecommunications Conference Bridge.*** Another special for an employer, this unit can set up automatic (emergency) conferences with up to 64 parties - all in full duplex conference, or operate in "dial up" mode, where up to 64 parties can call a designated number and be connected to the conference. A new feature (not available on any existing conferencing system) is the ability of any conference party to mute their channel (at the bridge, not their phone) if they are in a noisy area that would make the conference difficult for others. This unit has resulted in sales worth a great deal of revenue for my employer. This unit is completely analogue, and is far better than any digital system I have used.
- ***Multi Projector Controller.*** A controller for multiple slide projectors (used for advertising in shopping centres and the like), which controlled the timing for slide changes, controlled turn-on for the lamps to prolong their life, and jamming sensors to prevent the projectors from melting (and yes, this did happen with the original simple controller).
- ***Multi Sector Intruder Alarm System.*** This was (in its time) one of the most sophisticated alarm panels available, and had unique automatic arming and re-arming features that even now (with computer control) are not commonly available.
- ***Sound Recording Studios.*** I have designed two recording studios, one for myself and a business partner, and another on contract. Both were deemed very successful, and had excellent acoustics.

- ***Software.*** I have written various software programs, most for "fun", and one commercial product as well. Various of these are in daily use (world-wide), and the only comments I get are positive. I have also developed computer controlled telecommunications test sets and the software to drive them.
- ***Digital Circuits.*** While I am not a fan of digital audio, I have developed many digital control systems, and several of the products above are based on logic control.

As I said above, a small sample. Please contact me if you have anything that I might be able to help you with.

Terms & Conditions

As can be expected, these are negotiable on a case by case basis. I will generally be happy to operate on a Payment for Service basis, or a royalty based on the number of items sold and my contribution to the product.

I do not expect to have to use the services of lawyers for any contract, unless the value (to you or me) justifies this. A sample contract is available upon request.

Absolute confidentiality is guaranteed for all negotiations, designs and products.

Projects Index (Design Samples)

e-mail ESP

Main Index

Copyright Notice. This article, including but not limited to all text, logos and diagrams where applicable, is the intellectual property of Rod Elliott, and is Copyright (c) 1999, 2000, 2001, 2002. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page Created and Copyright (c) 04 Oct 1999.

The Audio Pages

Elliott Sound Products — Privacy and Ethics Statement

The Internet by its very nature is distributed and unregulated. While this is undeniably its greatest asset, it also eliminates many of the tangible cues to a merchant's credibility, which are normally quite obvious in the real world. In the hope of replacing these lost "consumer comfort factors", and preferably without the need for government intervention, Elliott Sound Products (ESP) voluntarily commits to follow a Universal Standard of Ethics when transacting business on the Internet.

As a reputable member of the global Internet commerce community, ESP hereby agrees to ...

Ethics

- Deliver what we promise and promise only what we can deliver.
- Deal honestly and fairly with customers.
- Accurately represent who we are and what we do.
- Fulfill any warranty claim within a reasonable period of time.

Credit Cards *

- Charge only the amount required for the goods ordered and appropriate postage, in accordance with the published prices.
- Never apply any undisclosed charge, repeat charge, or other charge unless specifically requested by the purchaser to do so (for example, add additional items to an order).
- Not apply any surcharge for use of a credit card.

***** ESP reserves the right to correct errors in orders, where the total cost has been incorrectly calculated by the purchaser, the incorrect pricelist was used, or other errors as may occur from time to time. You will *always*be advised of any changes and the reasons for those changes.

Complaints

- Promptly respond to all customer complaints
- Employ our best efforts to fairly resolve all legitimate complaints in a timely fashion

Advertising

- Ensure to the best of our ability that no documents containing false statements are distributed on-line

- Promptly substantiate any statements or advertised claims upon request
- Not undertake any unsolicited electronic advertising or bulk e-mailing of any kind

Disclosure

- Disclose basic features and benefits of products in language that is easily understood by the consumer - bearing in mind that such descriptions are commonly of a technical nature because of the products I sell
- Disclose all terms, conditions, financial and contractual obligations applicable to the transaction including warranties, returns and refunds
- Disclose the price, type of currency and all fees, surcharges, and shipping costs associated with any transaction
- Disclose availability of goods and services, timing of billing, and realistic shipping expectations

Adherence to Laws & Regulations

- Abide by all laws of the Australian Federal and State Governments regarding credit card transactions
- Abide by the rules and regulations pertaining to the Merchant Credit Card Facilities provided by the Commonwealth Bank of Australia or other financial institution as may be used in the future.

Your Privacy

This is one thing that terrifies many potential Internet purchasers, since they are never sure what the merchant will do with their personal information. We know that you are concerned about your privacy and so is ESP. If you provide information about yourself or your business, such as name, postal or e-mail address or other data, this information will be added to our internal data records. Any information received from you (either by mail, fax or in any electronic format) is used internally by ESP alone, and is not shared with *anyone*.

Privacy

- ESP's privacy policy is absolute. Your name, details of purchases and address are absolutely confidential, and will not be disclosed to anyone who does not need this information other than for the authorisation of a credit card purchase, unless I am lawfully obliged to provide such details to a government agency or other statutory body.
- Your credit card details shall not be stored or archived on any computer (unless fully encrypted), and will be kept safely and securely only in the original purchase form and credit card voucher. All card details are destroyed by shredding after the statutory holding period has expired (as established by my bank's policies).
- Used facsimile (fax) developer rolls are cut into small sections and heated to melt the 'ink' before disposal (these otherwise contain a perfect negative image of all pages printed by the fax machine).
- ESP will not send unsolicited e-mail, nor provide your e-mail address to any other party for any reason whatsoever.
- Where your information is supplied in relation to an order for purchases, the information will be used solely for that purpose. ESP will not share personally identifiable information with any third party without your express consent.

On the subject of privacy - because ESP is operated from my home address, I do not advertise or publish the street address or telephone number of the business. This is sometimes seen as evidence of a business that has something to hide - this is not my intention - I am entitled to some privacy as well, and that is the sole reason for this omission. If you have a reason for needing this information, all you have to do is ask.

So far, I have not had one complaint from a customer, nor from my credit card handling facility, PayPal or Paymate. I am determined to keep my reputation - honesty and integrity are the only tools at my disposal to achieve this.

Page created and Copyright © 02 Feb 2002

Elliott Sound Products — Spam, Scams & Security

Copyright © 2003 - Rod Elliott (ESP)
Page Updated 21 Oct 2003

Index		Update
Spam	There's a lot more to this vile abuse than first meets the eye	Apr 03
Scams	Some of these are mind-numbingly dumb, but people still get caught out	07 Mar
Security	Is yours at risk? Some things for you to consider (and one you would *never* have guessed!)	Sep 03
Main Index		

Page created 25 Apr 2003

Elliott Sound Products **Editorial - I Am As Mad As Hell *!***

Last Updated 25 March 2004

Keep watching this space for more examples of Hi-Fraud

Contents

- The Lead-Free Solder Directive
- My Own Little Gripe
- The Triphazer - are people being ripped off (again)?
- Death of an Electronics Magazine
- Counterfeit Transistors ***(Moved to a separate page - 03 Apr 2002)***
- Burn-in - Myth or Magic??
- A Funny Cable Story
- German Hi-Fi Repair (New York, USA) (From a reader) - Archived
- The Final Insult (to our intelligence and everything else !) - Archived
- Gaincard, Power Humpty (I'm not kidding) and other Fairy Tales - Archived
- Magic Lacquer - (or "I Simply Don't Believe This") - Archived
- Mains Lead Madness - Archived
- Gryphon Exorcist (aka Criminal Extortion) - Archived

The Lead-Free Solder Directive

The EEC's RoHS (Reduction of Hazardous Substances) directive is due to become mandatory in 2006, and while the overall idea seems reasonable, it is flawed in the extreme. There is little or no evidence to support the claims that lead is 'leached' out of solder by acidic groundwater, and since lead is almost impervious to most acid and alkaline attacks, there is no reason to believe that there is any real hazard. Indeed, high temperature tin-lead solders (those with more than 85% lead) are *exempt* from the directive. So highly leaded solder is less of a problem than the 40% lead solder normally

used? I don't think so!

The costs to manufacturers will be very high, with higher energy usage needed to achieve the higher temperatures needed. Higher temperatures mean greater thermal stress on PCBs and components, potentially reducing their life expectancy. There is a great deal of concern over the RoHS directive worldwide, since any product made after July 2006 cannot be sold in the European Community if it uses lead based solder.

The alternatives are many - so many in fact that all they have achieved is confusion. Some patented lead-free solder formulations are totally incompatible with lead based (and some other lead-free) solder alloys, and a small mistake by a service technician could render a part or an entire PCB assembly unusable because of unreliable solder joints caused by incompatible alloys.

Bismuth in particular is a problem if any lead based solder is inadvertently used, and lead contamination of a tin-bismuth-silver based solder results in joints with very poor thermal properties, and a high probability of faulty joints appearing shortly after repair work. Such lead contamination can result from the use of PCBs or components with lead based solder 'tinned' leads or tracks, or from the application of lead based solder during repairs. If accidentally contaminated, there is a likelihood that an entire PCB may have to be scrapped unless there is a method to de-contaminate the board - as far as I know, no such method exists.

In addition to the problems of the extra heat needed (most lead-free solders require at least 217°C soldering temperature), lead-free solder has poor wetting properties so PCBs and leads must be completely clean, and the solder joints do not have the same appearance as those made with conventional solder, making visual inspection that much harder. Fluxes are not as effective (and burn easily) at the higher temperatures.

A web search for 'lead free solder directive' gives over 6,800 results on Google, so there is plenty of information to allow you to make up your own mind. Although some Japanese manufacturers have already made the switch, there is nothing so far that shows any real advantages, but disadvantages abound.

There is a consensus that the RoHS directive will achieve little that is genuinely useful, and IMO it will create so many problems that it should be rescinded. It is interesting to note that for high reliability applications (computer servers, telecommunications, military, etc) the directive does not apply, or will be postponed for several years, and this in itself is very telling - they *know* that there will be reliability problems with components and the solder joints themselves, and have provided exemptions for those areas where reliability

will cause major problems for industry or military applications. The EEC obviously does not care that consumer goods will be less reliable and cost more, and has shown yet again that there is (or seems to be) almost a conspiracy to either prevent or make it as difficult as possible for 'outsiders' from competing in the European market.

What of the poor hobbyist? Those who are just learning soldering skills will damage PCBs and components, as will many professionals who are unused to the high temperature and comparatively poor looking solder joints. Having used lead free solder for some basic testing, I can confirm that it is difficult to use by comparison, and produces a joint that looks bad. It seems to be strong enough, but I know of no-one who actually likes it.

The 'benefit' of the RoHS Directive is that 0.6% of the world's lead production will be removed from solder, where it appears to do no harm to anyone. For some bizarre reason, the EEC would rather use bismuth (which has not been banned), but the fumes are thought to be toxic (see Material Safety Data Sheet for more information). The effects of contamination are unknown, but it is certainly more likely for bismuth to be leached from solder than lead.

A good idea? I think not.

My Own Little Gripe

There is a natural flow-on effect from this into some (or all) of the other topics I have covered, and I hope that at least a few readers will understand that these editorials are as much an attempt to bring some sanity back into audio as they are an attack on dishonest practices.

People are very often far too trusting of those who will take every advantage, but will brand as a sceptic or fool those who will stand up and shout "*Bullshit*". I know this from past experience with some of my editorials, and I am somewhat saddened that I am accused of being closed-minded or a simple techno-freak who utterly fails to understand the "finer points".

How is it possible that those who are selling something (that is usually outrageously expensive for what it is) are hailed as heroes who have saved the audiophiles from mediocre sound, while those with no hidden agenda, are not selling anything that is technically dubious, and have been dedicated to audio excellence all their lives are castigated, flamed, and accused of (somehow) trying to spoil the enjoyment of the consumer?

This makes no sense to me, and I know I am not alone. Rationality is neither a crime nor a sin, but some of the vested interests would have you believe that it is both. Don't accept it.

Interestingly, the claims against my stance are almost always made in a public forum, and often by the vested interests I am standing against, although regularly "disguised" in some way, typically by not noting their affiliations in their posts. Very few (and I really do mean very few!) people who disagree have sent an e-mail or attempted to contact me directly in any way. There have been exceptions, and these are often reasonable people who have useful information that may well make a difference. I have modified several articles based on proper (substantiated) information that has been supplied over the period this site has been in operation, and will continue to do so whenever such info is made available to me - I cannot test and verify every claim made, so rely on others to provide the details that make my evaluations (and comments, for and against) complete, and as factual as possible.

The Web is littered with outlandish claims, grossly overpriced "accessories", and completely unsubstantiated claims for "products" that you "cannot live without". These are nearly always identifiable by the complete lack of technical information, or pseudo-scientific technobabble, designed specifically to confuse the reader. If you are unsure - try a Web search on the topic to see if there is any technical explanation for the claims made.

The above does not mean that all such products are horse feathers - some may have a genuine value, but beware of the panacea - very few products will work with any system without proper integration. A rock placed on top of an amp or speaker is unlikely to reduce distortion and colouration, and it cannot compensate for room acoustics ... but rocks are cheap, so you can experiment to your heart's content

If it costs a lot of money, and will cure everything from ingrown toenails to poor imaging, then beware - you are about to lose your cash. Eventually you might convince yourself that it has made a difference (no-one likes to be taken for a ride), but careful analysis (with and without the "gizmo") is needed before you can be certain.

Triphaser

Are people being ripped off (again)? The answer appears to be yes ...

There is a device called the Triphazer, that according to its manufacturer will "transform"

your system. Allegedly, it works just as well on top of the line audiophile hi-fi equipment as on a $99 boom box. A truly remarkable device, methinks. I hope the designer is more capable with electronics than writing English - the punctuation is pitiful (I saw one reference that suggested that they should have their apostrophe license revoked :-)

A few quotes (indented text - all are verbatim, including all punctuation and spelling errors, etc.) from the makers of the "Triphazer" (aka Triphasor, Triphaser), followed by my comments.

> Using proprietary circuitry they [Triphazers] improve resolution and lower distortion
>
> All components and cables, have inherent nonlinearities. Amplifiers, preamps in the record and playback chain all contribute to very small distortions that are sometimes called SKEW or SMEAR. Because all designs are limited by this integrated SKEW, Triphazers have been designed to reverse these non - linearity's,
>
> We can repair this skew back into the original shape or condition that it was in when it left the studio mic or concert hall microphones.

For a start, cables do not have non-linearities, other than in frequency response. This is generally only apparent with difficult loudspeaker loads, where there are significant impedance variations, especially at higher frequencies. Response anomalies are very rare in interconnects, at least with normal lengths and within the range from DC to 100kHz. No-one has ever been able to measure distortion introduced by any cable.

No device can remove non-linear distortion of any kind, unless it has the original signal as a reference - this is how negative feedback works, it uses the amplifier to detect differences between the input and output, and make appropriate corrections. The Triphazer does not have this ability, so the claim is blatantly false.

Try this one for a laugh (from the FAQ page) ...

> What will it do for sound reinforcement (public address, rock bands)?
>
> They do a lot! Playing substantially louder and the clip lights never come on! Musicians comment they can hear each other better and request a lower monitor levels, fatigue is reduced tightness improves because the members can hear nuances ..that musicians need to work closely together. Triphazers remove large amounts of distortion that an audio system would normally reproduce and have to be dissipated by the speaker drivers," causing over heating in the voice coil , listening fatigue and confusion. It is this "Removing of Artifacts" that result's in increased efficiency and leaves more energy available for the fundamental musical notes that were at the core of the musical signal before being altered during the record, production and playback process.

This passive device will make a PA system louder, with less power, as well as reduce speaker voice coil dissipation (and again remove distortion!). If this were true, then everyone (and I mean everyone) on the planet who has ever used a PA system would have one. Last I saw, there are but a few references on the Web, not the thousands one would expect. I did find some references on a pro audio forum - all were scathing (to put it mildly).

Another quote ...

> Used in PROFESSIONAL SOUND REINFORCEMENT placing TRIPHAZERS after the mixer provides all inputs with improved clarity and freedom from Feedback! Freedom From Feed back! YOU HEARD ME! FREEDOM from FEEDBACK!

Oh really! Acoustical feedback is always a problem, and there are many solutions that provide some relief. Parametric equalisers, frequency shifters (typically about 3Hz) or a combination will suppress feedback, but according to the above claim, the Triphaser ***eliminates*** it. Bullshit!

Acoustic feedback is the result of sound from the loudspeakers reaching the microphone and being re-amplified. A passive box cannot do this without either attenuating the signal (turning down the volume is a lot cheaper), or introducing one or more notches in the response at exactly the right frequencies (as one would with an equaliser). That it eliminates feedback entirely ("freedom from feedback" implies complete elimination) is again blatantly false. Presumably it takes at least 50 hours to do so, according to the next piece of "wisdom" ...

> What about break-in? What is the deal?
>
> That is a complex question... lucky for you we have the answers that can help you get the most out of your stereo, home theater, PA, or studio. We recommend 50 to 200 hours for preliminary break-in. But we have other technology's that can speed up the breakin time. We will be writing much more on the subject to set the record straight and to show our customers how to obtain the maximum from recording and audio system investments.

Hmmm. The concert will be over by the time the feedback is eliminated, which makes sense, because the PA system will be turned off by then :-) I have since seen claims that your system will sound better, then worse, then much better - sounds just like the claim for the "Magic Lacquer". In the time allowed, your ears and brain will get used to the sound - the sound will not change. No passive circuitry requires break-in, it is completely in the mind of the beholder (belistener ??).

Conclusions

The Triphazer is now patented, and rather than describe it here, I suggest that you look at US Patents and Trademark Office, and search on Patent No 6,486,750. In short, the device **cannot** do what is claimed for it, and is a pure placebo (albeit an expensive one). The claims made in the patent are laughable, and no patent should have been granted IMO.

The Triphazer is not even a Zobel network. It is essentially several pieces of wire of different lengths encapsulated in epoxy, and presumably the inventor has worked out a method to make 3 parallel signal paths act as 3 separate circuits for different frequencies. Unfortunately, none of this is funny, and while it is possible that the inventor really does think it works, he is very much mistaken. If you were to connect anything in parallel with a short piece of wire (for example), the piece of wire will be dominant to the point where the additional circuitry will make zero audible or measurable difference to the signal, regardless of frequency. Different frequencies within the audio range do not 'choose' to travel by different paths - not even if you ask them nicely.

Tritium responded to my request for some technical details by sending a collection of "unsolicited" e-mails saying how good it was. I specifically asked for information, not rubbish. Other responses I have had point to bad reviews, general criticism, and everyone who had tried to get any technical detail has been fobbed off with the same e-mail collection as I received, and no real info at all. I would have serious doubts about having any dealings at all with this company - anyone who has a real product has nothing to hide, and this lot seem to have everything to hide (and no real product).

As of 24 March 2004, I have not had *one* additional response from the manufacturer, and no-one has presented any information whatsoever - for or against - other than the section below, which is actually funnier that the Triphazer nonsense.

---

05 Feb 2002

I did get some information from a reader (a Patent Attorney - amongst other things - as it transpires). Naturally enough, I shall not name the source, but he had this to say (reproduced by permission of the author) ...

