VALVE AMPLIFIERS

FOURTH EDITION

MORGAN JONES

Newnes

VALVE AMPLIFIERS

FOURTH EDITION

MORGAN JONES

Newnes

Table of Contents

Front-matter

Valve Amplifiers

Valve Amplifiers
Fourth Edition
Morgan Jones

ELSEVIER AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD • PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO Newnes
Newnes is an imprint of Elsevier

Copyright

Newnes is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

225 Wyman Street, Waltham, MA 02451, USA

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library
Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress
ISBN: 978-0-08096640-3

For information on all Newnes publications visit our web site at www.newnespress.com

Printed and bound in the UK
11 12 13 14 15 10 9 8 7 6 5 4 3 2 1

Preface

Almost 40 years ago the author bought his first valve amplifier; it cost him £3, and represented many weeks' pocket money. Whilst his pocket money has increased, so have his aspirations, and the DIY need was born.

Although there were many sources of information on circuit design, the electronics works gave scant regard to audio design, whilst the Hi-Fi books barely scratched the surface of the theory. The author, therefore, spent much time in libraries trying to link this information together to form a basis for audio design. This book is the result of those years of effort and aims to present thermionic theory in an accessible form without getting too bogged down in maths. Primarily, it is a book for practical people armed with a calculator or computer, a power drill and a (temperature-controlled) soldering iron.

The author started a B.Sc. in Acoustical Engineering, but left after a year to join BBC Engineering as a Technical Assistant, where he received excellent tuition in electronics and rose to the giddy heights of a Senior Engineer before being made redundant by BBC cuts. He has also served time in Higher Education, and although developing the UK's first B.Sc. (Hons.) Media Technology course and watching students blossom into graduates with successful careers was immensely satisfying, education is achieved by class contact – not by committees and paper chases.

Early on, he became a member of the Audio Engineering Society, and has designed and constructed many valve pre-amplifiers and power amplifiers, loudspeakers, pick-up arms and a pair of electrostatic headphones.

It is now 18 years since work began on the 1st edition of *Valve Amplifiers*, yet much has changed in this obsolete technology since then.

The relentless infestation of homes by computers has meant that test and measurement has become both cheaper and more easily integrated, either because it directly uses the processing power of a computer, or because it borrows from the technology needed to make them. Thus, the Fast Fourier Transform has become a tool for all to use, from industrial designer to keen amateur – enabling spectrum analysis via a £100 sound card that was the province of world class companies only 20 years ago. As a happy consequence, this edition benefits from detailed measurements limited primarily by the author's time. Computer modelling is now freely available – exemplified by Duncan Munro's PSUD2 linear power supply freeware.

The spread of Internet trading has made the market for valves truly global.

Exotica such as Loctals, European ‘Special Quality’ valves, and final generation Soviet bloc valves are now all readily available worldwide to any Luddite with the patience to access the Internet – we no longer need to be constrained to conservative (but expensive) choices of traditional audio valves. Even better, many of the 1950s engineering books that you thought had gone forever are available from the second-hand book sellers on the Internet.

Paradoxically, although digital electronics has improved the supply of valves, other analogue components are dying. Capacitors are the worst affected by the lack of raw materials; polycarbonate disappeared in 2001, and silvered-mica capacitors and polystyrene are both endangered species. Controls have succumbed to the ubiquitous digital encoder, so mechanical switch ranges have contracted and potentiometers face a similar Darwinian fate. It is particularly galling to discover a use for Zeners just as major semiconductor manufacturers stop making them.

Despite, or perhaps *because* of, these problems, valves and vinyl have become design icons, both in television adverts and the bits in between. The relentless hype from manufacturers of audio servers that favour convenience over sound quality has forced manufacturers of CD players to justify their products on sound quality (*and* convenience, because although nobody mentions it, a CD player is unable to wipe your entire music library at the drop of an operating system). CD and vinyl are now the only reliable sources of quality audio – which is perhaps a step forward from the 1980s when it was FM radio and vinyl.

Note for the MP3 generation: That shiny 120 mm disc was invented for storing music (such as Beethoven’s 9th Symphony) at far higher quality than a compressed download. Try it some time – you might even like it.

Dedication

The author would like to dedicate this book to the dwindling band of BBC engineers, particularly at BBC Southampton, and also to those at BBC Wood Norton, of which he has many colourful memories.

Acknowledgements

Special thanks are due to Euan McKenzie who undertook the onerous task of proofreading at long distance on short notice and in record time to an appropriately low uncertainty.