> Ahhh, yes, patent procedure. You see, patent applications are generally not published (i.e., available through the databases) until 18 months after the filing date, a practice that has only recently been adopted by the US (it used to be that *nothing* was published until the actual grant of a patent in the US, often many *years* after the filing date - a strange practice much deplored by the rest of the world's patent practitioners). So it is no surprise that your search did not yield the application in question.

> That said, one must understand that the patent system (in the US or anywhere else!) does not require that the invention actually works as claimed! The statutes merely require that the invention is novel, is not obvious to a person skilled in the art, and is industrially applicable (i.e., can be manufactured and sold). The inventor has no burden of proof that his invention has any real benefit.
>
> Sadly, many patents are granted for devices or methods that are absolutely ludicrous, especially in the US. If you want to have a good laugh, look up US Patent 6,025,810, granted(!) for a "Hyper-Light-Speed Antenna". Some quotes: "The present invention has discovered the apparent existence of a new dimension capable of acting as a medium for RF signals ... ***The present invention takes a transmission of energy, and instead of sending it through normal time and space, it pokes a small hole into another dimension, thus, sending the energy through a place which allows transmission of energy to exceed the speed of light ... It has been observed by the inventor and witnesses that accelerated plant growth can occur using the present invention***." One can only speculate what type of plants the inventor was smoking at the time. But yes, this idiocy was worthy of a US patent!! Other examples are legion.

The above is verbatim, and has not been altered in any way (other than the application of bold and italics for emphasis). Methinks the Triphazer is no less silly than the "hyper light-speed antenna", and only misses out on this mysterious "other" dimension ... I suspect that this is planned for the Triphazer Mk II

Please feel free to roll on the floor with laughter at any of the claims made about the Triphazer and the antenna alike, but note that ESP accepts no responsibility for any injury sustained as a consequence of such mirth. If you feel that it is all too much, please stop reading this material immediately, and seek help from a medical professional (either that, or take two aspirin and if symptoms persist, e-mail me in the morning).

You would be well advised to bear the implications of the above in mind when ***any*** manufacturer claims that a patent is pending or has been granted for the "offering" in question. Just because they have a patent does not mean that the product or invention has any merit whatsoever ... "pokes a small hole into another dimension" - I mean, really this is just toooo much - ROFLOL

Dead EAT

Well, the inevitable has happened, and EAT is no more. The magazine is officially bankrupt, and all copyright material has been purchased by the remaining Australian electronics magazine, Silicon Chip. This is a sad time for those who supported

Electronics Australia for so long, but a richly deserved reward for those who caused its demise. It is to be hoped that they never re-enter the world of electronics publishing - may they sweep streets for eternity.

Australia has (had!) the distinction of having what I believe was the second longest running electronics magazine in the world. The magazine began life as Wireless Weekly, and subsequently went through several name changes. These included Radio, Television and Hobbies, then Electronics Australia. The magazine has ceased to be, and the remainder of this section is now redundant and has been removed.

Counterfeit Transistors (and more)

Please note - this section has been moved to a separate page. Click here for the article.

Burn-In - Myth or Magic?

I have managed to bite my tongue on this topic for quite a while, but a passion for common sense finally overcame me, and I had to write something.

First, A Contrary View
After many exhaustive hours of listening tests, I have determined that when a cable is burned-in, it is actually ruined. No cable should be used for more than a few hours, as the stresses on the insulation and the agitation of the copper molecules cause permanent changes to the structure of the cable - these changes are invariably for the worse, and fresh unused cables can be proven by listening tests to be superior in all respects.

The characteristics of the insulation change very subtly as the cable is stressed by signal voltages, and this has an as yet unexplained effect on the stereo imaging, and in particular causes veiling of the high frequencies and a loss of presence in the upper midrange. In extreme cases, the authority of the bass also suffers, with the lower registers lacking speed and power.

All the above defects are rectified by substitution of a new set of cables - the brilliance is restored and the finer details are brought back into startling realism. Bass speed is improved tenfold by a brand new unused mains cable, and new interconnects have a profound effect on the upper frequencies where detail is paramount.

I could say all these things (I just did :-) and I would be lying through my teeth. I could easily expound on this theory, lodge a few posts at some of the audiophile "watering holes", and offer a range of affordable leads in 10-packs (enough to last for a month or more of normal listening) so that fresh leads can be used when the others start to sound "tired". If I did my homework and used all the right words, how many sets of leads do you think I would sell? I already know the answer - "A lot". I am appalled at the rubbish that is fed to the audio fraternity by charlatans and frauds.

Sure, I could do it too (I would certainly make a lot more money than I do from the stance I take at the moment), but I would not enjoy it in the least. The reason I would not enjoy it is simply that it would be dishonest and fraudulent - no-one would actually be able to prove me wrong (as is the current situation), but I would be unable to back up my claims with facts that were in any way meaningful. I would simply throw in a few mathematical terms, a bit of random molecular theory, or perhaps make dire mutterings about resonances and how they become more significant as the cable ages. From reading the postulations of the opposing viewpoint I know that a significant number of people would believe me, and why not? A significant number of people already believe that burn-in makes a good difference, so why would anyone who was disillusioned by a set of recalcitrant cables be unwilling to accept that my idea was either wrong or any more ridiculous than the others "facts" that are out there?

My credibility would suffer badly from such an exercise, and I would do a great deal of harm to the hobby that I (and so many others) enjoy. It is most regrettable (IMHO) that some of the others don't have the same attitude, and are just out to make a quick buck.

We are not talking about facts when the discussion turns to cable burn-in, we are simply digressing into the realms of magic - anything that cannot be explained by the "facts" as they are currently known to exist *is* magic. The concept of a vast machine that can fly through the air is pure magic to a primitive people who have never seen such a thing before. The concept that cables sound better after burn-in (or is that when brand new?) is magic to those who have little knowledge of electrical principles, but love music.

When there is a gaggle of reviewers and manufacturers out there telling them that "something" is so, why would they disbelieve these self proclaimed experts? Who else is there to turn to for help? When the uninitiated think they can't hear a difference, what are they to do? To admit that they hear no difference is liable to have them "cast out" by their peers for the heinous crime of being cloth-eared. A sorry state of affairs indeed, and one that ensures that the unscrupulous will not only survive but prosper.

We keep hearing that cables (some will say *all* audio equipment) should be subjected to various techniques to "stabilise" them. This is generally referred to as burn-in, and after the treatment the item(s) supposedly sound better. To aid this process - of course - many entrepreneurs have slaved away for whole minutes to create CDs with pink noise or some other signal "specially designed" to do the job properly.

So far, I have not seen a shred of evidence that any so-called treatment has any effect whatsoever, other than a psycho-acoustical phenomenon known as "getting used to the sound". This indicates that it is the owner's ears that get burned in, and has nothing to do with the cables.

OK, so I am claiming that there is no change in the cable. I have measured cables (as have many others before me), and normally expect to find three main characteristics and two that are not relevant to audio. These are (respectively) ....

- Resistance - influenced by the length and diameter of the conductors, and to a very much smaller degree by the purity of the copper used
- Capacitance - influenced by the distance between the conductors and the insulation material. The capacitance is also proportional to the length of the cable.
- Inductance - influenced by the cable length, diameter, spacing, and the amount of twist between the conductors
- Self resonance - in any cable suitable for audio this is insignificant, as it is (or should be) so far out from the audio spectrum that it will have no effect whatsoever
- Impedance - all cable has a characteristic impedance, and like self resonance it is meaningless for audio unless interconnects or speaker leads are many kilometres in length - this is unusual.

To some degree, the above comments are tempered a little when radio frequency interference (RFI) is present, but it will ultimately be the way the cable is terminated that makes a difference (rather than the cable itself).

It must be understood from the outset that cables are not very smart. In fact, they are bereft of any knowledge of anything. Indeed, their own existence is unknown to them, and their memory is much shorter even than that of a goldfish. This rather generalised statement applies to the conducting and non-conducting (insulating) materials alike.

A cable has no interest in the current flowing in it (or not) unless it is greater than the current carrying capacity of the conductors, in which case it will get hot (or perhaps only warm). This increases the resistance, but only for as long as the overload lasts, and until

the cable returns to ambient temperature. This will take a few minutes at the most.

As long as the temperature is kept well below the melting point of the insulation (or the copper), no permanent change occurs. This is an extreme example, since in practice most cables are at room temperature, and may gain but a fraction of a degree even at maximum amplifier power. Any current that may have flowed at some time is instantly forgotten.

Likewise, the insulation is not the least interested in the voltage that may have existed between the conductors once it has gone away - again making a valid assumption that the output voltage from the amplifier will not cause the insulation to break down, allowing the signal to arc between the conductors. There are some very minor effects with all insulators (dielectrics), where a short memory effect can be noted, but this is not at all significant for audio, and even less so in the long term.

The end result of this is that cable burn-in is an invalid concept. More than just invalid, it is an attempt to convince you (the buyer) that the reason the expensive cable(s) you just bought don't make any appreciable difference, is that they haven't been given the necessary treatment, so you should buy this CD (or some other overpriced piece of equipment) to rectify the situation.

The simple fact of the matter is that changes in room temperature will cause a far greater variation in the characteristics of a cable than pink noise applied for a minimum of 37.5 hours. At the end of the "treatment" the cable will still exhibit exactly the same resistance, capacitance and inductance as before - so what has changed? And the answer is .... nothing.

There are electrical principles that exist despite any marketing hype. The hype and bullshit does not affect these principles in the least, and there is absolutely nothing you can do to a cable with a normal signal that it will remember or that change its long term characteristics.

What about the other products?

There is no doubt that loudspeakers require some amount of use before they settle down. This is because new speakers will have a stiffer than normal suspension, and it takes some time for this to stabilise. There is no correct time, and no signal that is better than another (as long as there is roughly equal energy in each octave band). Each speaker will be different, and some will have been subjected to test waveforms and will already be optimum (or close to it) when they are purchased.

Over time, the suspensions will "relax" more and more, and will eventually fail - hopefully only after many satisfying years. The sound will change very slowly during the lifetime of the speaker, but largely goes un-noticed because we get used to the subtle changes as we age as well.

Transistor (bipolar or MOSFET) amplifiers generally require a minimal "burn-in" period. This is not really burn-in at all, but a period of time to ensure that polarised capacitors (i. e. electrolytics) have achieved their normal operating state. This varies from one amp to another (even from the same manufacturing run), and as electrolytic capacitors settle down, their leakage falls and capacitance often increases slightly. This process is actually repeated each and every time the amp has been turned off for any period of time.

I do not advocate leaving amplifiers on permanently, as this is a waste of power, and the above effects should not be noticeable in a well designed amplifier.

Valve amplifiers are another matter. Valves change their characteristics quite dramatically in the first few hours of operation. They stabilise after this, and then go into a very gradual decline as they age. Eventually (this can range from weeks to years, depending on how hard the valves are driven), the decline becomes much greater and the valve becomes unusable.

With both speakers and valves, the ageing curve is similar to the discharge curve of a cell or battery. A simple graph as shown in Figure 1 is typical of all three devices - the vertical axis is either stiffness (for a loudspeaker), emission (for a valve) or voltage for a battery.

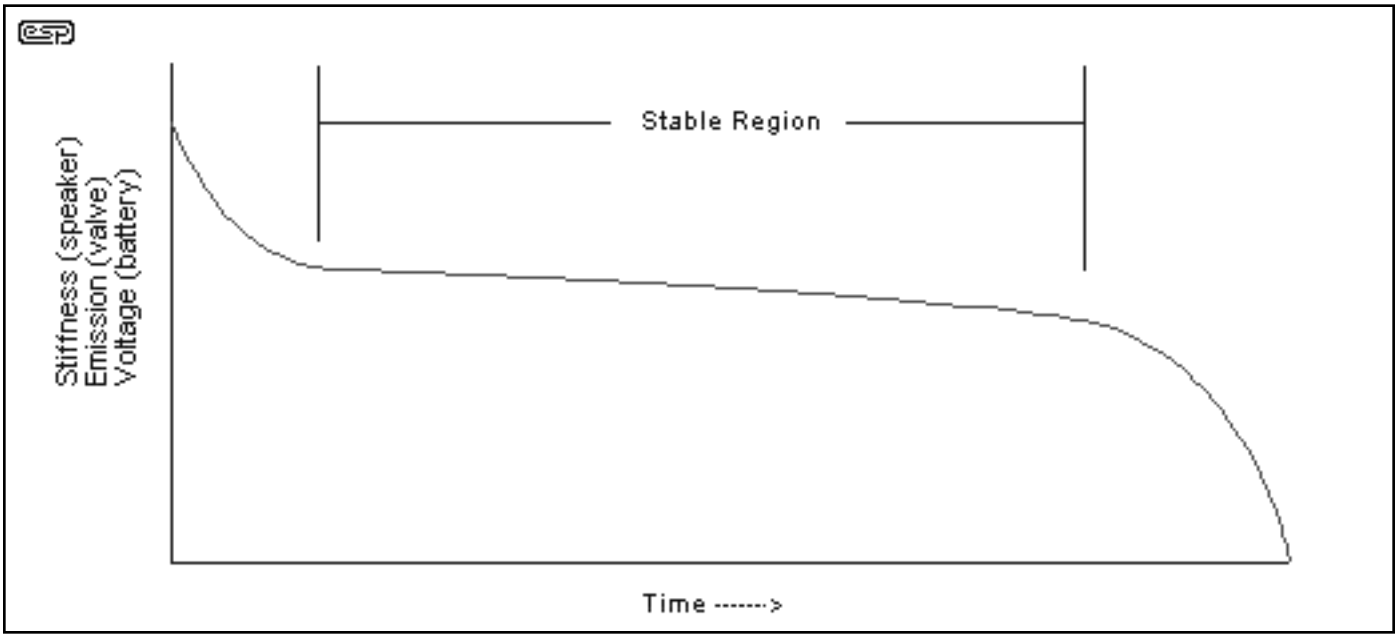


Figure 1 - A Typical Ageing Curve

I really hope that you find this information useful. Unlike some of my other editorials I have not flamed anyone (although there will be some who may claim they have been flamed by inference), but have simply stated a few facts of life. I fear that a few will take this entirely the wrong way and I cannot change this - another fact of life. I fully agree that there are some things that cannot be explained at this time, but cable burn-in is not one of them. There is simply nothing to explain.

Should anyone have some data they would like to share showing a measurable change in the basic characteristics of a cable from its original measurements after a burn-in period, I will be happy to include this information here. My stance on this is that I have not experienced any changes that I could detect - perhaps someone else has (as doubtful as this may be).

Cheers,
Rod Elliott

A Funny Cable Story (And It's True)

This story comes from a reader in the UK and is reproduced verbatim - only the name of his mate has been changed ...

I know that most of what's talked about cables is crap, but how's about this for a story -

A mate of mine decided his system was sounding crap. Since I'm considered (by others,

not me) to be the "authority" on all things hi-fi in my neck of the woods ('cos I own a soldering iron I guess) he wobbled his way towards me one Saturday night in the pub and moaned about how some sales rep in a hi-fi shop in the city had told him that he needed to spend about two hundred quid (GBP) on cables to get the sound he wanted.

To prove this, he trucked all his gear to the shop, the little twerp installed the new cables for him, and Metallica leapt out of the speaker boxes and assaulted my mate about the lugholes. During the course of the conversation, it came out that the twerp had spent ages "pushing" the new cables onto the connectors. Pushing? Never heard of an RCA connector THAT tight. I had a look the next day and what he'd actually done was spend ages twisting the new RCA plugs soldered to the "audiophile" connectors onto heavily-oxidised chassis mounted jacks to get rid of all the crud. We dug out my mate's old cables, stripped all his gear, cleaned the connectors properly and plugged everything back in. Result? His old home made connectors now sound better than the shite sold to him by the twerp.

Interestingly, he's a big lad is my mate Fred, and he's not a fan of being ripped of at all. Last I heard he was taking a day off humping billets of steel round an engineering works to go and see the twerp and have a few words ....

The moral of the story? As always, charlatans beware! Especially when the guy you've just ripped off weighs about eighteen stone (252 pounds or 115kg), and has mates.

(Some time later from a following e-mail) After a bit of foot shuffling and looking at the floor, Fred got his brass back. I suggested that he pushed for recompense for the "damaged" phono connectors on the back of his kit as well. He didn't. I guess he's not an unkind man at heart. Still, when you're that size it's hard to tell - nobody seems inclined to test the theory.

On a serious note, it's just another prime example of the unscrupulous taking advantage of a guy with a bit of brass who wants something decent but doesn't really know what he's talking about. To add insult to injury, the pleb tried to convince him that he had to shove hard as the new connector was stiff. You know what? When the CD player and the connector were both placed on the counter in front of him and he was asked to demonstrate, he had no trouble sliding one smoothly onto the other. Tch! These "audiophile" cables. Loosen up in no time don't they ?

Now he's taking me out for a few beers to celebrate getting his money back and having a great-sounding system for free. Poor lad! He's obviously forgotten about my strange but well-known gastro-financial condition - my capacity for ale and takeaways

spontaneously quadruples when someone else is paying.

My thanks to the reader who submitted this tale. Has anyone else had a similar experience?

A few explanations are in order here, as the story is from the UK, and some of the terms may be unknown elsewhere (in alphabetical order) ....

beer - the stuff most commonly quaffed in pubs :-D
brass - money
GBP - Pounds (as in the currency of Great Britain, not to be confused with "Fred's" weight)
lughole - ear
pleb - plebian, an ordinary person, (also a vulgar person, a boor - probably the intended meaning here)
pub - a place for drinking alcoholic beverages, and generally having a riotous time (or not, as the mood takes one).
shite - I think you can work this one out for yourselves (hint - leave off the 'e')
twerp - a despicable person / a cad (also, a bit of an idiot, a twit)

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000 except where noted below. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and Copyright © 14 June 2000 Rod Elliott. Updated Dec 2001 - added triphazer, and additional material, archived GHF, and moved "gripe" column./ 17/29 Jun - clarified a few points, and added some new info.

Elliott Sound Products **Contributions To The Audio Pages**

How to (and why) Contribute

The how to contribute is fairly easy. Send me an e-mail describing the article or project (no attachments please). Provided it fulfils the basic requirements (see below), I will request the full text and any diagrams to be sent as an e-mail attachment.

I will then edit your descriptive text (if and where necessary), insert your diagrams in the places you indicate and advise you of the URL so you can see the finished product before it is indexed. Once we are both happy with the result, I will index the material and your masterpiece is on line.

The why is also simple. I have a page that gets over 2000 visitors a day. This means great exposure for you and your ideas.

Terms and Conditions

These terms and conditions must be read before you fill in the Submission Request document, which is printed from your browser, signed and sent to ESP. All relevant details are on the form.

NOTE: The Submission Request is not mandatory, so you don't have to fill it in if you don't want to. I suggest that you do (and keep a copy for yourself), as this provides you with some degree of copyright protection. Obviously, if someone wants to steal your idea there is not a lot you can do to stop them, but at least you have proof of ownership of the idea, and the date it was submitted. I will also be able to support you if necessary if the form is completed.

The terms and conditions below are not intended to scare you off! The Audio Pages have gained considerable popularity because of the content and value to the readers. Any material that you contribute should have the same intent - to share your ideas, to educate the newcomer and to provide a benefit to the audio community. This is our hobby, our enjoyment (even our passion), so let others benefit from your experiences and help me to establish the best DIY audio web site in the world.

The basis of all of the stuff below is very simple. You don't offer material that is not yours, I don't steal your ideas, and you don't steal mine. We work together to everyone's advantage. I admit that some of the stuff below looks heavy going, but it is actually quite straightforward and in simple terms means that I will not accept responsibility if you try to pass someone else's work off as your own, and get caught out.

Terminology

So you know exactly what I mean by some of the terms, here are my definitions:

Submitted material	this includes any printed material or other hard copy, all software, text files or drawings in any format.
Published Material	includes any or all of the submitted material, and also includes the page format in Hypertext Markup Language (html), the ESP logo, any text or drawings prepared by ESP or others working on behalf of ESP and the final copyright notice.
Relationship	includes any business partnership, memorandum of understanding or other business relationship or any verbal agreement that provides a financial benefit to any person.
Submission Request	the form that declares your ownership of the submitted article, and includes your full name, postal and e-mail addresses.

Ownership

All material submitted must be your own. Reprints of existing articles and/or the work or rework of others is not acceptable without written authorisation from the original copyright holder. Any person submitting material for publication indemnifies Elliott Sound Products and Rod Elliott (hereinafter referred to as ESP) against any legal action that may be taken as a result of the publication of the material either in whole or in part. ESP shall not be held accountable for any breach of copyright, as submitted material is accepted in good faith that no prior claim is held against the published work.

Where the material submitted warrants it, you may be asked to fill in the Submission Request form, and return by post or fax to ESP. The form declares your ownership or other rights to the material. This form is available as an HTML page, and may be submitted with any material if you would like to do so. It is expected that this will not be necessary in the majority of cases, but the choice will generally be yours.

Copyright

The copyright of the original submission of any published material remains your own absolutely, and may be re-used or re-published at your discretion. The final published

material, including but not limited to any drawings or additional text added by ESP shall have joint copyright, and the final published material may not be re-used, sold or re-published either in whole or in part in any form whatsoever without my written consent.

I will not sell, re-use or republish your original submission or the final published material without your written consent, either in whole or in part. Should ESP design a printed circuit board or any other product based on any submission that is subsequently offered for sale, you will be compensated at a rate that will be negotiated beforehand.

As the primary copyright holder, you may request the withdrawal of your published material at any time. I will delete the material from my site completely, within an agreed time period. I reserve the right to retain a copy of the material for my own reference, but will not disclose any part of the material to any other party without your written permission.

All credit for the work will be given to the author, and ESP's only credits will be for editing and additional comment or explanation if necessary.

Content

Material that is offensive, inflammatory or defames other persons or organisations is not acceptable, with the sole exception of exposing fraud or deception. If you choose to do this - make sure you have your facts straight. ESP cannot and will not take any responsibility whatsoever if you are wrong, and will withdraw any article without notice if requested to do so by any Government agency or other authority acting for or on behalf of anyone who claims (rightly or otherwise) that they have been damaged in any way by the published material. You will be named as the author if this information is requested, and such authorities shall be informed of your sole responsibility for the material as set out herein.

Articles or projects should be related (directly or indirectly) to audio, and shall be non-commercial in content. Advertising of any sort is not acceptable. Application Notes are a special case, and may be based on specific manufacturers' data, and do not need to have anything whatsoever to do with audio.

References to commercial component outlets or resellers are only accepted where a specific component is unavailable elsewhere, and where the seller is unrelated in any way whatsoever to the author.

Philosophy

Material does not have to agree with my own published philosophy, but I may have second thoughts if it is diametrically opposed. For example, an article about the "great benefits of super conducting mains leads" (power cords) will probably not be acceptable, since I have grave doubts about such things. However a well constructed article on the differences that can be measured between different power leads is most likely quite OK. Note that differences and benefits should be quantifiable - purely subjective comparisons are useless unless properly conducted double blind testing has revealed that differences do exist. Where possible, any such article should be sufficiently well developed to be able to explain the differences in real (as opposed to imaginary or indefinable) terms.

Editing Rights

Editing shall include spelling and grammatical checking. Australian English spellings will be substituted for others where appropriate (for example, "travelling" versus "traveling"). If any change affects the original meaning or intent, you must advise me of this as soon as the discrepancy is found and it will be corrected.

Drawings that do not fit the standard browser page width, or are of a style that is considered unsuitable for any reason will be redrawn or rejected. If rejected, you may redraw the diagram and resubmit it.

Payment

This is very simple - there is none (with one exception, see below). I do not make money from my pages (other than from the sale of PCBs for my projects), and with no advertising have no means of payment whatsoever, unlike magazines. The Audio Pages are for the enjoyment of everyone, and in the best traditions of the WWW are (with very few exceptions) completely free. By the same standards, I expect no payment from contributors, for exactly the same reasons.

Product Sales

The only exception to the "no payment" policy is where you have boards, kits or other material of any kind that is offered for sale. Since you will get the benefit of the wide circulation of The Audio Pages, I will expect a percentage of the profits of sales made from this site. This is negotiable on a case by case basis, and will generally be based on an honesty system. Any breach of good faith will ensure the immediate withdrawal of the material without notice.

As noted above, the same applies if I make a PCB or kit from your submitted material. I will then pay you a negotiated fee for the right to use your idea. This will again usually be a percentage of the selling price or a single upfront fee based on mutually agreeable terms.

Acceptance of Terms and Conditions

If you submit any material for publication, it is accepted that you agree to the terms and conditions as set out herein. Your failure to read these terms and conditions in no way relieves you of any obligation nor obligates Rod Elliott (including but not limited to his heirs and assigns) to any penalty imposed due to breach of copyright or any other matter whatsoever. You agree with the terms and conditions and accept full responsibility for all published material based on your submitted material.

Format of Submitted Material

The material you submit should be as follows:

Text	An ASCII text file is preferred. Since the risk of computer virus infection is high with document files, I prefer not to have anything submitted in this format. Since I will not open any Microsoft Word or similar file in anything other than a text editor, any formatting you may apply will be lost anyway.
Photos	JPEG is the preferred format for all photographic images. These may be greyscale ("black and white") or full colour. The size must be compatible with the standard web browser page sizes, and should be able to be displayed without scrolling on a standard 1024 x 768 pixel display. File size should be limited to less than 10kB (10,000 bytes) if possible.
Drawings	All line drawings (including schematic diagrams) must be in GIF (Graphic Interchange Format), and should be formatted with black text and lines on a white background. Like photos, they should fit in a standard browser window without scrolling (if possible). File size should be limited to less than 10kB (10,000 bytes) if possible.
References	Where the submitted material contains significant reference(s) to other published or unpublished works of any party, these must be stated if at all possible. Material that obviously uses references that are not acknowledged will be rejected, or placed on hold until the necessary references are supplied.
Printed Material	Any printed material must be in a format that can be scanned and read with basic OCR software or converted into an image file.

Software	Any software, spreadsheet or other computer based material that is submitted should be designed for the Microsoft Windows (98SE or above) operating systems, as I cannot verify operation for any software designed for other operating systems. Software must be certified as virus-free, and 'adware' is absolutely banned. Where this is not possible or available, suitable "screen shots" should be provided to allow me (and other readers) to see what the program is meant to do. (It is unlikely that non-M$ software will be published.)
Completeness	The submitted material should be complete, without any omissions that could make the material unusable to the average reader.
Supporting Material	I need your full name, postal and e-mail addresses, and a brief description of the material. Use of the Submission Request form (see below) ensures that I have all the information I need, and you will have a permanent record of what was sent (and when). This establishes your copyright to the material, and is also for my own records. This information will not be disclosed to any party other than a government body or other authorised agency with a genuine need to know, other than at your request. Should this information be requested (or demanded) from me, I will let you know as soon as possible.

Thank you for taking the time to read this information, and for your interest in submitting material for publication. With your continued input and support, it becomes possible to build the most informative and useful web site in the world for all things audio.

View and fill in the Submission Request form. Use of this form is voluntary but recommended. All instructions are on the form.

---

e-mail ESP

Main Index

Copyright Notice. This article, including but not limited to all text, diagrams and/or images, is the intellectual property of Rod Elliott, and is Copyright © 2000-2005. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created and copyright © 29 May 2000. Updated June 2005 - minor undates added./

The Audio Pages

Elliott Sound Products **Frequently Asked Questions**

Updated 01 Feb 2006

Over the past few years I have answered many, many e-mails. Some ask questions of the circuits and articles presented here, but I also get a lot of general questions as well. The biamp article still creates quite a few questions by itself, and along with the others, I have added this FAQ page. Some of these FAQs used to be located in the Readers' Feedback page, and have been moved to make them more accessible.

Some of these questions are (almost) exactly as asked, while others are a "composite" of many similar questions. I shall leave it to you to decide which is which (assuming you care, that is)

Main Index

Contact ESP

FAQ Index

- Bi-amping and Tri-amping
- Amplifiers
- Mobile (Car) Audio
- PCB Information
- Miscellaneous

Bi-amping / Tri-amping

Q: Can I get a schematic for your phase-coherent crossover?

A: Yes. The Linkwitz-Riley version is currently the favourite, and is available on the Projects Page

Q: I am thinking of tri-amping. What do I need to look out for?

A:Be very careful of DC power-on transients, which will destroy tweeters. A capacitor (not less than 20uF) in series will prevent damage, but preferably do not use an electrolytic (bi-polar or otherwise) unless you are willing to accept (possibly) compromised sound quality or (more commonly) relatively short life. Note that many amps have a "de-popping" circuit (generally a timed relay), which will eliminate the DC "transient", but usually offers no protection against O/P device failure! (Read the updated section on tri-amping in the Biamp article) You might also want to have a look at the article on Class-A amps, which are ideal for the top end.

It is also worth looking at a DC protection scheme, such as P33 or P111.

Q: I want to bi-amp my system, but don't want to remove the internal crossovers. Will this work ?

A:Yes, but with some caveats. You will not get the same power gains, since both amps will be reproducing the full frequency range, but you will get some degree of benefit in terms of intermodulation distortion, etc. Overall, this is not the ideal solution, since it is almost the same as bi-wiring, but uses more amplifiers. IMHO it is a waste of time and money.

Q: Can I biamp without an electronic crossover?

A: No. All you will be doing is "active bi-wiring". See the previous Q&A

Q: If I disconnect the internal crossovers in my speakers, what should I look out for?

A: First, make sure that your electronic crossover is set for the same frequency. Second, for a 3-way system, you have to keep the mid to high passive crossover section (unless you are going to tri-amp). This will involve determining the circuit of

the existing crossover, and only disconnect the low to mid+high section. If you do not know how to do this, seek help from someone who does. **This is important!**

Q: My passive crossovers use impedance correction circuits. How do I disconnect the crossovers and leave these in place?

A: You don't. These are needed so the passive crossover is not affected by the speaker impedance variation. Electronic crossovers are not affected by impedance changes, so the correction circuits are redundant. Although they waste energy and can be removed, leaving them in place will usually not hurt anything - just don't expect an improvement.

Q: What about (notch) filter circuits? These are often used to remove objectionable tweeter resonances, so these still have to be there, right?

A: Probably not. If you are replacing a 6dB or 12dB passive crossover with a 24dB electronic circuit, you may well find that these are no longer needed. Because of the original crossover slope and the impedance interactions, tweeter resonance effects are common. They are all but eliminated by the steep slope of an electronic crossover. If the tweeter resonance still causes a problem, it is better to remove the offending frequency electronically. Passive notch filters can make the amplifier load "difficult", and may cause instability in some amps.

Q: In thinking about the sound difference between passive and active setups, I have come up with a question that perhaps you would give me your thoughts on.

I am wondering whether the main reason an active crossover sounds so good is the fact that you truly CAN have same-length speaker cables to each driver, whereas in a passive setup you don't?

Just as background, I have a stereo power amp sitting behind each speaker. One channel drives the bass and the other drives the mid+high crossover. With your active crossover in place, an interconnect feeds the power amp and then two, same-length speaker cables connect to the bass and mid drivers.

In the passive setup, the low-pass section for the bass driver has 2 coils in series - which are many metres of wire - whereas the 12dB high-pass slope of the mid-range

driver has a cap in series. Therefore the length of the cable carrying the signal from power amp to bass driver is many times the length of the mid-driver signal cable. I understood the "Audio Rule" was that you must have same-length speaker cables or you screw up the sound stage ... so, by definition, passive crossovers cannot obey this rule due to the coils that they use.

A:Such a long question deserves a long answer (and needs it, too) The difference in path length with passive xovers is large, but the electrical delay is small (other than the phase delay that is common to all crossover networks). I still think that the major benefits of an active crossover are the phase coherence, lack of impedance effects on the crossover, and the fact that each amp has a limited frequency range and you get extra headroom compared to a single amp. The active version is completely unaffected by the driver characteristics (which change with frequency, power level, etc.), and in turn affect a passive crossover's performance.

The equal length cable theory is a bit of a myth really. You can prove this to yourself by running a 10 metre and a 5 metre cable in parallel (or other numbers that remain passably sensible)

If the "rule" were true, then you would hear the difference, but the propagation delay is only nanoseconds, so you won't hear any change. Electric current flows along cables at around 2.25 x 10 ^8 metres/second worst case, so if one had a 1 metre and a 100 metre cable in parallel, this represents a 440ns (nanosecond) delay between the two. This would create a 1 degree phase shift (delay actually, but let's not split hairs) at a little over 6 kHz. Even at over 20 kHz, the shift is only about 4 degrees - you get more than that acoustically by moving your head a few millimetres.

The cancellation caused by a 440 ns delay between parallel conductors is 0.08 dB at 100 kHz (which is negligible), so at normal audio frequencies it may be completely ignored.

Q: I have heard that bipolar electrolytics are commonly used in crossovers. You and just about everyone else don't like them, so why are they still used?

A:Price. They are cheap and have high capacitance in a small volume. However they are not reliable at high power, and will degrade in time. It can be Ok to use them in impedance correction circuits, since they are not in the signal path (they

are in parallel with the driver and amplifier when used like this). However, because of the potentially high current they will be subjected to, their life may be reduced quite dramatically

---

Amplifiers

Q: Can I increase the supply voltage on your amps to get more power?

A:All power amplifiers have limitations on the maximum supply voltage and minimum load impedance. Although you can use parallel output devices, the driver transistors, Class-A driver and input stages also have limitations.

For reasons that remain completely unclear to me, everyone wants to operate P3A (in particular) at above the design voltage rating, which is ±35V (or ±42V into 8 ohms **only**). 32VAC (for example) will give ±45V, and the amplifier will fail. Maybe not today, nor tomorrow, but you will stress many transistors beyond their design limits, so eventual failure is guaranteed.

Power dissipation (and the likelihood of output device failure) is based on many things, and these were taken into account when the amp was designed. As voltage increases, the probability of device failure rises exponentially.

The dissipation into a resistive load must not be used to determine safe operation, and the worst case load is a loudspeaker near resonance, where the voltage and current are 45°out of phase. Instantaneous peak output transistor dissipation is *doubled*under these conditions, and if the transistors are hot and pushed to their limits, then failure is a certainty.

200W transistors will be driven into potential destructive operation with ±42V into a 4 ohm load (peak dissipation is in excess of 200W). ±35V is a safe and 'transistor friendly' voltage, and the amp will give many years of faithful service at that voltage. Even at ±35V, peak dissipation will reach about 150W into 4 ohm loads!

All of my designs are deliberately conservative, since many are built by novices, and it is obviously better if they work reliably. All attempts to obtain more power by increasing supply voltage (and with little or no understanding of the stresses on the transistors) reduce reliability, and place the amp and speakers at risk.

For P3A, ±42V is the **absolute maximum**voltage - not 45V, and not anything that is greater than 42V. Period!

To use a higher voltage, the input transistors, current sink, Class-A driver, drivers and output devices must all be upgraded, then the input device bias will change, and performance may be reduced.

Amplifier design may appear to be 'trivial', but it is not, and any change to any operational parameters should only be done with a full understanding of the ramifications.

It also needs to be understood that seemingly large gains in power are not worth the decrease in reliability. P3A at ±35V will provide an instantaneous peak power (before the supply voltage collapses) of just under 75W (remember too that the supply voltage will normally be higher than expected because of transformer regulation). Also remember that the full +ve to -ve supply voltage can (and does) appear across the transistors, so with a ±42V supply, all driver and output transistors must be able to withstand 84V.

If the supply voltage is increased to ±42V, peak power increases to 110W. It may *look*like a lot more power, but it is only 1.6dB - you need a **10dB**increase for an amp to sound "twice as loud", and that requires going from 75W to **750W**. 1.6dB is barely audible as an increase, and is not worth risking the reliability of the amplifier for.

Q: Would you consider a design for a 1,000W amplifier?

A:No, I would never consider a 1kW amp. No loudspeaker (other than a small number of very expensive professional types) can handle that much power, regardless of claims. Look at a 1000W electric heater - feel how hot it gets and how quickly. Note that the element is thick resistance wire on a ceramic former. The resistance wire glows red hot after only a few seconds.

Since only about 1% of power is converted to sound, the voicecoil has to do the same work as the electric heater. But look at a voicecoil - it is thin wire on a thin aluminium or Kapton former. At 1kW it may last 10 seconds if you are lucky. At low frequencies, the voicecoil will leave the gap, the speaker will distort badly, and quickly fail. In some cases, the spider or surround can be torn, or the voicecoil can suffer impact damage from colliding with the rear pole-piece

Unfortunately, the advertisments and even data sheets lead one to believe that many drivers will take that kind of power. The vast majority will not. I have tested a (claimed) 500W driver that distorted (and would have quickly overheated) with

only 100W input. It was a massive affair, having dual voicecoils, dual spiders, dual magnets and double the distortion.

Although it is useful to have headroom in an amp, it is far better to use smaller amps and an electronic crossover. This makes the requirement for 1kW amps simply go away for all domestic hi-fi systems.

Q: What is power bandwidth and what relationship does it have to overall amplifier sound quality? Does it have some relationship to transient intermodulation distortion, even if indirectly?

A:This remains a contentious question. Power bandwidth is (in amplifier terms) the bandwidth that the amp can provide full power (usually measured at the -3dB frequency). It is closely related to the slew rate of an amp. If any signal is faster than the amplifier can handle, then intermodulation products are generated (TIM). Few modern amplifiers suffer this problem - even if the amp can only provide full power up to 10kHz, this will not cause slew rate limiting, as the signal levels are so low. There is some evidence that such a limited bandwidth creates other audible effects, possibly due to the much lower gain at higher frequencies, meaning less feedback and higher distortion. Just about every designer on the planet will give you a different answer :-(

Q: When an amplifier gets a complex audio source (such as an orchestra soundtrack) does this create greater load conditions at the speaker end and adversely affect damping ratios determined in the normal manner.

A: An amplifier does not care one iota about the complexity of the signal. A single frequency or multiple frequencies all at once make no difference. At any point in time, there is only one voltage present, and the amp will amplify it. This of course is only true if there are no frequencies that cause a rate of change of the signal that is outside the bandwidth of the amplifier. There are two distinct (and separate) things to deal with - the instantaneous value of voltage, and the rate of change of this voltage. If either is outside the amp's capabilities, you will get distortion (of one form or another). Output impedance is (to some extent) frequency dependent, and varies with the feedback ratio - this in turn is reliant on the open loop bandwidth of the amp. After this, it gets complex !

Q: What I mean is does an amplifier's output impedance actually rise during complex musical production or is this all nonsense?

> **A:** This is nonsense. The amplifier's impedance is affected by frequency (so will be different at different frequencies), but the complexity of the music has no direct effect on the impedance.

Q: Does any of the above have anything to do with why valve and solid state amplifiers can sound different?