Thermionic design cannot proceed in a vacuum, so the author is grateful for the perceptive insights and insults freely offered by Stuart Yaniger over the recent years.

An annual celebration of awe and wonder has been the European Triode Festival. This delightfully civilised bacchanalia has humbled the author with splendid works of art and engineering whilst at the same time reassuring him that he was not alone. Thank you, Christian, for first inviting me, and even more thanks to subsequent organisers for successfully maintaining the momentum.

Finally, the author would like to thank those readers who took the time and trouble to breach the publishing citadel and give the author hugely useful feedback.

Chapter 1. Circuit Analysis

In order to look at the interesting business of designing and building valve amplifiers, we need some knowledge of electronics funmentals. Unfortunately, fundamentals are not terribly interesting, and to cover them fully would consume the entire book. Ruthless pruning is, therefore, necessary to condense what is needed in one chapter.

It is thus with deep sorrow that the author has had to forsaken complex numbers and vectors, whilst the omission of differential calculus is a particularly poignant loss. All that is left is ordinary algebra, and although there are *lots* of equations, they are timid, miserable creatures and quite defenceless.

If you are comfortable with basic electronic terms and techniques, then please feel free to go directly to Chapter 2, where valves appear.

Mathematical Symbols

Unavoidably, a number of mathematical symbols are used, some of which you may have forgotten, or perhaps not previously met:

$a \equiv b$

a is totally equivalent to *b*

$a = b$

a equals *b*

$a \approx b$

a is approximately equal to *b*

$a \propto b$

a is proportional to *b*

$a \neq b$

a is not equal to *b*

$a > b$

a is greater than *b*

$a < b$

a is less than *b*

a*≥ *b

a is greater than, or equal to, *b*

a*≤ *b

a is less than, or equal to, *b*

As with the = and ≠ symbols, the four preceding symbols can have a slash through them to negate their meaning (*a* ∋ *b*, a is not less than *b*).

√ *a*

the number which when multiplied by itself is equal to *a* (square root)

a^n

a multiplied by itself *n* times. $a^4 = a \times a \times a \times a$ (*a* to the power *n*)

±

plus or minus

∞

infinity

°

degree, either of temperature (°C), or of an angle (360° in a circle)

|

parallel, either parallel lines, or an electrical parallel connection

Δ

a small change in the associated value, such as ΔV_{gk}.

Electrons and Definitions

Electrons are *charged* particles. Charged objects are attracted to other charged particles or objects. A practical demonstration of this is to take a balloon, rub it briskly against a jumper and then place the rubbed face against a wall. Let it go. The balloon remains stuck to the wall. This is because we have charged the balloon, and so there is an attractive force between it and the wall. (Although the wall was initially uncharged, placing the balloon on the wall induced a charge.)

Charged objects come in two forms: negative and positive. Unlike charges attract, and like charges repel. Electrons are negative and positrons are positive, but whilst electrons are stable in our universe, positrons encounter an electron almost immediately after production, resulting in mutual annihilation and a pair of 511 keV gamma rays.

If we don't have ready access to positrons, how can we have a positively charged object? Suppose we had an object that was negatively charged, because it had 2,000 electrons clustered on its surface. If we had another, similar, object that only had 1,000 electrons on its surface, then we would say that the first object was more negatively charged than the second, but as we can't count how many electrons we have, we might just as easily have said that the second object was more positively charged than the first. It's just a matter of which way you look at it.

To charge our balloon, we had to do some work and use energy. We had to overcome friction when rubbing the balloon against the woollen jumper. In the process, electrons were moved from one surface to the other. Therefore, one object (the balloon) has acquired an excess of electrons and is negatively charged, whilst the other object (woollen jumper) has lost the same number of electrons and is positively charged.

The balloon would, therefore, stick to the jumper. Or would it? Certainly it will be attracted to the jumper, but what happens when we place the two in contact? The balloon does not stick. This is because the fibres of the jumper were able to touch the whole of the charged area on the balloon, and the electrons were so attracted to the jumper that they moved back onto the jumper, thus neutralising the charge.

At this point, we can discard vague talk of balloons and jumpers because we have just observed electron flow.

An electron is very small, and doesn't have much of a charge, so we need a more practical unit for defining charge. That practical unit is the *coulomb* (*C*). We could now say that 1 C of charge had flowed between one point and another, which would be equivalent to saying that approximately 6,240,000,000,000,000,000 electrons had passed, but much handier.