> **A:** It can. Valve amps usually have a power bandwidth that is almost independent of feedback, and is limited mainly by the valve electrode capacitance(s) and the output transformer construction. Have a read of the article about valve and transistor amps "Valve amps - do they really sound different". There are quite a few other factors (output impedance especially) that have a greater influence on the perceived sound quality.

Q: I suspect that many or all of the above potential problems are greatly reduced by bi-amping or tri-amping.

> **A:** There is no doubt on this score. Anything that reduces the demands on an amp helps, but it helps even more when the speakers are driven from a defined and constant impedance - this does not happen with most passive crossovers. Biamping is more about the speakers than the amps. Very few (if any) modern amplifiers are stressed by the full range signal, although impedance dips caused by passive crossovers may be responsible for some subtle (or not so subtle) effects on the sound. This has not been proven (to my knowledge).

Q: Can I operate the 60W amplifier (or any of the others) into a 4 ohm or 2 ohm load for more power?

> **A:** Four ohm loads will often be acceptable, but I strongly recommend using paralleled (or more powerful) output transistors. It may also be necessary to use a small heatsink on the drivers, as their dissipation is usually more than doubled. Two ohm loads are very difficult for any amp not designed specifically for this impedance. I do not recommend using any of the published circuits for 2 ohm loads.

Q: Is is better to use transistors or op-amps for amplifiers built today?

A: It depends on what you want to achieve - a well designed discrete power amp still outperforms any power opamp, but good opamps outperform (technically, at least) any discrete circuit for preamps. However, the P37 (DoZ preamp) circuit is a good indicator of what can be done with only a few parts, and that measures well and sounds very good.

There is no simple answer.

Q: I don't have an oscilloscope, so how can I know if my amp has crossover distortion?

A: A reader sent me his idea for this, and it is both simple and ingenious. You need a very clean sinewave source of between 100 and 150Hz. Set up the amplifier providing some power into a resistor load - about 1V RMS into 8 ohms is usually enough. A cheap piezo tweeter connected across the load will make harsh clicking noises if crossover distortion is present. Adjust the level to zero to make sure that it is not spurious noise.

If you are adjusting the quiescent current on an unknown amplifier, set it so that the clicking noise just goes away (or stops decreasing). You must check that all transistors remain at a sensible temperature. (Thanks to Raymond Quan for the idea.)

Q: As far as I can tell, there is next to nothing on PWM amplifiers at your site, despite the fact that recently there seems to have been some dramatic improvements in chip(set)s that enable truly amazing compact (size of a credit card) amplifiers. I find it curious that you don't seem to have addressed PWM amplification at all.

A: You are (almost) correct - there is not a lot, but there *is* a very comprehensive article that discusses PWM/ Class-D amps.

There are many problems with DIY Class-D amps, and the possible need for surface mount components is only a small part of it. Making/debugging something like that requires considerable expertise in digital signal (or switchmode) analysis, and absolutely requires the use of an oscilloscope. The required parts are also hard to get and relatively expensive, especially in small quantities.

There is nothing curious about it - many people have problems getting a simple discrete amplifier to work, and I don't even want to contemplate the questions I'd get if I had a PWM amp project. I agree that it is interesting, and in a few years most high power amps will be PWM, but at this stage I would be a fool to develop and publish one !

When PWM amps do become available from me, they will be fully built and tested.

For more information on the design of amplifiers, see the various articles (and references) I have published.

Mobile (Car) Audio

Q: i have a 1200watt car amp and 2 600watt car speakers with no car. witch brings me to the problem how do i hook it all up in the house. if you have any advice e-mail me at e-mail@address.com. tanks!

A: In case you were wondering, that was an actual e-mail I received (some are a lot worse!). You will need a big power supply, probably with a 12V car battery (situated outside the house, please!) for backup and to provide for surge currents. A battery charger will not be helpful, because there will be excessive noise on the 13.8V supply (that's the actual charge voltage for a 12V battery, and **all**car amps are power rated at 13.8V).

I get quite a few of these questions, and quite frankly, I'm sick of them. This is not a trivial undertaking - one person even suggested that he already had a wall transformer, and wondered if that would work - it wouldn't. To those who may be thinking of asking ... don't. There are many forum sites that deal exclusively with car audio systems, and you should ask them, not me.

Q: Do you have a design for a car power amplifier. I want to use your 60W amp from the 12V supply, what do I need to do?

A: For useful power from conventional hi-fi amplifiers, you need to increase the supply voltage. This requires a switch mode power supply.

Q: Are you planning to publish a suitable switching supply suitable for any of your

amplifier designs?

A: Yes, a 300W supply design is available on the Projects Pages. However ... there is a fundamental problem with switch mode supplies - the ferrite transformer core. Unless you can get one that is quite similar (and achieve the same primary inductance), the supply will just blow up. Unfortunately, these cores are not readily available to hobbyists, but they can be obtained. In addition you need high speed (Fast or Ultra-Fast) diodes, and again, these are not easy to get. The layout and construction of switching supplies is such that they are difficult for the home constructor inexperienced in building them.

Q: About the Project 89 Switchmode power supply for car audio ... Can it be modified to work (step-down of course) from 220VAC (120, 240, etc.) rectified to 300VDC? If not, where can I find a design for a +30/-30 SMPS rated at 500W or more?

A: The SMPS is not suitable for 220V (or any other mains voltage) operation without major re-design. It is an entirely different matter to make a high voltage step-down system, and apart from PC type supplies, I have not seen a schematic anywhere on the web for an off-line switcher. As a DIY project they become rather daunting (and dangerous), and it is highly unlikely that I will ever attempt a design for publication.

Q: Can I use your preamp or crossover in my car system? I am especially interested in the Linkwitz-Riley crossover / sub-woofer controller / quasi parametric equaliser.

A: Yes you can. You need to establish an "artificial" earth (ground), at 1/2 the supply voltage, and input / output capacitors must be used (make sure you check the polarity of electrolytics). There is a project for a simple preamp and artificial earth in theProject Pages. This can be adapted to your needs. Alternatively, you can use Project 69 (Switchmode power supply for preamps) to obtain ±12V at up to 45mA

Printed Circuit Boards

Q: Can I get a copy of the PCB artwork for any of the projects?

A: No. I do not supply artwork to anyone. The PCB layouts take considerable time

to develop, and I must not only recoup the time spent, but I also need to be able to pay the costs of operating the website. This is my primary source of income, so I don't think it is reasonable to expect me to give away my intellectual property. No-one else who sells products will give you their design information either.

Q: With all of your boards, I have to run wires for the pots. This is a real pain. Why don't you use PCB mounting pots?

A: I dislike running wires as much (or more) than anyone, but if I have the pots on the board, you are limited to using the same type of pot (some are not readily available in other parts of the world). In addition, you would be limited to using the same layout as I designed for - the pots would be the distance apart that I designed the board for, and the PCBs would be considerably larger (and therefore more expensive). You would be much more restricted, and would not be able to use the layout you want.

Q: I sent a faxed order, but I didn't receive any confirmation. The first I knew that it was received correctly was when the boards arrived. Why didn't you let me know?

A: I normally only contact you if there is a problem with the card details, if I am temporarily out of stock (rare, but it has happened), or if I can't determine what you want. If the fax is clear, and I can validate it properly, then the order is processed and despatched, usually within 2 (±2) days of receipt (i.e. it may be immediate, or could take up to four days, depending on my workload).

Q: Can you give me an estimate on the cost of assembling Project (insert any project number)

A:No. The cost of parts varies from one supplier to another, and even more energetically between different countries. It is simply not possible for me to try to maintain any sort of price list, as there are too many projects, too many suppliers, and too many countries and currencies. This is your job, mine is to produce the projects

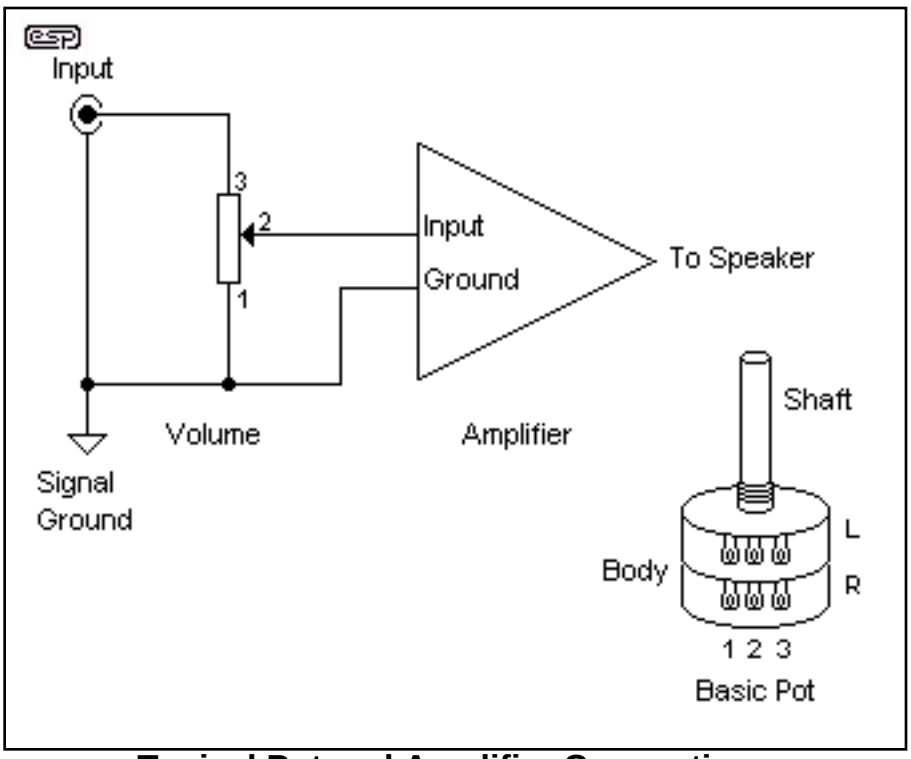
For power amplifiers, a reasonable estimate can be made by pricing the power transistors and doubling it to cover the cost of the smaller devices and passive components. The heatsink, power transformer, filter capacitors and case will normally exceed the cost of the amplifier itself by a wide margin. The availability (and cost) of these varies widely from one country/ supplier to the next.

Q: Do you have power supply boards for any power amplifiers ?

A:No. The power supply for power amps consists of a transformer (too large to mount on a PCB) a bridge rectifier (which needs to be chassis mounted for cooling) and filter capacitors. If I designed a board for filter caps, it would cost a ridiculous amount of money because of its size, and you would have to use caps of the same physical size as I designed for. This would be extremely restricting, and would make component selection much harder than it should be. Power supplies are best hard wired using thick wiring and you have far greater flexibility for mounting.

Q: How can I add a volume control to your Project "XX" power amplifier? (XX is any amp project number)

A: This is a *very*common question, and is fully explained in the article about potentiometers (pots). However, I will also cover it here, as I might save some e-mails. The basic schematic for a volume control and a drawing of a pot are shown below ...



Typical Pot and Amplifier Connections

For more information, see the article Beginners Guide to Potentiometers. Generally, a value of between 10k and 25k (or thereabouts) is fine, and log pots are "traditional". The above article shows how to make a linear pot behave more

logarithmically than a "real" log pot, and this is usually very much better than cheap to medium priced log alternatives. Expect to pay fairly serious money for well matched conductive plastic log pots, or get the same performance from a budget carbon pot suitably modified.

Miscellaneous

Q: I have been doing some research into what my next hi-fi upgrade might be and i have been reading about how vast am improvement can be made with the filtering of AC power.

According to 'everyone' such a device can lower noise floors on a system so that all you hear is the music and no longer the system. It makes sense that the cleaner the ac signal the cleaner the output signal will be although i have been unable to find any really hard evidence that this is the case (in terms of graphs or whatever).

What do you know or think about this? I have heard of several types of filtering methods, ranging from some product called a ***** to battery banks to run amps from cleaner DC.

A: First, with no music, place an ear next to the speaker drivers. If the noise you hear is just a gentle hiss, then you are already at the system noise floor, and there is not much you can do to improve this. Hums and buzzes are probably due to wiring, and re-routing the leads keeping power separate from interconnects (etc) may improve matters.

Unless you have recognisable evidence of mains interference, external filters and conditioners serve no purpose. I'd steer well clear of anything that claims to be a 'magic bullet', and batteries may be noisier than a good mains supply in some cases (the chemical reactions create electrical noise). Mains filtering is only effective and useful if the system's background noise is audible from the listening position, and is the result of mains interference.

If there is no interference and the system noise is inaudible from the listening position, then nothing needs to be done, as you do not seem to have a problem.

Q: Can we obtain tremendously high voltage gain using a FET or valve (vacuum tube) with a current source load or bootstrap circuit

A: Using a current source (or bootstrap) load with valves or MOSFETs will make

them more linear, but does not increase the voltage gain by anywhere near as much as with bipolar transistors. The voltage gain is almost an illusion, largely because the bipolar transistor is a current device (output current is determined by input *current*, not voltage).

Q: What is the reason that bipolar transistors have such a high voltage gain when used with a current source load versus valves or MOSFETs

A: Transistors are current controlled current sources, and the others are voltage controlled current sources. Because the current is controlled by a voltage, valves and MOSFETs have a much lower voltage gain, and high impedance loading (however applied) will not cause a huge change in the transfer characteristics. (This *is*the simple explanation, by the way)

Q: What is your opinion on using an "off-line" power supply, using a capacitor to drop the mains voltage, then regulating using zener diodes. Will it affect the sound? For better or worse?

A: DON'T DO IT !These supplies are extremely dangerous, and are intended only for equipment that is not connected to anything else (for example, stand-alone mains appliances). This method must never be used for audio equipment, since there are always connections to the outside world.

Q: Will an amp that is built with film caps will be better than with ceramic caps? Can you advise any sources on the Net about how caps affect sound and what ones are better for using in audio projects?

A: Film caps versus ceramics is no contest. Ceramics are not particularly stable, and their capacitance varies with voltage, temperature and whim. They are by far the best for bypass applications, but should never be used for coupling or in filter circuits and the like. For coupling and any filter application, always use film or "metallised" types. There are claims that some film caps sound better than others, but I know of no definitive test that has ever proven this (for listening tests - many measurements will show differences, but these have never been proven to be significant in an blind test).

Q: I am thinking of using Brand X amplifiers, and Brand Y cables. What do you think?

A: I will (generally) not endorse any brand of amp,cable or other product in these

pages, although the occasional comment may be in order. Basically, my response is to try it and listen to how it sounds.

Q: I want to build one of your published projects. Are there kits or circuit boards available?

A: I currently have PCBs available for quite a few of the more popular projects, and eventually hope to be able to offer complete kits as well. This remains some way off at the moment because of the cost of purchasing large quantities of expensive devices. Check the Purchase PCBspage for the latest information.

Q: I am confused about transformer voltages and VA ratings, and I want to build a power supply for an amp I am building. What do I need?

A: This is a very common question, and the answer is different in nearly every case. The article on Amplifier Design has a few details, but there are extensive answers in the Linear Power Supply article, as well as the beginners' articles on transformers.

The VA rating is simply the product of volts and amps (hence VA) - it is not called "Watts" since the VA rating and Wattage are very different due to the rectifier and filter caps. A rule of thumb for Class-AB amps is that the transformer should be rated at a minimum of 1.5 times the maximum power from the amplifier(s). For Class-A, the ideal is that the transformer is rated at a minimum of 4 or 5 times the RMS amplifier power rating. It must be remembered that the load in Class-A is continuous, and a transformer that is too small will overheat and be ruined.

Q: Which of the amplifiers in your projects pages sounds the best?

A: Perhaps surprisingly (perhaps ??), I don't recommend any one of them over any other. The Audio Pages are about experimenting, learning and building. All the amps (and preamps) work well and sound good, and it is up to you to decide which one you want to build.

Q: Do you do custom design work?

A: Yes. Send me an e-mail, and I will see what can be done. I do not have the time (and you don't have the money) for development of projects for personal use. My time is very limited at the moment, so I may not be able to help at all.

Q: Do you plan to produce a project for an AM/FM hi-fi tuner?

A: No. These are quite reasonably priced as commercial offerings, and are too hard (i.e. almost impossible) to align for the home constructor with no RF test equipment and specific skills in this area.

Q: What does 4k7 (100R, 2n2, etc) mean? Does this mean 47k (100 ohms, 2.2nF, etc)?

A: 4k7 means 4.7k - this is basically a European standard and is used to ensure that the decimal point is not missed. 100R means 100 ohms, and is used so the omega symbol does not have to be inserted all the time. 2n2 (and similar) means 2.2nF, and the same logic applies as with resistors.

Q: What does 0R47 mean?

A: 0R47 means 0.47 Ohms - another variation to the European standard and again used to ensure that the decimal point is not missed.

Q: What does 2.2nF mean" What is "nF"?

A: A nanofarad is $1x10^{-9}$Farad, and is equal to 0.001uF. This is the preferred nomenclature for capacitors between 1000pF (1nF or 0.001uF) and 1uF (1000nF).

Q: Do you have time to even answer one of these pain in the butt questions?

A: It would seem so

Q: On a different subject altogether, I heard that you helped set up the SAE (School of Audio Engineering). Is this true?

A: Yes. Along with John Burnett and Tom Misner. I was only there for a short while, until I became aware that the operation was not (IMO) above board. Without going into detail, it transpired that I couldn't work with Tom Misner for a number of reasons, not the least being his (self acknowledged) lack of ethics. As a result, I quit on the spot, and John (Burnett) followed a few months later.

For more information on the topic, have a look at John's education FAQ page - Lenard Audio FAQ.

Main Index

Contact ESP

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2000-2006. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro-mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use of the material herein is prohibited without express written authorisation from Rod Elliott.

Elliott Sound Products **ESP Links Page**

Main Index

Recommended Links

coldamp	coldamp - Class-D audio power amplifiers, switchmode power supplies and other innovative audio products. coldamp is a company dedicated to design and manufacture high efficiency audio amplifiers and other complementary elements for audio systems.
GUITAR Tips Online	**Discover the fastest way to learn the electric or acoustic guitar - online.** Learn from the best - Chris shows you how to play the guitar in exactly 30 days, and you can bypass traditional guitar lessons and achieve real results! This is a very good course, and is genuinely recommended.
SSS	Stones Sound Studio - Artisan audio art series speakers, speaker design, plans, drawings, pcb & crossover design, components, active filter design, speaker driver t/s measurements. Russell Storey, senior design engineer, acoustical consultant, high end audio products sales and service.
NOT GOOD ENOUGH .ORG	What can I say about this site? It is everything that the government "watchdogs" are not, and is probably one of the best consumer action sites around. NotGoodEnough is an Australian operation that does everything it can to get consumer problems sorted out with manufacturers, retailers, banks, airlines, etc. **Highly Recommended** (International membership is available!).
LENARD	The Lenard Audio site has an excellent educational section, and belongs to a friend of many years. You can see some of the commercial products I worked on in the past (and/or am working on at present) with John, and browse through the education sections to understand some of the fundamentals of audio and its application.
Class-A Amps	Geoff Moss (my defacto editor and contributor) now has re-instated his page, primarily devoted to Class-A amplifiers. Geoff has all the details on the John Linsley Hood classic (both versions) plus a lot more Class-A amps and a great deal more helpful information on various amp topics.

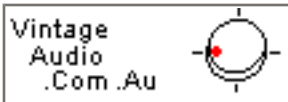
RANE	Rane Corporation - probably one of the most open and informative pro audio manufacturer websites ever. There is a wealth of excellent design information, covering a wide range of topics. Not only for professionals, it is just as applicable to hi-fi. Check out the PDF library - I hope you have lots of disk space :-)
The Self Site	Douglas Self - well known designer of amplifiers and audio equipment generally, published many times in Wireless World (now Electronics World) - This is an excellent reference site for design information.
Audio Myths	Many of the things that are taken for granted in audio are myths, and on good days have the credibility of alien abductions or miracle cures. The rest of the time ... well, find out what someone else has to say.
Steve Ekblad	Audio Related Internet WWW and FTP sites. One of the all-time great links pages
True Audio	TrueAudio - Info on speakers and their design + loads of other stuff
ThieleSmall	ThieleSmall dot com has a driver database that is second to none. Jan has put a lot of work into this, and for the speaker builders it will become a very useful resource as it gets known
ValvesUnlimited	For those who just love FETs with pilot lights - I am one of them, does that say something?
CyberTracker	Singapore's Unofficial Audio/Visual Technology Pages
DIYaudio	The name says it all, really. Also some more good links
Linkwitz Labs	Siegfried Linkwitz (of Linkwitz-Riley crossover fame) has some extremely useful information, especially on the design of bipolar loudspeakers and other speaker topics. Some of this information will change the way you think about reproduction of audio.
AudioXpress	AudioXpress is the home of the renowned Speaker Builder and Audio Amateur magazines. Subscription info and sample articles are available, plus a lot of other info.
eCoustics	Easily find the latest news, articles, and reviews on thousands of consumer electronics and hi-fi audio/video products from the most popular, informative, and reliable A/V web sites.
morrison	Don Morrison Audio - Creator of the E.L.A.D. (Electronic Line Amplifying Device) which has had rave reviews, as well as some interesting articles and nice speakers.
Musical Fun!	One of my link exceptions! This page has a collection of very funny stuff - mostly orchestra musician jokes (pity the poor viola player!), but much more, too. Don't miss the first story: "The Bangkok Piano Recital Review"
GMArts	Another excellent Australian site, this one for guitar players. Lots of info about guitars, amps (etc).
Electronics Calculators	Online electronics calculators - work out resonance, capacitance, inductance, resistance and time constants on-line.
Audio Calculators	Audio calculators for all occasions. There is a huge range here - something for everyone

SatCure	Satellite equipment, MacIntosh computers, and (of course) audio make up Martin Pickering's site. Well worth investigation.
Decibel Dungeon	Nick Whetstone has another audio site that is worth a look, especially if you are interested in turntables and other DIY activities.
SSC Electronics	DC converters for portable or car applications, SSM2017 mic preamp ICs (yes, they have them), plus many other useful items.
AUDIOHOLICS.COM	Audio equipment, reviews and thoughtful articles and home theatre setup hints. Includes a forum, FAQs, and more.
Phase Linear	The official Phase Linear site - Owners' manuals, service information, photos, schematics, history and general information
GedLee	Loudspeakers, home theatre, acoustics & noise control and sound quality are the main areas of expertise, and GedLee is the home of *SPEAK* loudspeaker design and modelling software, and the book 'Audio Transducers'.
Bill Collison	Vented Subwoofer information, showing that it *can* be done, if you are willing to make the effort. There is quite a lot of useful information, allowing you to make a high performance vented system and avoid the pitfalls.
CATENA SIMETRIX	An absolutely top notch simulator. The free version has a limited number of nodes (connections), but is still sufficient for nearly all your simulation needs..

Additional Useful Links

mozilla	Want the best browser and e-mail client on the planet? Of course you do. Get rid of the rubbish you use now, and grab a copy of the finest internet package available anywhere mozilla.org
Audio Analogue	From Italy, you can see Federico Paoletti's Unofficial Audio Pages
MEIsearch	Mobile Electronics Industry Search. This is a mobile electronics search engine and website directory including car audio, mobile video, navigation, mobile MP3, vehicle security and other related websites.
How Stuff Works	This is a great site for just messing around in - not much audio, but lots of other stuff
Soaps for Precious	Soap? Yes, you read that right, soap. Quality handmade soaps with no animal products, and using only the finest natural ingredients.
Hobbytron	Hobbytron carries a wide range of electronics fun and toys. The site features secure ordering, and has kits and test equipment, MP3 players, and much, much more. Also has extensive DIY links.

Exchange Links

	Process Sensors - NIR Moisture Sensors, Moisture Gauges, Moisture Analysers. Process Sensors Corporation (PSC) is a manufacturer of moisture gauges and sensors, providing worldwide industries, accurate and reliable moisture measurements for quality control in manufacturing processes.
	Vintage Audio .Com .Au, Australia's own Vintage Hi-Fi web site where you can see, read about & buy great pieces of vintage audio from the 70s and 80s. All the great brands of the period like Pioneer, Sansui, Marantz, Rotel, Luxman and more.
	Datasheet4U: is a free electronic engineering tool that enables you to locate product datasheets from hundreds of electronic component manufacturers worldwide.
	AudioSold: High end audio, home theater, audiophile, classifieds, DIY & vintage audio
	Canuck Audio Mart - Canada's largest online HiFi/Audio Classifieds Site! Classifieds, Forum, Personal Ads, Links, Etc.
	Home Theater Reviews: Information on home cinema systems and equipment. Get help with choosing your Home Theatre design, setup, accessories and more!.
HotVsNot	Web Directory of Resources, including Audio, Free Submissions are Welcome
	Audionova has a huge links section divided into different categories to make it easier to find what you are looking for (such as DIY, driver manufacturers, software, etc.). A very good resource.
	Summit Batteries - lots of battery data, chargers, etc., and links to battery suppliers all over the world.
delabs	***delabs*** has around 200 electronic circuits in pdf document format or png image format.

More Links - The Audio Miscellaneous Collection

As of March 2000, I have suspended further additions to what is now the "Miscellaneous Links" page. I will only include links here to those sites that (a) satisfy my criteria for content and accuracy of information, and (b) include a reciprocal link to my home page (there are some exceptions, but they are rare for any new links).

Last Update - 30 Apr 2005

Humour Index

It is so often beneficial to have something to laugh at (other than management and the antics of our elected representatives), that I have this humour collection - expect it to grow. Contributions from readers are most welcome.

Please Note: I don't want to offend other readers, although the Joke Collections already have some risque offerings - please, if you don't like "off colour" jokes, don't go there.

ESP Index

Satcure UNIQUE Audio Products

ESP's Humour Collection

Dynamic Range Versus Ambient Noise	By George Izzard O'Veering - some light hearted reading for the technical types
Waltzing Matilda	An "interpretation" of Australia's most famous song
HAL9000	Stanley Kubrick, eat your heart out
OS Airlines	If Your Operating System were an airline (perish the thought)
Monty Python	A trio of the more famous (and suitably bizarre) sketches from those well known lunatics
Murphy's Laws (and Others)	A collection of laws by the likes of Murphy, and joined by various others. If you ever wanted to know what rules your life - here are a few answers.
Joke Collection - 1	An array of jokes to appeal to (or otherwise) just about anyone - Please note that some of these offerings may contain language that could offend.

Joke Collection - 2	More jokes for your enjoyment. This also contains amusing news headlines, hospital emergency room anecdotes and .... jokes!

Please note that Murphy's Laws and the joke collections (1 and 2) may be considered "public domain" - feel free to copy

Elliott Sound Products **Reader Feedback**

Some of the responses I have had from readers. It is most gratifying that so many of you have taken the time to respond. While all e-mails are answered, the answers to the offerings here are not reproduced, as they are basically my personal thanks to the sender. These letters are reproduced verbatim - no editing has been done at all, except to remove the writer's name (and in some cases additional info or questions).

Please note that I have basically stopped updating these reader response pages, as the task was becoming overwhelming. If you have something you *really want* others to see, ask me to publish it (rather than the other way 'round as it was before).

For more readers' responses (and my answers), please see Readers Letters

---

Main Index

More Readers' Letters

Website Feedback

This is a small sample of the comments I get from readers. These are reproduced verbatim, with no modification to the text at all (hence some interesting spellings and typos :-)

Thanks for such a comprehensive treatise on bi-amp.

First of all i must congratulate you for your page, it's far the best i've visit on the net, because the didactics, the projects and the humour.

First of all I want to congratulate you for this great web-side. It's really the best audio/ hobby side I've encounterd. The first time I saw it, I printed out all interesting articles and projects and I could not sleep that nigth till 6:30 because I had to read them all!!! Twice

or three times cause I learned a lot while reading. Everything is indeed well written and has a beautiful style.

Impressive website - finally an engineer/audiophile who hasn't lost touch with reality. I thoroughly enjoyed reading everything!

Nice site. Informative and easy to navigate. I'm interted in seeing some pcb layouts when you get around to it.

Hello, I read your articles on the net and it is so interesting.

Thanks for this very informative overview on bi-amping.

Since I was browsing during office hours I didn't sign your guestbook, but felt I should compliment you on your site - easy to browse and full of sensible advice...

I enjoy your page very much! Please present an article on a digital delay circuit!

I have seen your website but I am not sure what business you are in. I absolutely agree with you about biamping. I have had many arguments with people over biwiring. I very much enjoy your article and feel it has more truth than most articles I have read on audio.

Hello, thanks for the very good article and caveats for the unwary.

Let me say, first of all, how much I enjoyed reading the articles on the ESP website. I found myself agreeing with almost everything you have written.

I have really enjoyed reading your project pages. I do wish that pcb's for the 60 watt amp were available.

Thanks for your splendid guidelines about amplifiers on your website. I didn't know about the nonlineairity of darlingtons and mosfets because of all the designs with them, till I had read your website, so I asked friends of me working in the electronic "scene" and they comfirmed it.

Thank you for a very good article.

I think you have an excellent site. It is one of the best I have found. I wish it were around 3 years ago. I find the articles to be very informative - I hope to see more and perhaps contribute to your site.

Hi .. my name is (name supplied) i,m from egypt , and my age is 21 year I very happy to contact your site and thanks for you to add this site in the internet.

I made your surround decoder, and works well.

Loved your website and the effort that you have put into it. Congratulations and my best wishes.

I surfed over to your site for the first time in a while. All the DoZ information was interesting. Some of it was over my head, and all of it is beyond my current skills and resources. But it was entertaining. Some "audiophiles" should be fairly irritated with the death of their misconceptions regarding their beloved Zen components.

It was great pleasure for me to read your articles of amplification theory and amplifier desigh particularly.
While I am not totally agree with your conception of FETs and MOSFETs usage in HI-FI, your approach to schematics is very reasonable and "classical". The same as mine.

Thanks for good web site. The capacitance multiplier is a great circuit! I have made a Pass zen-amp. Now I am making "death of zen"-amp. The things you have said about the Zen are true. I am not fanatic about any topology, though class-A is most interesting to me. I expect to get better definition and precision with your amp.

I just finished the Linkwitz-Riley cross-over. I built a pair of mid/tweeter bins with two mid-range and two tweeters (4 ohms) crossed over at 3000Hz. Then I built the L-R cross-over at 650Hz (34uF & 5K1 ohms). The bass driver is 97dB while the M/Ts are 94dB each. The Amps are a 125w into 8 ohms for the bass and 108w into 4 ohms for the M/T bins. It is a little top heavy, but when I build a 200w into 8 ohms amp, coupled with the 175w into 4 ohms amp currently driving the bass speaker I think it should be fine. I must say it is very "loud" and very clear in the voice and instrument range, without loosing any of the bottom end. It is an excellent result.

I just read your editorial, and I must say that it is very refreshing to hear someone finally talk some sense! It has always been my experience that audiophiles tend to subscribe to mystical cures to phantom problems they didn't even know they had! Thank you for your wisdom and straight forward approach.

I´m very interested in your "projekt 30" and I think I´m going to build it myself one. I´m recording on my computer and therefore I need a little outside mixer. I also saw that you had one chapter unfinished so I wonder if it´s possible that you can e-mail me some info about this when done? :)

Thanks for the very interesting web site. I look forward to visiting it often. It is nice to see valuable information given for free.

As ussual, your page is great stuff. ABout your Termal-fan cooling and the upcomming Termal Shutdown project. Please only use disctgrete devices (no Op-amp), so the circuit can be integrated into an Power-amp circuit, WITHOUT having to go to low-level PSU (like 12 V).

I am very pleased to have found your web site. I am an EE (Electronic Engineer) working as embedded designer/software developer and have developed a strong interest in professional music electronics since I started running my church's sound system . To my dismay I find that my formal education didn't really equip me for high quality audio equipment design. Over the last few months I have been searching the web for useful design information and practices. To my further dismay, I have found lots of "cookbook" type data but very little in the way of rigorous design. Most of what I have found so far is either based on component vendors application briefs or else the circuits presented don't appear to agree with what is claimed about them.

I find your articles interesting, well written and very informative. I especially like the fact that you lay out the pro's and con's of each design choice as you make it and don't try to hide the mistakes you have made. Too many of the "experts" I have read try to make you think their favorite scheme was handed down to them from on high and is the ultimate solution to all of life's problem. Real life is a lot messier than they allow for, especially since a solution that is a good fit for one problem may end up making a different one worse.
...
I encourage you to keep on disseminating sound (no pun intended, well maybe a little bit) audio circuits, exposing meaningless marketing hype in audio gear and the preaching the importance of testing (instrumenting) what you do.

You rock. I am so happy there are people like you in the world, preachin' the truth and laying it out clearly and plainly. Too many things i've been reading lately are written by people who i suspect don't really understand exactly what they are waving their hands at, and I know because all of a sudden things get too "complicated".

You have given me several enjoyable hours in my favorite coffeehouse so far. I'm slowly going to swallow the entire site, burp, and sit back with a complacent, enlightened grin on my face. What a relief that there's no magic where there shouldn't be, and all you gotta do is ask the right question and think a little.

Keep it up. Pass it on. Thank you.

I stopped by your site after someone posted a link to it from a guitar effects makers forum - so you probably know where I'm coming from here. Just like to say after reading through your reviews or the dodgy amps and spray etc, I was really surprised that at what crap people would believe.

I'm an electronics engineer, so I see and hear a lot of what you're talking about. Glad to see someone showing up these morons and exposing them for what they really are - frauds.

Great articles, keep it up

Main Index

More Readers' Letters

Elliott Sound Products — About The Audio Pages

This site was initially created some time in late 1998, and has progressed from a single page (a somewhat shorter version of the bi-amping article) to what you see today. I have gradually built up the content, and the overall site "map" has changed several times as I have tried to incorporate all the new stuff in a reasonably sensible manner.

As the site continues to grow, you will see more changes, but I will always keep the user interface as simple as possible to maximise loading speed. This is one reason that you won't see fancy mapped graphics, frames or other frills that might make the site look really cool, but at the expense of download times.

The overall philosophy of the site has never changed - keep to the facts, stay away from the constant efforts of the subjectivist camp to ever "improve" on what they have (almost always with expensive "tweaks" whose performance cannot be measured, only heard by people with "finely tuned ears" :-). Music is to listen to. Recordings are rarely perfect, the concept of reproduction ever matching a live performance is a myth. Listen to the music, not the equipment.

This is not to say that good equipment is not something to strive for, since the enjoyment is greater when it sounds good. I love to experiment, and many of the designs are experimental - in some cases just to prove a point (the DoZ is a perfect example). Sometimes these experiments backfire (the DoZ is a perfect example!), and I get a whole bunch of e-mail telling me how great it sounds.

How much of the great sound is purely the result of the reader having built it himself/herself? I honestly have no idea, but it doesn't matter. If people can get double the enjoyment from building and then listening to equipment then so much the better. In the long run it is all about enjoyment; of music, of making something and of life.

May you all enjoy building my projects as much as I enjoy bringing them to you.

Images

I have been asked many times about the way I create the circuit diagrams (or schematics, if you insist), and over the time the pages have been running this has changed. I currently use Protel to draw the diagrams (although I have used other methods in some of the earlier drawings). These are simply captured and pasted into Paintbrush for some minor touch-ups, and the final image is then exported as a GIF file. This method is a little time consuming, but I have found that the images are very clear, and I get consistent results.

Articles & Projects

The content of all the articles and projects is entirely my own unless otherwise stated. This extends to the philosophy of the site itself, which is mine and mine alone. This (of course) does not mean that others will not have similar ideas (many do), nor that I automatically disagree with the opinions of others who might have a slightly different opinion on the same subject. I have been corrected many, many times - for anything from spelling mistakes to errors in diagrams (I have even managed to get a few electrolytics backwards - oops!), and various people have assisted with additional information on a number of occasions.

I do not (knowingly) steal the ideas, drawings or other content of others, and any information from others is reproduced with permission and full credit is given to the original author. Contributions are encouraged, as I am determined to make the best audio web site around, and I cannot do it alone.

There is a very small number of images on these pages that seem to be in the public domain, and I have used some of these where appropriate. If any reader out there sees their image on my pages and is offended that I purloined it, let me know and I will remove it.

Sometimes you see an image that is just too wonderful to ignore - the picture here falls into that category. It was sent to be by a friend, and I am sorry to say that I know not where it came from. I just loved it on sight!

I do not use (or condone the use of) spam (the web kind or the canned variety), so you will never get bulk e-mail or cans of fake "meat-like substance" from me for any reason, so that image is appropriate in it's own silly way :-)

I just wish I knew where it came from so I could thank its creator. Whoever you are - my thanks and apologies for "borrowing" this image.

Cheers,

Rod Elliott

Main Index

Contact the Author

Created 09 Aug 2000

Elliott Sound Products **Meet The Author**

I (Rod Elliott / aka "the author" / aka "rode") am based in Sydney (Australia), where I have lived for the majority of my life. I was in a small country town in my early childhood, and having revisited it not long ago, I am immensely pleased that I no longer live there. I have travelled extensively, both for work and pleasure, and this helps to balance one's view of the world. Australia is still my favourite place to live though :-)

I am a QbE (Qualified by Experience) engineer - I do have some formal qualifications, but they are relatively meaningless without experience and background - I consider them to be "unimportant". I have worked in the electronics, audio and of late the telecommunications industries for over 30 years, and have developed a number of products which in their time received critical acclaim. At this stage of my career (Definition: to rush or hurtle with much speed but little or no control) I have no real desire to re-enter the rush and bustle of commercial enterprise, but one cannot still a lifelong passion.

As a result, I have produced these pages which have stimulated some interest in the subject and hopefully introduce some new thinking and ideas into what is rapidly becoming a tired old brain (Oy! Who put that bit in? I never said that!).

In case anyone was wondering, I am not an "audiophile" in the true sense of the term. I enjoy music immensely, and quality reproduction is naturally very important to me. Will I go off and spend hours, days or weeks testing different tweaks, cables or boxes of sand to "damp vibrations" - no. I will spend a considerable amount of time experimenting with ideas or building something new to play with, and every so often I actually get a chance to just listen. This does not happen often enough, unfortunately, since I have a web page to feed and actual work to be done - these things take their toll on available time.

When I do get the chance to "just listen", I do what a lot of audiophiles should do. I listen to the music. The reproducing system must be open, clean and transparent enough so that I don't hear it, only what comes from it. The extraordinary amounts of money that some people spend on systems is beyond my means, but what I listen to is what I have built myself, and I figure that a considerable sum has been saved in doing so. This is the major reason for the existence of The Audio Pages - to encourage others to do the same.