Simply being able to say that a large number of electrons had flowed past a given point is not in itself very helpful. We might say that a billion cars have travelled down a particular section of motorway since it was built, but if you were planning a journey down that motorway, you would want to know the flow of cars *per hour* through that section.

Similarly in electronics, we are not concerned with the total flow of electrons since the dawn of time, but we do want to know about electron flow at any given instant. Thus, we could define the flow as the number of coulombs of charge that flowed past a point in one second. This is still rather long-winded, and we will abbreviate yet further.

We will call the flow of electrons *current*, and as the coulomb/second is unwieldy, it will be redefined as a new unit, the *ampere* (*A*). Because the ampere is such a useful unit, the definition linking current and charge is usually stated in the following form.

One coulomb is the charge moved by one ampere flowing for one second.

$$\text{charge (coulombs)} = \text{current (amperes)} \times \text{time (seconds)}$$

This is still rather unwieldy, so symbols are assigned to the various units: charge has symbol *Q*, current *I* and time *t*.

$$Q = It$$

This is a very useful equation, and we will meet it again when we look at capacitors (which store charge).

Meanwhile, current has been flowing, but why did it flow? If we are going to move electrons from one place to another, we need a force to cause this movement. This force is known as the electro motive force (EMF). Current continues to flow whilst this force is applied, and it flows from a higher potential to a lower potential.

If two points are at the same potential, no current can flow between them. What is important is the *potential difference* (*pd*).

A potential difference causes a current to flow between two points. As this is a new property, we need a unit, a symbol and a definition to describe it. We mentioned work being done in charging the balloon, and in its very precise and physical sense, this is how we can define potential difference, but first, we must define *work*.

One joule of work is done if a force of one newton moves one metre from its point of application.

This very physical interpretation of work can be understood easily once we realise that it means that one joule of work would be done by moving one kilogramme a distance of one metre in one second. Since charge is directly related to the mass of electrons moved, the physical definition of work can be modified to define the force that causes the movement of charge.
Unsurprisingly, because it causes the motion of electrons, the force is called the Electro-Motive Force, and it is measured in *volts*.

If one joule of work is done moving one coulomb of charge, then the system is said to have a potential difference of one volt (V).

$$\text{work done (joules)} = \text{charge (coulombs)} \times \text{potential difference (volts)}$$

$$W = QV$$

The concept of work is important because work can be done only by the expenditure of energy, which is, therefore, also expressed in joules.

$$\text{work done (joules)} = \text{energy expended (joules)}$$

$$W = E$$

In our specialised sense, doing work means moving charge (electrons) to make currents flow.

Batteries and Lamps

If we want to make a current flow, we need a *circuit*. A circuit is exactly that a loop or path through which a current can flow, from its starting point all the way round the circuit, to *return* to its starting point. Break the circuit, and the current ceases to flow.
The simplest circuit that we might imagine is a battery connected to an incandescent lamp via a switch. We *open* the switch to stop the current flow (open circuit) and close it to light the lamp. Meanwhile, our helpful friend (who has been watching all this) leans over and drops a thick piece of copper across the battery terminals, causing a *short* circuit.
The lamp goes out. Why?

Ohm's Law

To answer the last question, we need some property that defines how much current flows. That property is *resistance*, so we need another definition, units and a symbol.

If a potential difference of one volt is applied across a resistance, resulting in a current of one ampere, then the resistance has a value of one ohm (Ω).

$$\text{potential difference (volts)} = \text{current (amperes)} \times \text{resistance (ohms)}$$

$$V = IR$$

This is actually a simplified statement of Ohm's law, rather than a strict definition of resistance, but we don't need to worry too much about that.

We can rearrange the previous equation to make *I* or *R* the subject.

$$I = \frac{V}{R}$$

$$R = \frac{V}{I}$$

These are incredibly powerful equations and should be committed to memory.

The circuit shown in Figure 1.1 is switched on, and a current of 0.25 A flows. What is the resistance of the lamp?

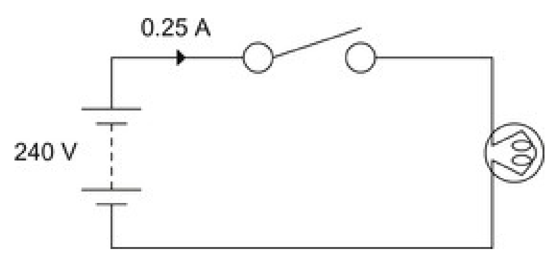


Figure 1.1 Use of Ohm's law to determine the resistance of a hot lamp.

$$R = \frac{V}{I} = \frac{240}{0.25} = 960\ \Omega$$

Now this might seem like a trivial example, since we could easily have measured the resistance of the lamp to 3½ significant figures using our shiny, new, digital multimeter. But could we? The hot resistance of an incandescent lamp is very different from its cold resistance; in the example above, the cold resistance was 80 Ω.