Apart from anything else, there is great satisfaction in building equipment. It may not have the glamorous appearance and prestige name of some of the "finest equipment", but if it sounds great (to me and my friends - some of whom *are* audiophiles), then it is great. I don't need the brand names and fancy looks, I want good sound at a respectable price. This is something that is attainable for everyone who has the interest, and who decides to build their own equipment. Reading these pages won't hurt either ;-)

Anyway, enough drivel ...

History

For many years I had my own business, which was predominantly involved with musical instrument amps, PA systems and the like. During this period I also worked with a friend in a recording studio I designed and built,

called "Fly-By-Night Recordings". We did a fair bit of advertising work (bleccht) and recorded a bunch of bands. One of the recordings we made has recently been re-mastered and released on CD, so the quality must have been there.

I played in a few bands for a while (guitar, then bass), but soon realised that I was much better at the technical side than the musical. I still play for the fun of it, but alas, don't get as much time as I would like for that, either. I spent quite a while mixing for live bands, and toured Australia a few times - life on the road with a rock band has to be experienced to be believed!

I worked for a spell with (and helped to start) the SAE (School of Audio Engineering) - one of the most recognisable of the audio engineering schools, and was teaching students the basics of recording - how to mic a drum kit, mixing techniques and that sort of stuff. See the FAQ page for a bit more (and a useful link) on that topic. After that was a school that taught electronics (a privately run affair called The School of Electronics), where I taught the basics (and not so basics) of electronics, primarily analogue of course - digital was still in the future at the time. I still see some of my former students, and a great many of them are still involved in the electronics industry.

I then spent some years at a computer company, initially in the repair centre (this was when it was still worthwhile to repair computer boards), then into a Research and Development role (as its sole team member!) designing telecommunications products. I then moved to a telecommunications company, where after a while, I was back to teaching again - don't you just hate when that happens?

I have become my own boss again, so ESP is now a full time business - this is something I have been hankering for, and it has now happened - not without some trepidation of course, but I will be able to devote more time to the things I enjoy the most.

My System

My personal system has been through many changes (as you would expect), but in its present form it was fairly stable for about 4 years (up until September '99, when some major changes were introduced).

I am using a modified fairly standard type CD player (premium Burr-Brown opamps instead of the originals), and also a direct drive turntable with a fairly old but still magnificent moving coil pickup cartridge. There is the mandatory cassette tape machine (rarely used except for some of my old live recordings), and an FM tuner. A DVD player is a stable part of the system - just don't assume that because they can play CDs that you should use one for this. They do it, but not as well as the 'real thing' (this depends on the DVD player, of course - some are outstanding!)

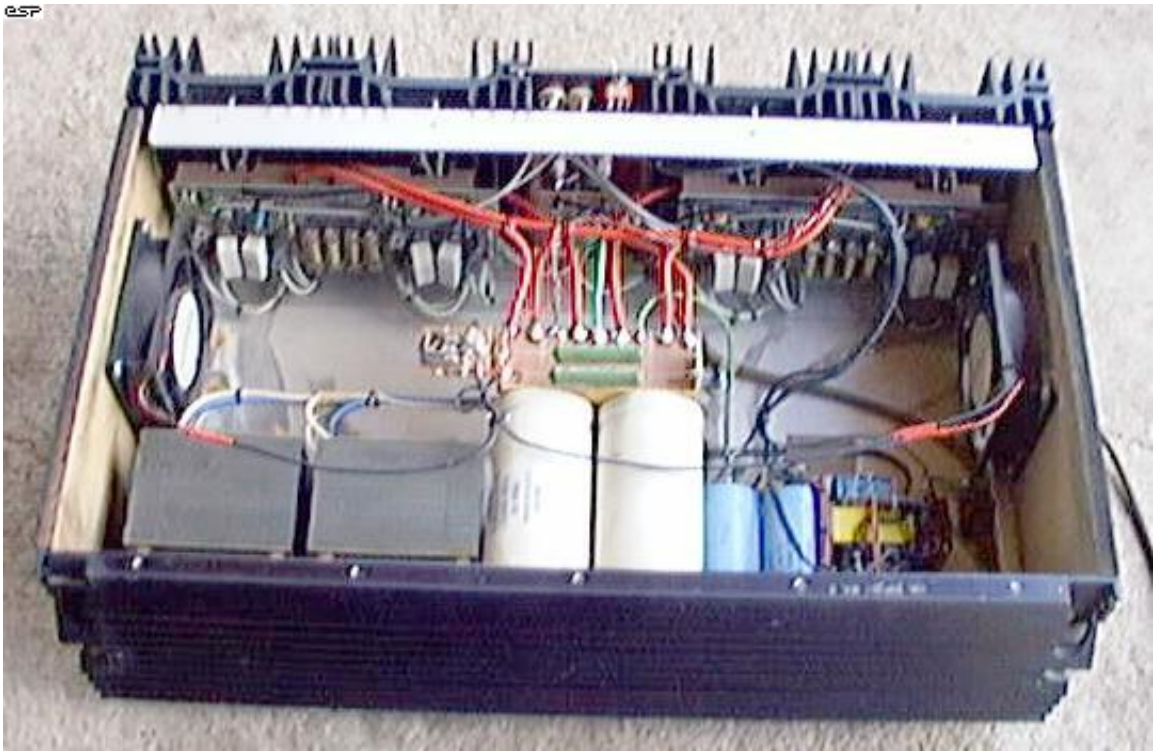
Nothing really special in the source department, but the preamp and control unit are of my own design, and the system has an inbuilt phase-coherent crossover network (at 300Hz and 3000Hz), feeding a triamped system using four of the 60W (P3A) amps (modified) described in the Project Pages. The Control Unit shown below does the following ...

- Provides tweeter amplification (20W per channel)
- Contains the 3-way Linkwitz Riley electronic crossover
- Houses the phono preamp (Fully shielded - bottom left corner)
- Signal shorting relays and DC protection relay for the tweeters
- Switched power for all power amplifiers in the system (controlled from the preamp)

I was using my valve preamp, but now use the Project 88 preamp, which I think actually sounds better! This feeds directly into the crossover. It also allows me to use the phono preamp (the preamp unit does not have its own), and now everything is there.

As you can see from the photo, almost all of the circuitry uses the PCBs I offer for sale (the power amp is an early prototype of the P3A amp board). If you want to know what the valve preamp looks like (because this is what I used to use) have a look at the VP103 page.

The main power amp is something else again. I had actually forgotten how much work I put into the thing, and the insides surprised me when I took it apart to replace the input and output connectors.



As you can see, the power supply is a fairly robust affair. The two large transformers (there are two little ones for the preamp too) are 200VA each (400VA total) and were specially made to my design quite a few years ago. The amps are as described in my projects pages (the P3A amp), but have been upgraded, using 200W TO3 power transistors and run a +/- 40V supply, with 18,000uF of capacitance per side. The smaller electrolytics you can see on the right were for the preamp, but are no longer used. The amp can happily supply 70W into 8 Ohms from each of the 4 amps, and will do about 80W peak on normal program material.

The fans are thermostatically controlled, and only ever run when we get one of the famous Australian hot days. Because the power amp runs a fairly high quiescent current (around 80mA per amplifier), it gets rather warm even when idle, and the safety margin for temperature was not large enough for my liking.

All input sources (including the turntable, using the phono preamp in the new control unit), now go through my P88 preamp before the crossover, and that has all the characteristics one would desire - dead quiet, excellent imaging, etc.

The speakers and control unit are described in the My New Speaker Project article. This tells just about anything that you might want to know about them, so I shall not ramble on here. All electronics in the control unit (crossover, tweeter amp and phono preamplifier) are as described in the various published projects.

To combat the lack of low bass (< 50Hz with the new boxes), I am using an EAS (Electronically Assisted Sub-woofer) of my own design (seeProject 48 for details) which has extended the bottom end dramatically. Powered by a 400W Class-D amplifier, I can get -3dB at 15Hz or less, but raised this to 25Hz to prevent excessive cone excursions caused by assymmetrical signal waveforms (these shift the effective DC operating point). There is very little that is audible (or recorded) below 25Hz, and not being a great fan of pipe organ music, I don't think that I am missing much in the frequency range. The sub is now equalised with the P84 Subwoofer Equaliser, and is flat to 20Hz in my listening room.

Test Gear

The photo below shows the bench mount gear. There is also the mandatory collection of hand held meters - both analogue and digital, capacitance and inductance, etc. (not shown, 'cos they are boring :-)

From left to right, you can see my distortion meter (the large black box with big meter), and on top of that is an arbitrary waveform generator, frequency counter and tone-burst generator perched at the summit. Next is an audio generator (now well over 25 years old, and still going strong), then my digital oscilloscope. My ancient transistor tester is next, with a function generator on top of that. At the top of that stack is my quasi-parametric equaliser (with its back rudely turned!). Next in the line is a dual tracking power supply, with a combination test

unit on top of it (audio generator, small power amp, monitor speaker, and triple power supply). At the top of that pile is a 10 band spectrum analyser. At the extreme right is my test amplifier, with 3-way variable crossover, phase correction and impedance control on two of the three amps. An old tuner (now retired and replaced by another no quite as old) sits atop that. (Sorry, that was boring too.)

The inevitable (for me) bits of "stuff" are occupying any remaining space. I don't know why, but any time I get a clear flat surface (regardless of size), it seems to be covered with things within minutes. I suspect that I may be the cause, but this cannot be proven :-)

Me, And Some of My Toys

This is not a particularly good photo, but I am reluctantly forced to accept that I do look more or less like the photo suggests. Given my location (Australia) the DownUnder sweat shirt is fairly appropriate.

A couple of my favourite toys are visible - the pantograph engraving machine in front, and radial arm saw behind. There are obviously others, but I won't bore you further with the details.

I hope you enjoy my offerings, and those of my growing list of contributors. I wish you happy reading.

Cheers, Rod

Main Index

Contact the Author

Last updated 04 May 2003 - New P88 preamp, minor reformat./ 02 Feb 2006, minor text changes

The Audio Pages

Elliott Sound Products **ESP Philosophy**

Copyright © 2004 - Rod Elliott (ESP)
Page Created 05 Jan 2004

Articles Index

Main Index

Contents

- 1. Introduction
- 2. What is Truth?
- 3. Truth Versus Belief
- 4. What is Science and Belief?
- 5. Perpetual Motion in a Different Form?
- 6. Definition of Proof
- 7. Conclusion
- 8. References

---

1. - Introduction

In as far as it is possible, the articles presented on the ESP site are aimed at one thing - the truth. To be more exact, it is really my version of the truth, since truth is neither absolute nor tangible. There will be areas where it is later proven that I am (or was) wrong - I accept and welcome this.

At various times, I have been berated because I do not 'trust my ears', or that I have claimed that a product is fraudulent without having tested it at all. Guilty as charged - there are simply too many products that should be tested, and in a lot of cases the

claims are so preposterous that testing is clearly unnecessary. While my ears remain a much used and relied upon test instrument, this is true only for my own tests - to be able to say anything with authority, I need hard evidence, not just add to the confusion by making unsubstantiated claims based on what I think I hear.

There are many products that claim *massive* improvements over and above the standard offerings, without offering one single shred of evidence to show that the claim has any basis in reality. While it is apparently perfectly acceptable for the vendors of such products to make their claims unhindered, for some reason it is not alright for me to debunk these claims without having listened to the product myself (indeed, testing is often claimed *not* to be needed, or should be avoided!). In many cases, it really is not necessary to do so - the claims made have no relationship with reality, and wasting any time at all on such products is pointless.

Most commonly, magic component claims are riddled with pseudo-science, gobbledygook 'technical' specifications or references to proven characteristics of materials that, while passably interesting, have nothing to do with that material's performance in a piece of electronic equipment. This is a popular technique to convince the non-technical reader that there is real science involved, when in fact nothing could be further from the truth.

I cannot change the way that people think, and I have no desire to even attempt to do so - I present my material as is, and by virtue of the fact that I wrote most of the material, it comes with my own prejudices built in. To some readers, this is unacceptable (mainly because I don't agree with them), and is seen as a perfectly good reason to launch an attack ... most commonly in forum and bulletin board sites, and rarely by anyone who has ever bothered to send me an e-mail to discuss their point of view (or offer some real data to prove their point). There have been several cases over the years where I *have* been challenged, real data *has* been offered, and I have done further testing or investigation and made changes to the affected page(s) as a result.

There are great differences in the hearing ability of different people. This changes from day to day, and is affected by mood, alcohol and other substances, minor illnesses and many other factors. That some people will hear things others cannot is a given, but no-one can hear sounds that are well below the general ambient noise floor, except under well controlled conditions and perhaps with specific sounds. A signal at 1dB SPL cannot be audible in the presence of another sound of similar frequency (range) at 110dB SPL, yet it is not uncommon to find claims that some individuals can hear sounds that would relate to -10dB SPL or less relative to the overall sound pressure. For example, if a 0dBV signal produces sound (via a power amplifier and loudspeaker) at 100dB SPL,

then -100dBV is 0dB SPL - the threshold of hearing. Anything below -100dBV will be inaudible in *absolute* terms, and completely inaudible relative to the overall SPL.

Please Note: It is very important that one thing be made perfectly clear. No individual person or specific group of people is referenced herein by inference or deed, although it is possible that some readers may think that there is an oblique reference to them, their behaviour or their beliefs. This is not the case - this article refers only to generalised examples of particular behaviour, and no specific or individual person or group is stated or implied. There is one exception, which follows ...

---

2. - What is Truth?

Krishnamurti said in 1929 (in a speech where he disbanded the Order of the Star - of which he was the head) ...

> **I maintain that Truth is a pathless land, and you cannot approach it by any path whatsoever, by any religion, by any sect. That is my point of view, and I adhere to that absolutely and unconditionally. Truth, being limitless, unconditioned, unapproachable by any path whatsoever, cannot be organised; nor should any organisation be formed to lead or to coerce people along any particular path. If you first understand that, then you will see how impossible it is to organise a belief. A belief is purely an individual matter, and you cannot and must not organise it. If you do, it becomes dead, crystallised; it becomes a creed, a sect, a religion, to be imposed on others. This is what everyone throughout the world is attempting to do. Truth is narrowed down and made a plaything for those who are weak, for those who are only momentarily discontented. Truth cannot be brought down, rather the individual must make the effort to ascend to it. You cannot bring the mountain-top to the valley. If you would attain to the mountain-top you must pass through the valley, climb the steeps, unafraid of the dangerous precipices.**
>
> Reproduced with the kind permission of the Krishnamurti Foundation of America [1]

This is such an eloquent description of truth that it almost certainly cannot be bettered. It is just as relevant to the pursuit of excellence in audio as anything else, although I'm sure Krishnamurti would have considered this to be rather trivial compared to the spiritual context he referred to. Trivial or not, the fact remains that the essential principles of which he spoke are just as valid in the field of audio as spirituality, and indeed anywhere that several competing versions of 'truth' exist side by side. This is rarely a peaceful coexistence - examples abound in the world around us, and we have seen the terrible destruction of life (and property) in the name of one truth or another.

We have also seen the almost religious fervour with which some people will defend their point of view or favourite piece of sound equipment. Likewise, we see attacks (sometimes very vicious) on those who do *not* believe in pseudo-science, magic components or rocks (special or otherwise). To criticise is seen as heresy by some, and pity he who has the gall and audacity to openly state that $10,000 speaker cables (for example) are not the path to Nirvana.

Such extravagances may make their owners think their system is something Heaven-sent, however it is completely unreasonable and unacceptable that they will attack someone else who claims that the cable had nothing to do with the sound, or for another to claim that such sound is impossible to achieve *without* the 'special' (or magic) cables/ components used. It is simply impossible for either party to win such an argument, and it is best for all if they could agree to disagree and leave it at that.

Much of what is at issue is based on the oft-repeated claim that there are many things that simply cannot be measured, and only our ears will be able to pick the subtle improvements offered by the product. There is no point asking anyone making that claim to prove it, nor is there any point trying to convince the person otherwise. It is their belief, and to them it is true.

However ... How much of their belief has come from elsewhere? To what extent is the belief something that has been formed after years of listening to music on different systems and taking detailed measurements, versus listening to the proclamations of the Black Knight 'gurus' whose belief system was built on an unknown premise? Does the cable vendor (for example) genuinely believe that his product is as good as he claims, or it the whole exercise a cynical ploy to separate the buyer from his hard earned cash? We simply do not know the answers to these questions, and further probing for the answer will only infuriate the persons involved. The net gain is zero - to all parties.

---

3. - Truth Versus Belief

The problem is that each of us has our own truth, and this is arrived at by many and various means. In some cases it is nothing more than a belief, and it must be accepted by others that this is valid for the person who believes. Although we may challenge the belief, it is rare (to the point of being virtually unheard of) for people to simply drop one belief system for another.

In other cases, our truth is the result of research, testing and measuring. Should we find no evidence that a particular type of cable or component makes an audible difference,

then that becomes our truth. This must also include pure analysis, without measurement or listening test. For example, if it is suggested that demagnetising plastic materials makes a difference to the sound of an audio system, there is little point in anyone with an analytical mind testing this. Since it is well known that no commonly used plastic can be magnetised, there is no point 'demagnetising' it. Simple analysis indicates that demagnetisation and a resultant improvement in sound quality are impossible, when the base material *cannot* be magnetised in the first place. Perhaps there is some other mechanism at work, but it is certainly not magnetism.

But, what of the beliefs of those who have demagnetised non-magnetic materials, and say that the difference is astonishing? Are they demented, or have they succumbed to the placebo effect? Could they be right? I can't answer this, and nor can anyone else. To me, they are the unknowing victims of a fraud, and have chosen to believe that their hard earned money was indeed well spent, and not wasted making the fraudster rich. It is important to understand that despite this, *their experience is real to them!*

Neither you nor I can convince them that they are wrong, any more than they will be able to convince me that the effects are real. If it cannot be measured (bearing in mind that the digital signal can be analysed byte-by-byte) then I shall remain unconvinced, since there is zero proof to support the contention. Billions of characters of electronic and conventional (paper) storage are devoted to arguments between the different 'camps', be it religion, politics, audio, cars, 'free' energy, conspiracy theorists etc., etc.

No-one wins these arguments - ***ever***. They are un-winnable in any forum, since no-one can convince anyone against their will. All that anyone can do is to explain in a reasoned and (hopefully) rational manner that *they* believe that the claims are false, giving sound reasoning and clear details of the facts as they are understood. In any case, the 'true believer' will not be swayed by any argument, however rational or well reasoned - and lest you may think that I am approaching this from my own perspective, this applies both ways.

A cable believer will be just as incapable of persuading me to change my mind as I will be incapable of changing his - we each have our own truth, and it is completely true - but only to the one person. I base my beliefs on what I can measure, since I believe that measuring equipment can resolve details that cannot be heard under any conditions. For example, although I know that I can measure the microphony of a cable (i.e. it's ability to act as a microphone when vibrated), I also know that the signal level is infinitesimally small when the same cable is loaded with the typical impedances of the source and load. Some claim that this is in fact audible, and to them, this may be true. It may also be true that they *think* they can hear it, and a double blind test (DBT) would show that they

can't. We cannot force anyone to perform a DBT, exacting measurements, or anything else against their will.

There is no answer - people will continue to believe what they want to believe, and all anyone can do is to present the facts as they know (or believe) them to be. One can do no more.

---

4. - What is Science and Belief?

Two excellent articles that explain the concepts of beliefs are "What is Science, anyway?" [2] and The Belief Engine [3]. These two articles are recommended reading, but there is a great deal of other information available from any search engine.

What is Science?

"Science is the systematic enterprise of gathering knowledge about the world, and organising and condensing that knowledge into testable laws and principles." This rather elegant description is from biologist E.O. Wilson's book "Consilience" [4]. That laws and principles will change over time is a given, since we are always learning. Very little science (none that I can think of) has grown from pure belief - the gathering of evidence is hard work.

In short, science is knowledge based - all scientific principles are based on what we *know* rather than what we may believe. Science is a process of assembling knowledge, correlating that with other knowledge, and testing the outcomes to determine a set of rules (or 'laws') that govern the interactions between materials and/or energy. Without the testing and verification, there is no science. If the tests are not repeatable by others, there is no science. Anything that cannot be verified or measured, relies on opinion or subjective reports (hearsay) falls into the following category ...

What is Belief?

Simply stated, a belief is based on something that we cannot prove, and requires *faith*. We do not 'believe' in the existence of air or water, for example, since we *know* that they exist. Some may *believe* that they can live on air or water with no other sustenance, and this belief will stay with them until they either die of malnutrition or come to their senses. Thus, it may be demonstrated that faith is not enough, and that we cannot live on air or water alone - such beliefs are not real to the majority of people. Other beliefs that are not so easily disproved are much harder to deal with. The longer a belief has been held, the harder it will be to convince the believer that s/he is mistaken.

What of God or magic components? These are beliefs - neither can be proven to the satisfaction of any observer, so there will be those who believe in God and those who do not. Some will believe that God is different from the (false) God of other believers, that S/He has different rules, is benevolent, fearsome, omnipotent, or just 'there'. Others (agnostics) may take the wait and see approach, which according to some believers guarantees the agnostics will suffer eternal damnation. There are no facts either way, so no-one can actually be right *or* wrong - they simply have different beliefs, and will not change readily.

Likewise in audio (and indeed many other pursuits) - some believe in magic components, others do not. The non-believers may be agnostic or full-bodied non-believers, but the end result is much the same. Although it is possible to prove that a great many of the magic component claims are false, this proof will only affect those who are already non-believers or agnostics. The believers will claim that the proof is bogus, just as the non-believers will claim that the reasons given that the magic *works* is bogus.

I like to think of the believers in magic components as 'naysayers', since (to me) they refuse to see reality. To them, I am the naysayer, since I refuse to believe that there is any magic in components, and I believe that all audible effects can be measured.

There are some that fall into a special category I call the 'Black Knights' - as in the Black Knight of Monty Python fame, who refuses to accept that he can no longer fight, even after he has lost both arms and legs. Some of the Black Knights of the audio world have been brutalised and scorned by many, and some observers might say with considerable justification. Their beliefs are so strongly held that no argument, no rational discussion of the facts, no possible reason, will sway them from their often ludicrous postulations. Rarely (if ever) does the Black Knight have a website with detailed diagrams, test results and measurements. Often (if they *do* exist), the data will be lacking certain critical information, without which it is essentially meaningless. There will nearly always be other people who agree with the Black Knights for reasons of their own, and who believe that the claims are reasonable and rational. They will leap to the defence of a downed Knight regardless of the rationality of the detractor's claims, and most often also without a shred of hard evidence to back up their position.

To me, this behaviour is denial - they deny that there is anything wrong with their claims, they deny that any proof to the contrary is valid, (regardless of the qualifications or reputation of the person providing the proof), and they will deny that there is anything to deny. The arguments used are usually ad-hoc, often (ever so slightly) off-topic, and almost always emotional rather than rational. In exactly the same way, a creationist will deny that the proof of evolution is valid, or a conspiracy theorist will deny that any proof

of man's moon-walk is valid.

To take it to an extreme, the Flat-Earth Society (and yes, it does exist) refuses to believe or accept that the photos taken from space prove anything (they will typically claim they are all faked), and despite the fact that not one of them has ever fallen off the edge of the world, it is still flat - regardless of the awful paradox that creates with air travel, the existence of the horizon, etc. These are the ultimate Black Knights - are they completely mad? I don't know, and I don't much care either, the truth be known.

Naturally, some of the claims made by (some of) the Black Knights will be valid and may be proved easily, and this makes it much harder for the layman to separate fact from fiction. Only by experimentation and education will people be able to see what is (or is likely to be) true, and what is based on faith or is complete fabrication. In some cases, a claim can easily appear to be completely rational and have scientific validity, but without evidence and repeatable test results it should be treated with suspicion.

---

5. - Perpetual Motion in a Different Form?

The proponents of 'overunity' or Perpetual Motion Machines (PMMs) have a similar philosophy of life to that of the magic component proponents and/ or vendors. Essentially, they feel that their discoveries are so important that *no-one* should overlook them, yet both groups will reveal nothing tangible about how their creations actually do what is claimed.

Any information that is provided requires that you must take it at face value, and only after a purchase will you discover the truth - namely that it does not work, or at least, does not work as claimed. These people seem to have a disconnect with reality, and believe that established science fails to see the importance of their work, is blind to the benefits, or even deliberately withholding the truth.

Having purchased the magic component or plans for the PMM, in many cases the buyer may convince himself that it works, for to do otherwise is to admit to having been duped.

The claims in both cases are very similar - you will get benefits that are way beyond what simple physics would indicate, and these people will protest that established science simply does not understand the principles involved. No amount of mathematical or measured performance proof will ever make a difference to their beliefs, and will be simply deemed wrong, inappropriate or misguided. They will vehemently deny that their own claims are wrong, inappropriate or misguided, and will reject any proffered proof out

of hand.

In many cases, the PMM brigade will defend their position with words alone, and would rather continue with endless debate than spend some time in the lab or workshop fabricating their invention (and thus proving the claims), and rendering all further words superfluous.

The magic component believers will do exactly the same thing, but with a twist. They do build equipment and test their inventions, but refuse to perform meaningful measurements, double-blind tests or anything else that would prove or disprove the claimed benefits once and for all.

Both groups are rather like the alchemists of old, and think that all they need is sufficient faith and purity of thought to achieve things that are in fact virtually impossible. The scientific approach is considered 'defective', because it does not embrace their beliefs.

There are many examples that will be brought up where science has been proven wrong in the past. Splitting the atom or exceeding the speed of sound are two classic examples, since everyone used to consider both to be impossible. Now that they have been shown to be perfectly possible, the same logic is applied to their own creations ... "*Since the scientists were proven wrong then, so shall they be in the future when my claims are examined properly.*" The flaw (from their perspective) is in current thinking, which disallows the proof that their PMM, cable or 'universal distortion and noise canceller' (etcetera) will actually do what they think it does.

Then there is the 'undiscovered law of physics' that can be brought to bear if a detractor is particularly persistent. While it is quite possible that there really is an undiscovered law, it will not negate the current and well proven concepts of conservation of energy, thermodynamics, or any other basic principles. Any such new law (should it exist in the future) will add greater understanding of existing principles rather than make the currently used physical laws obsolete.

To these people, Einstein's special law of relativity should read ...

$$E = mc^2 \pm 3dB$$

... and the uncertainly factor allows them the latitude needed to enable their process to function as claimed. Sadly for them, this is not the case.

In a similar vein, there are many who believe that everything is black or white, true or

false. There is no grey (which translates to uncertainty). In fact, the grey area is usually where the real truth is to be found, this is the area that should be examined most closely, for in doing so one will learn a great deal about himself, and potentially a great deal about the way the world works.

In this context, there used to be a claim that "you can't have trees and grass too" - this was the common cry of early sheep farmers in Australia. This belief resulted in the wholesale removal of trees, and the result is now decimation of the land - massive soil erosion, rising salt tables that threaten (or make useless) more and more land each year. The claim in itself is true, but without examining the grey area, the actual *reasons* were not seen, and trees were simply cut down.

So, what are the reasons? I present one possible reason that is potentially quite simple, although I *do not* claim this to be the final answer - just one of many possibilities that will only be seen when the grey area is explored ...

- Sheep are cold climate animals.
- Australia (especially inland) gets very hot.
- The sheep will head for the coolest areas they can find. Under ... trees.
- While there, in the relative cool, they will eat!
- Before long, there will be no grass under the trees.

Now, it is not *because* of the trees there is no grass underneath, but because the sheep have eaten it all while trying to keep cool. In addition, sheep are hard hoofed animals, and Australian native grasses evolved with only the native mammals - none of which is hoofed!

Removing the trees was obviously a big mistake, and the real reason for no grass beneath the trees was neglected in the rush to 'solve' the 'problem'. In fact, trees may allow grass to grow better, especially in the areas shaded from the morning sun - I can see this in my own backyard! I can likewise lament this when I have to mow the very thick grass thus shaded, which grows like crazy during summer. In the areas that get full morning sun, the grass growth is much more subdued.

Ok, but what on earth does this have to do with electronics, audio, PMMs or anything else?

Good question, simple answer, but with complex understanding. The grey area of all things is that which is unknown. There is no immediate right or wrong, true or false, real or unreal. To get to the answers is difficult, time consuming, and in some cases

expensive. It requires one to think, long and hard, to discuss and to examine and reject theories. It is not the easy path.

On the other hand, to accept someone else's theory (e.g. "Oh, don't even *try* to understand it - it's magic!") is easy, but you then fall into the very trap that Krishnamurti warned against. 'Truth' is packaged, sanitised and instutionalised, you are warned that you must never question that which is 'written in stone'. This applies to politics, religion, finance, hi-fi and almost anything else you can think of. To accept the 'word' is the easy way - you no longer *need* to question, examine or experiment, or to search further for enlightenment. You have been told - it must be true because so many others also believe, and anyone who questions this truth is an heretic.

This principle was used to great effect by the Spanish Inquisition, and is used today to push the current fiscal policy, someone's belief in 'world order', and at the trivial end of the scale, to attempt to silence those who criticise the high end cable vendors.

A new belief can come about so easily that it is scary ...

> Someone thinks they hear a difference between two components. At first, they may be unsure, so will ask someone else. The second listener may say "I do believe you're right - I'm *sure*I hear a difference too." At this stage, they may (or may not) actually hear a difference, and it is more likely that they have fallen victim to the 'experimenter expectancy' effect.
>
> Along comes another - he may be unsure, but it's two against one now, so he may well bow to the pressure and *think*that he too hears a difference (this is called self-delusion, and is far more common that most people realise). Anyone who comes along after that either goes with the flow, or admits he has 'tin ears' and is ostracised from the new clique.
>
> Advance a few years, and this vague belief has become the truth!

Those who refuse to believe this (new) truth are naysayers, have tin ears and are deluded! **Of course** they are! The truth has been laid before them, and they refuse to see it. They are blind to the truth, and insist on trying to explain away the magic with mere science! By now, the clique has grown to cult status, it brandishes its truth as fact, and the heretics who refuse to believe are ostracised, categorised, criticised, and put aside - until they come to their senses. This is identical to the old tale of "The Emperor's New Clothes" [5], where everyone claimed to behold fine garments that never existed.

Does any of this sound familiar?

---

6. - Definition of Proof

Despite the reservations some may have, proof is not difficult to explain. Simply stated ...

> Proof may be defined as an experiment or test result whose outcome is demonstrably different from that predicted by random variations as may be found in similar devices or processes (i.e. the results must be statistically significant). To be valid, the experiment or test procedure mustbe repeatable, using equipment that need not be identical but has similar specifications, is not modified in any way that will produce the outcome independently of the characteristics of the device under test, and does not rely on guesswork in any form.

From Dictionary.com [6] we get the following ...

1. The evidence or argument that compels the mind to accept an assertion as true.
2. a The validation of a proposition by application of specified rules, as of induction or deduction, to assumptions, axioms, and sequentially derived conclusions.
   b. A statement or argument used in such a validation.
3. a. Convincing or persuasive demonstration: (*was asked for proof of his identity; an employment history that was proof of her dependability*.)
   b. The state of being convinced or persuaded by consideration of evidence.
4 Determination of the quality of something by testing; trial: put one's beliefs to the proof.
5 Law. The result or effect of evidence; the establishment or denial of a fact by evidence.

In other words, if anyone takes a measurement that shows a particular behaviour (positive or negative) with a particular component, then anyone else with similar equipment should be able to duplicate the results. Proof of any behavioural characteristic does not imply anything else. For example, if I were to prove that certain component leads were magnetic, then that (and *only* that) is relevant. If I then postulate that this degrades the sound quality, there is no proof of this, merely speculation or guesswork based on the original finding. It is necessary to then *prove* that there is a degradation, and if I cannot (or will not) do so then my postulation remains an opinion only - it is not a fact, it has not been proven, and may or may not be true.

Most people who attempt any form of proof will make mistakes. They may overlook something, or be unaware of a characteristic of one or more of the components. Their

proof will be flawed, and will almost certainly be challenged.

Those who take a scientific approach will (usually) re-examine the new data, and may decide that it is irrelevant or important. If the latter, they will usually look more closely at the claim and their proof, and often find that they were indeed wrong. They may well be miffed (I have been on several occasions), but will generally publish a correction once the new data is correlated and the experimental processes are repeatable.

By contrast, there are those who rely on postulation (or proclamation) alone, and do not even attempt to prove their claims. Given the lack of evidence the claims must therefore be viewed with suspicion, since without any offered proof, there are no facts to work with. Emotive comments, unsubstantiated 'evidence', and hearsay don't amount to a hill of beans in the real world, and any complaint that modern equipment is not sensitive enough to measure the difference is simply bollocks.

"I heard a difference" is not evidence, merely a claim or hearsay, unless the test was performed using a properly conducted Double-Blind Test methodology, with a statistically significant outcome. The majority of magic component vendors and proponents eschew the DBT, and will use all manner of (usually ridiculous) arguments to support their position, such as "the DBT equipment degrades the sound so much that the test is meaningless" or "DB Testing is too stressful, so you never get the correct result". This is simply rubbish, and should be seen in every case as a priori evidence that the claims made are grossly exaggerated or simply balderdash.

A proposition is knowable 'a priori' if it is knowable independently of experience [7]

In some cases, it is entirely possible that a test methodology really does not exist to prove or disprove the claim, but the proponents of magic components will never, *ever* attempt to devise a method to prove that what they say is true, for to do so would almost certainly amount to shooting themselves in the foot. They know that they are talking through their hats much of the time, but they certainly don't want you to know that - especially if they want you to buy their product.

In some cases, the vendor may genuinely believe that his magic component works. He may be passionate about it to the point of being fanatical, and have such faith in the benefits that he is blind to all reason. He *cannot* accept any proof that the device simply doesn't work. The only 'proof' needed is letters of commendation (no, not condemnation) from users or reviewers, but there will be no technical information, no graphs or charts showing the before and after effects. No test equipment will ever be brought to bear, for this may shake the faith when it is demonstrated that no change can be found. It is

important to understand that this is still technically fraud, regardless of the beliefs of the vendor. A device must actually perform the described functions, and this should be based on evidence, not hearsay.

---

7. - Conclusion

So, to summarise the general philosophy of The Audio Pages and myself, I can safely state that ...

- I do *not* believe in magic components, fantastic 'capabilities' of certain interconnects, speaker or mains cables, magic lacquers, overunity or perpetual motion machines, or other amazing claims that are made regularly on the Web.
- If a component makes an audible difference, then I believe that it will be possible to measure the difference between the standard and improved component. It is unlikely (IMO) that any non-measurable differences are audible, but it is possible that a measurement technique may not have been thought of to quantify some subjective differences. This is an area that I am looking at closely in a number of areas (mainly loudspeakers).
- There is considerable evidence that a great many measurable differences in components exist that are not audible, but virtually nothing to support the opposing viewpoint.
- It is highly unlikely that there is a flaw in current science or thinking that prevents PMMs or 'magic components' from working. Note that it is the flaw in *thinking* that causes the problem - not the 'undiscovered mechanism', which is presumably somehow obscured by current thinking, thus preventing its discovery by the many followers of such theories. Hmmm. I'm not sure that **I** even follow that one
- The opinions expressed on these pages are primarily my own, and although some articles have been contributed, the content is substantially in agreement with my own ideas and beliefs.
- Contrary to what some people may claim, I do not automatically denounce the theories of others when they do not agree with my own. Indeed, there have been cases where I set out to prove that something made no difference, only to find that it is important, and does have audible consequences (time alignment, for example).
- As a matter of course, I keep an open mind on a great deal of material, and just

because I cannot disprove it does not make it 'wrong'. There are, however, some claims that are just too outrageous or stupid to even consider, and it is no more being 'closed minded' to scoff at such claims than it is to scoff at the Flat-Earth Society's beliefs.

There are a great many things that I don't know, and as I grow older and (hopefully) wiser, I realise that the more I know, the more there remains to know - this will no doubt continue until the day I die. The primary aim of the ESP site is to educate and enlighten - and I have had so many e-mails of thanks that I can only conclude that a lot of people have found benefit and knowledge from the articles and projects.

This was always my aim, and it is hoped that the majority of readers will appreciate and welcome the information presented. For those who think that the site is biased and does not fit with their beliefs, then I can only suggest that you go elsewhere - I will no more change my beliefs than you will yours. Argument is futile, wastes a great deal of time, and rarely if ever achieves anything worthwhile.

Just because you disagree with one or more of the things said on the site does not mean that nothing is valid. All are welcome to browse, and to join and participate in the ESP Forum (whilst abiding by the rules). The greater the participation, the greater benefit to all who visit - this is not a place for 'I win, you lose' arguments - it is intended to be a place where everyone can win.

It must be remembered at all times that *any claim that seems too good to be true, almost certainly is!* There is no appeal against the laws of physics, and any claim to the contrary is false, and should be an instant warning that something is seriously wrong. While it is more than possible that there are still things that have not been discovered, they are unlikely to be of such massive importance that our listening habits will be forever changed - i.e. they will be evolutionary rather than revolutionary. Any such discoveries are far more likely to be made by well equipped laboratories than backyard tinkerers or 'magic component' suppliers.

---

8. - References

1 Krishnamurti Foundation of America

2 What is Science, anyway? - Robert L. Park (web site)

3 The Belief Engine - James Alcock (web site)

4 Consilience - Edward O. Wilson, Random House, Reprint edition (March 30, 1999) ISBN: 067976867X

5 The Emperor's New Clothes - and other tales of Aarne-Thompson type 1620 edited by D. L. Ashliman (web site)

6 The Internet Encyclopaedia of Philosophy (web site)

7 Dictionary.com (web site)

Articles Index

Main Index

Copyright Notice. This article, including but not limited to all text and diagrams, is the intellectual property of Rod Elliott, and is Copyright © 2004 except where noted. Reproduction or re-publication by any means whatsoever, whether electronic, mechanical or electro- mechanical, is strictly prohibited under International Copyright laws. The author (Rod Elliott) grants the reader the right to use this information for personal use only, and further allows that one (1) copy may be made for reference. Commercial use is prohibited without express written authorisation from Rod Elliott.

Page created 06 Feb 2004

POWERED BY html GEAR

Get This Gear

Elliott Sound Products

Name:

E-Mail:

Homepage Title:

Homepage URL:

Referred By:

Search Engine

Location:

Comments:

Continue

Lycos Yellow Pages

find a business near you.