We could now work the other way and ask how much current would flow through an 80 Ω resistor connected to 240 V.

$$I = \frac{V}{R} = \frac{240}{80} = 3 \text{ A}$$

Incidentally, this is why incandescent lamps are most likely to fail at switch-on. The high initial current that flows before the filament has warmed up and increased its resistance stresses the weakest parts of the filament, they become so hot that they vaporise, and the lamp blows.

Power

In the previous example, we looked at an incandescent lamp and rated it by the current that flowed through it when connected to a 240 V battery. But we all know that lamps are rated in *watts*, so there must be some connection between the two.

One watt (W) of power is expended if one joule of work is done in one second.

$$\text{power (watts)} = \frac{\text{work done (joules)}}{\text{time taken (seconds)}}$$

$$P = \frac{W}{t}$$

This may not seem to be the most useful of definitions, and, indeed, it is not, but by combining it with some earlier equations:

$$W = QV$$

So:

$$P = \frac{QV}{t}$$

But:

$$Q = It$$

So:

$$P = \frac{IVt}{t}$$

We obtain:

$$P = IV$$

This is a fundamental equation of equal importance to Ohm's law. Substituting the Ohm's law equations into this yields:

$$P = \frac{V^2}{R}$$

$$P = I^2 R$$

We can now use these equations to calculate the power rating of our lamp. Since it drew 0.25 A when fed from 240 V, and had a hot resistance of 960 Ω, we can use any of the three equations.
Using:

$$P = \frac{V^2}{R}$$

$$P = \frac{240^2}{960}$$

$$P = 60\ \mathrm{W}$$

It will probably not have escaped your notice that this lamp looks suspiciously like an AC mains lamp, and that the battery was rather large. We will return to this later.

Kirchhoff's Laws

There are two of these: a current law and a voltage law. They are both very simple and, at the same time, very powerful.
The current law states:

The algebraic sum of the currents flowing into, and out of, a node is equal to zero.

$$0 = I_1 + I_2 + I_3 + \cdots$$

What it says in a more relaxed form is that what goes in, comes out. If we have 10 A going into a node, or junction, then that much current must also leave that junction – it might not all come out on one wire, but it must all come out. A conservation of current, if you like (see Figure 1.2).

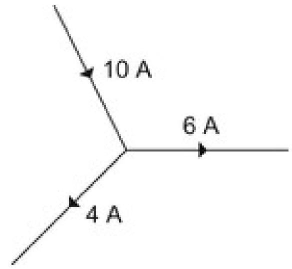


Figure 1.2 Currents at a node (Kirchhoff's current law).

$$\begin{aligned}
\text{current flowing into the node:} \quad & I_1 = 10\ \text{A} \\
\text{currents leaving the node:} \quad & I_2 = 4\ \text{A} \\
& I_3 = 6\ \text{A} \\
\text{total current leaving the node:} \quad & I_{\text{total}} = 10\ \text{A}
\end{aligned}$$

From the point of view of the node, the currents leaving the node are flowing in the opposite direction to the current flowing into the node, so we *must* give them a minus sign before plugging them into the equation.

$$\begin{aligned}
0 &= I_1 + I_2 + I_3 \\
&= 10\ \text{A} + (-4\ \text{A}) + (-6\ \text{A}) \\
&= 10 - 4 - 6
\end{aligned}$$

This may have seemed pedantic, since it was obvious from the diagram that the incoming currents equalled the outgoing currents, but you may need to find a current when you do not even know the direction in which it is flowing. Using this convention forces the correct answer!

It is vital to make sure that your signs are correct.

The voltage law states:

> *The algebraic sum of the EMFs and potential differences acting around any loop is equal to zero.*

This law draws a very definite distinction between EMFs and potential differences. EMFs are the *sources* of electrical energy (such as batteries), whereas potential differences are the *voltages dropped* across components. Another way of stating the law is to say that the algebraic sum of the EMFs must equal the algebraic sum of the potential drops around the loop. Again, you could consider this to be a conservation of voltage (see Figure 1.3).