Get A Blog Now!

html GEAR ADD-ONS TO ENHANCE YOUR SITE

help

HTML GEAR
- Home
- Gear Manager
- Live Examples
- Pro Live Examples

PROFESSIONAL GEAR
- Event Gear
- Feedback Gear
- Guest Gear
- Headline Gear
- Link Gear
- Poll Gear
- Text Gear

BUSINESS SERVICES
- Credit Card Processing
- E-Mail Publishing

PERSONAL GEAR
- Event Gear
- Feedback Gear
- Guest Gear
- Headline Gear
- Link Gear
- Poll Gear
- Text Gear

GOODBYE ADS!

Our NEW Professional Gears offer MORE features, MORE customization, and NO ADS or branding! Sign up now and add as many Gears to your site as you'd like. Packages start at just $5 a month!
Sign up now! | Learn more

FREE GEAR

Headline Gear

Make news! Keep your site fresh and informative by displaying headline on the topic your choice.

- Pick topics to match your site
- Increase traffic with fresh content
- No HTML skill required

GET GEAR SEE SPECS

PROFESSIONAL GEAR

Professional Feedback Gear

Keep your customers satisfied! Give them a direct line for comments and suggestions.

- Keeps your e-mail address private
- Auto-response option for instant reply
- No ads or HTML Gear branding

Professional Gear plans start at only $5.00 per month!

GET GEAR SEE SPECS

NEWS

Search With Hotbot!
Add a Hotbot Search to your Website.
It's quick and easy! Get Searching!

RECYCLE-A-BIKE

Recycle-A-Bicycle is a non-profit organization dedicated to youth development in New York City

» Lycos.com © Copyright 2006, Lycos, Inc. Lycos is a registered trademark of Lycos, Inc. All Rights Reserved.

About Lycos | Help | Jobs | Advertise

Your use of this website constitutes acceptance of the Lycos **Privacy Policy** and Terms & Conditions

Hosted by tripod

Search: Tripod Web by LYCOS Go!

POWERED BY html GEAR Guest Gear Get This Gear

Elliott Sound Products - The Audio Pages

Next Page »

Sign this Guestbook | Manage your Guestbook

Page 1 of 168

Go To Page: Go

Sunday 05/28/2006 4:26:05pm

Name: Zaman

E-Mail: vmx786@yahoo.com

Homepage Title: n/a

Homepage URL: http://n/a

Referred By: Search Engine

Location: Bangladesh

Comments: Fantastic! Dynamic!

Sunday 05/28/2006 4:05:07pm

Name: Derick Markwell

E-Mail:

Homepage Title: ICON Engineering

Homepage URL: http://www.iconeng.com.au

Referred By: Forum

Location: Perth

Comments: Great articles on your website re active xovers.
Maybe well do business when I build mine.

Wednesday 05/10/2006 0:09:42am

Name: rodel

E-Mail: roddeasis@yahoo.com

Homepage Title:

Homepage URL:

Referred By: Friend

Location: Philippines

Comments: Amazing website. Hopefully PCBs will be available in our country.

Tuesday 05/09/2006 2:55:32pm

Name: Michael Menina

E-Mail: mike_menina@hotmail.com

Homepage Title:

Homepage URL:

Referred By: Forum

Location: Laguna, Philippines

Comments: Great site!!!Im planning to purchase P3A PCB from you soon.Still checking if the transistor (and some alternatives) is available here in our country.Keep up the good work!

Sunday 05/07/2006 7:07:46am

Name: farshid

E-Mail: soshianet47@yahoo.com

Homepage Title:

Homepage URL:

Referred By: Search Engine

Location:

Comments:

Next Page »

Welcome, Guest
Webring Home - Help/Support - Sign In

The Audiophile Webring

Join this Ring

Instructions

- **Step 1**
  Fill Out Ring Application
- **Step 2**
  Review & Submit Application

New to WebRing?
Read important information about what joining a WebRing Ring means here.

Step 1: Fill Out Ring Application

Enter the website address, title and description of the site you would like to add to this Ring and click **Next**. You should click on the **Help** link above to review what joining a Ring involves.
Not sure about "navigation code"? Click here for more information.

Enter Your Site Address (URL)

http://

Note: If accepted, you must place the Ring's navigation bar on this website.

Title Your Site (50 characters or less)

Describe Your Site (400 characters or less)

We need some additional information to make it easier for you to manage your WebRing site information. You appear to be **new to WebRing**, or you are not logged in.

If you're an existing member, Log In now (Remember, you only need **one** User ID even if you want to join more than one Ring).

Don't have a WebRing User ID yet? - No problem, signing up for your User ID is fast, easy and **FREE**. We collect almost no personal information, the User ID is to help you manage your WebRing sites easily and securely. You only need to do this once, and then you can join as many Rings as you'd like using the same User ID. Then when you want to manage anything to do with any site you'll have them all in one convenient location! And don't worry about security, all the information

you enter below will be kept confidential by WebRing and **not** passed on to the Ring's manager (except the User ID of course, so they can communicate with you!).

Required fields:

WebRing USER ID:		Choose an ID using letters, digits or underscores '_' only (32 maximum).
Password:		Choose a password that is hard to guess. Capitalization matters!
Password (repeat):		Now type it again so we can verify it.
Email:	1	Enter the email address where WebRing **correspondence** should be sent. PLEASE ensure your email address is valid. We do **not** share your address with any individuals and we do **not** share your email address with third party companies, and we **only** use your email for WebRing operations.
Retype email:		
Birthday:	[select one] ,	Enter your birth date.

Optional fields:

If you forget your password, we can use the security question you enter below to verify your identity.

Question: Enter a simple question that others cannot answer easily.

Answer: Enter an answer that you can remember ***easily***.

Website host?: ☐ Would you like to learn more about letting SimpleNet host your website? Check here and we'll have them contact you.

By submitting your registration information, you indicate that you agree to the Terms of Service and have read and understand the Privacy Policy. Your submission of this form will constitute your consent to the collection and use of this information by WebRing and its affiliates.

Next Cancel

Copyright © 2001-2006 WebRing, Inc. All rights reserved. Terms of Service - Help
Notice: We collect personal information on this site.
To learn more about how we use your information, see our Privacy Policy
WebRing server hosting provided by SimpleNet.
WebRing Logo design by Sean McCullough.

Welcome, Guest
Webring Home - Help/Support - Sign In

The Audiophile Webring

Ring Manager: jazsaxman RingMaster's site

Join This Ring

Home > Entertainment & Arts > Audio and Visual Equipment

The Audiophile Webring is for anyone with an interest in high-end home audio and home theatre equipment. It allows you to show what your system is comprised of, and the music or movies you enjoy listening to and watching. Sites with audiophile links and commercial audiophile products or audiophile magazine sites are welcome.

Ads by Google

Audio Racks Super Sale
Get Free Shipping & Discount Prices on Audio Racks. Many Styles & Sizes
www.Audio-Video-Furniture.com

Stereo Cartridge Store
Buy Top-Brand Phono Cartridges Now. Variety of Models. Secure Site!
www.NeedleDoctor.com

Audiophile Speakers
Shop Fast. Buy Smart.
Shopzilla.com

M Audio Audiophile Usb
50+ PCI & USB Sound Cards Read Sound Card Reviews.
www.NexTag.com/Sound-Cards

Ring Information

Ring Stats

- Created 06/09/1998
- 83 active site(s)
- 44 page views today
- 202512 total page views
- details

Sites in Ring

< Prev 10 | Next 10 >

Search for [ ] in [This Ring] [All Rings]

Sponsor

Audio Racks Super Sale
Get Free Shipping & Discount Prices on Audio Racks. Many Styles & Sizes - www.Audio-Video-Furniture.com- Ad by Google

Ring Site(s)

Jarrell Audio Concepts : Home Audio and Home Theater
Huge resource of home audio and home theater reviews, forums, classifieds, how to guides, shopping and more. For the audio newbie to the audiophile.

The Audiophile Webring Home Page
The is the Audiophile Webring Home Page.

AudioOptions
Various audio leads Connect Bang & Olufsen to ipod DVD TV and more

Legendary Classic Audio Gear
Dedicated to classic audiogear. Stats, images, reviews, and much more. Hundreds of pages, thousands of specifications!

OMAK valve amplifiers and audio transformers
Manufacturer of high-end valve amplifiers and custom-

Ads by Google

Audio Racks Super Sale
Get Free Shipping & Discount Prices on Audio Racks. Many Styles & Sizes
www.Audio-Video-Furniture.com

Stereo Cartridge Store
Buy Top-Brand Phono Cartridges Now. Variety of Models. Secure Site!
www.NeedleDoctor.com

Audiophile Speakers
Shop Fast. Buy Smart.
Shopzilla.com

M Audio Audiophile Usb
50+ PCI & USB Sound Cards Read Sound Card Reviews.
www.NexTag.com/Sound-Cards

M-Audio sale at JRR Shop
Shop Our Hot Monthly Specials Today There's No Better Deal & Free Ship!
www.jrrshop.com

designed ultra-wideband audio transformers. Audiophile and DIY enthusiasts support.

Repair Service for High-End and Audiophile Equipment
We are a repair service for high-end audio equipment both tube and solid state. We repair all the classic models From Marantz, Fisher, Pioneer, Sansui, McIntosh, Scott,Dynaco Eico, as well as more recent tube, solid state and digital audio products . Complete information on our website. Pleae stop by.

Quadpages
How to rebuild and upgrade Quad II valve amplifiers

AudioNova - The W. W. Audio Resource!
AudioNova DIY Magazine - This Swedish online magazine is a source of information for the hands-on Audio hardware enthusiast. AudioNova W.W. Audio Links - The bigger the web gets the harder it gets to find what you want. Audio Nova makes it easy with 8000 links to audio manufacturers, forums, magazines and journals, DIY- sites, vintage, tube sites etc.

WoodPaW Furniture
This is the HomePage of WoodPaW Furniture, situated in the Capital of Canada, Ottawa. Among his many talents, this woodshop's owner has a love for music and unique wooden accessories.

Moon Audio
Custom Cables Modifications Consulting

M-Audio Products on Sale
10% Discount on Entire Order Money Back Guarantee and Free S/H!
www.DigitalProAudio.com

M-Audio Gear On Sale
Free Shipping , Huge Selection Low Price Guarantee, Make an Offer.
www.proaudioland.com

< Prev 10 | **Next 10 >**
Search for [ ] in [This Ring] [All Rings]

Ads by Google

Audio Racks Super Sale Get Free Shipping & Discount Prices on Audio Racks. Many Styles & Sizes	**Stereo Cartridge Store** Buy Top-Brand Phono Cartridges Now. Variety of Models. Secure Site!	**Audiophile Speakers** Shop Fast. Buy Smart.	**M Audio Audiophile Usb** 50+ PCI & USB Sound Cards Read Sound Card Reviews.

Copyright © 2001-2006 WebRing, Inc. All rights reserved. Terms of Service - Help
Notice: We collect personal information on this site.
To learn more about how we use your information, see our Privacy Policy
WebRing server hosting provided by SimpleNet.
WebRing Logo design by Sean McCullough.

Free Web Site - Free Web Space and Site Hosting - Web Hosting - Internet Store and Ecommerce Solution Provider - High Speed Internet

50MEGS LOGIN
Get a FREE Website
Click Here!

Click Here NOW!
NetZero Voice
UNLIMITED FREE long distance
CLICK HERE

Search the Web
Search
powered by YAHOO! search
close

Purveyors of Fine Audio and Video Products

Home Page

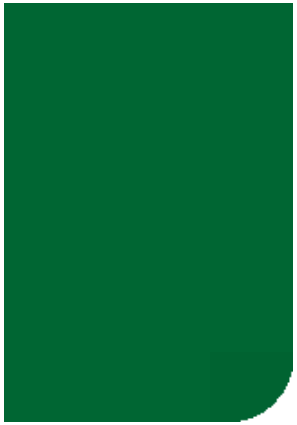
Photo Page

Contact Page

Welcome to The Speaker Shop on line. This site is a link to our sister sites with more info and pricing.

Home and Commercial Sound and Video Systems Designed, Installed and Serviced by Professionals..

Stop in to see fine audio systems coupled with the worlds finest video displays.
Over six theatre systems on display at all times.



Visit our Sister Site Here!

Free Satellite TV
Satellite Radio
Broadband Internet
Dial-up Internet
VOIP Phone
Enter
Cellular Services
Web Services
Computer Tips
Audio/Video Equipment
JAC Mall
JAC
Jarrell Audio Concepts
©2006-2008. All Rights Reserved
Privacy Webmaster

00828239

Welcome, Guest
Webring Home - Help/Support - Sign In

How would you like to get free web traffic to YOUR website?

Let WebRing do your web advertising. We're a unique internet service that helps people find your web site.

With a **WebRing** membership you can generate more web traffic to your website and **increase** your search engine ranking. It's quick, easy and FREE to become a member.

Just **Sign Up** and then **Join a Ring**. Search from almost 50,000 Rings in our directory to find **your** niche.

And if you can't find the right Ring for your website, then **create** your own ring and develop your own online community.

Because wouldn't you like to share your website with others?

Use our WebRing internet service to find great web sites, share ideas with peers and participate in topical communities. All for FREE. Get Started Today!

WebRing: We Bring the Internet *Together!*

WebRing News

- Be a Supporting Member
  - Surf ad-free!
  - Create a WebRing homepage!
  - Ring/Site Promotion
- News 6 August, 2005

Special Feature

Hurricane Katrina Relief

Featured Rings

Spring is for hobby & recreation. Make the cut with Fine Woodcarving, or find a dinosaur in The Paleo Ring. How about outdoors? Soar with the Falconry Ring, or ride a Quarter Horse. May is also Strawberry month, Egg month, Duckling month, Photo month, Bike month, and Physical Fitness month.

Hot Rings

Search WebRing

Browse Our Directory of Rings

Business & Finance
Investments, Industry

Computers & Internet
Blogs, Internet, User Groups

Cultures & Community
Issues, Seniors, LGB

Entertainment & Arts
Movies, TV, Books

Family & Home
Genealogy, Parenting, Gardening

Games
Computer, RPG

Government & Politics
Politics, Military, Parties

Health & Wellness
Fitness, Support

Hobbies & Crafts
Antiques, Collecting

Music
Rock & Pop, Rap, Country

Recreation & Sports
Travel, Cars, Wrestling

Regional
U.S. States, Countries

Religion & Beliefs
Christian, Islam, Jewish, Pagan

Schools & Education
High Schools, Colleges

Science
Astronomy, Biology, Pets

Relationships & Romance
Singles, Relationships

Browse Our Directory of Forums

Business & Finance
Investments, Industry

Computers & Internet
Internet, User Groups

Cultures & Community
Issues, Seniors, LGB

Entertainment & Arts
Movies, TV, Books

Family & Home
Genealogy, Parenting, Gardening

Games
Computer, RPG

Government & Politics

Hobbies & Crafts
Antiques, Collecting

Music
Rock & Pop, Rap, Country

Recreation & Sports
Travel, Cars, Wrestling

Regional
U.S. States, Countries

Religion & Beliefs
Christian, Islam, Jewish, Pagan

Schools & Education
High Schools, Colleges

Science

Events/Holidays

May 19th
19th Circus Day.
20th Flower Day.
22nd Skyscraper Day.
23rd World Turtle Day.

System Forums

Manager's Forum
- For managers to discuss issues, get help

General Member Forum
- For all users to discuss issues, get help

Proposed Enhancements
- Check here to review/ propose/comment on WebRing enhancements

Technical Support/Help
- Need help? You might try here

Top Forums

- Natural Cures They Don't Want ...
- The Guestbook Web Ring
- THE JAILHOUSE LAWYER'S WEBRING
- Let Freedom Ring
- Win Without War Webring
- .~. CrimsonEldritchs-[Darknes...
- FREE Advanced Health Plan
- Ultimate Photography WebRing
- Meditating Buddhists
- Mini-Dachs

more...

Hot Forums

· THE JAILHOUSE LAWYER'S WEBRING
· FREE Advanced Health Plan
· Win Without War Webring
· Let Freedom Ring
· THE JAILHOUSE LAWYER'S WEBPAGE...
· The AnnyBelle Foundation
· Aid for Animals
· RingMasters of WebRing.com
· Forums & Chat Ring
· Natural Cures They Don't Want ...
more...

Top NEW Rings

· German World War II Stories
· Comparative Tarot
· Leonbergers - Our Global Ties
· Kingdom Links
· Women Business Owners
· Classified Ads
· Make Money Online
· The Cross Of Love
· The Earth Peace Network
· online games, mmorpg, war game...
more...

Growing Rings

· WomenBloggers
· The Real Estate Ring
· Comparative Tarot
· RPG Ring
· YTB Travel Websites
· CHEERLEADING Webring
· Immigration and Border Issues
· Role Playing Games Webring
· Mormon Sites Pro-LDS
· A Mind, Body & Spirit

Politics, Military, Parties

Health & Wellness

Fitness, Support

Astronomy, Biology, Pets

Relationships & Romance

Singles, Relationships

· Advanced Scientific Health Web...
· FREE Advanced Health Plan
· Natural Cures They Don't Want ...
· Let Freedom Ring
· Win Without War Webring
· THE JAILHOUSE LAWYER'S WEBRING
· Mini-Dachs
· .~. CrimsonEldritchs-[Darknes...
· FREE Classifieds
· The Guestbook Web Ring
more...

Search for a Free Domain

WebRing is hosted by SimpleNet. You should be, too!

GO!

Search for the following tlds: .com, .net, .org, .info, .biz, & .us

Already have a domain or site?

Move it to SimpleNet

Ring

more...

Top Rings

· How to Draw Anime WebRing
· Gencturk.net Güzeller
· The Nikki Cox Webring
· AQ Frogzard Ring
· Advanced Scientific Health Web...
· Jane Austen Fan Fiction
· Little Rascals/Our Gang
· Natural Cures They Don't Want ...
· Movie Quotes
· THE JAILHOUSE LAWYER'S WEBRING

more...

WebRing Numbers

September 2005:
4.7 million unique visitors
21 million hits
46,800 Rings
1,004,000 **active** Sites
565,000+ registered **active** members
9,600+ contributing members .
View system traffic here.

View WebRing's Adult Content Policy

Copyright © 2001-2006 WebRing, Inc. All rights reserved. Terms of Service - Help
Notice: We collect personal information on this site.
To learn more about how we use your information, see our Privacy Policy

WebRing server hosting provided by SimpleNet.
WebRing Logo design by Sean McCullough.

Welcome, Guest
Webring Home - Help/Support - Sign In

POWERED BY SIMPLENET WEB HOSTING

WebRing - The *best* place on the web...

WebRing Links

- WebRing Directory
- What is WebRing?
- Join
- Ring Management
- Help
- Contact Us

You're viewing the complete list of WebRing Rings that member rode_sound has for the website you just visited. Please try give these a try.

The Audiophile Webring
[Join Now | Ring Hub | Random | **<< Prev** | **Next >>**]

Australian Electronics Web Ring
[Join Ring | List All Sites | Random | **<< Prev** | **Next >>**]

The Classic Stereo Equipment Web Ring
[Join Now | Ring Hub | Random | **<< Prev** | **Next >>**]

High performance phono amplifiers
[Join Now | Ring Hub | Random | **<< Prev** | **Next >>**]

High performance power amplifiers
[Join Now | Ring Hub | Random | **<< Prev** | **Next >>**]

Copyright © 2001-2005 WebRing, Inc. All rights reserved. Terms of Service - Help
Notice: We collect personal information on this site.
To learn more about how we use your information, see our Privacy Policy
WebRing server hosting provided by SimpleNet.
WebRing Logo design by Sean McCullough.