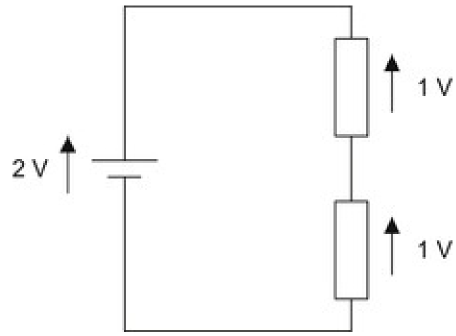


Figure 1.3 Summation of potentials within a loop (Kirchhoff's voltage law).

Resistors in Series and Parallel

If we had a *network* of resistors, we might want to know what the total resistance was between terminals A and B (see Figure 1.4).

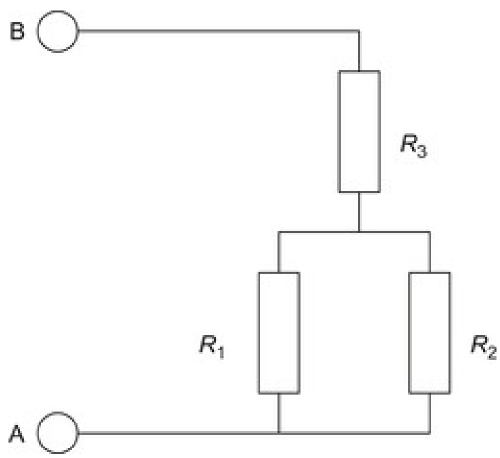


Figure 1.4 Series/parallel resistor network.

We have three resistors: R_1 is in *parallel* with R_2, and this *combination* is in *series* with R_3.

As with all problems, the thing to do is to break it down into its simplest parts. If we had some means of determining the value of resistors in series, we could use it to calculate the value of R_3 in series with the combination of R_1 and R_2, but as we do not yet know the value of the parallel combination, we must find this *first*. This question of order is most important, and we will return to it later.

If the two resistors (or any other component, for that matter) are in parallel, then they must have the same voltage drop across them. Ohm's law might therefore, be a useful starting point.

$$I = \frac{V}{R}$$

Using Kirchhoff's current law, we can state that:

$$I_{\text{total}} = I_{R_1} + I_{R_2} + \cdots$$

So:

$$\frac{V}{R} = \frac{V}{R_1} + \frac{V}{R_2} + \cdots$$

Dividing by *V*:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots$$

The reciprocal of the total parallel resistance is equal to the sum of the reciprocals of the individual resistors.

For the special case of only two resistors, we can derive the equation:

$$R = \frac{R_1 R_2}{R_1 + R_2}$$

This is often known as 'product over sum', and whilst it is useful for mental arithmetic, it is slow to use on a calculator (more keystrokes).

Now that we have cracked the parallel problem, we need to crack the series problem.

First, we will simplify the circuit. We can now calculate the total resistance of the parallel combination and replace it with one resistor of that value – an *equivalent* resistor (see Figure 1.5).

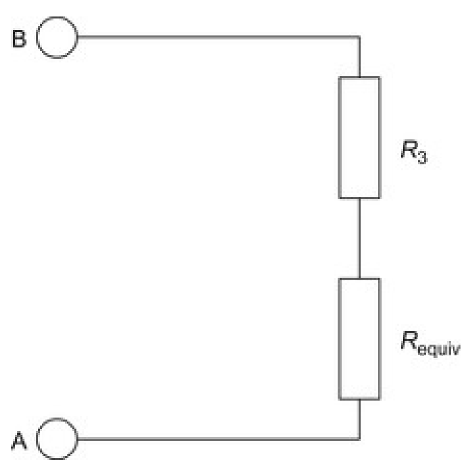


Figure 1.5 Simplification of Fig. 1.4 using an equivalent resistor.

Using the voltage law, the sum of the potentials across the resistors must be equal to the driving EMF:

$$V_{\text{total}} = V_{R_1} + V_{R_2} + \cdots$$

Using Ohm's law:

$$V_{\text{total}} = IR_1 + IR_2 + \cdots$$

But if we are trying to create an equivalent resistor, whose value is equal to the combination, we could say:

$$IR_{\text{total}} = IR_1 + IR_2 + \cdots$$

Hence:

$$R_{\text{series}} = R_1 + R_2 + \cdots$$

The total resistance of a combination of series resistors is equal to the sum of their individual resistances.

Using the parallel and series equations, we are now able to calculate the total resistance of *any* network (see Figure 1.6).

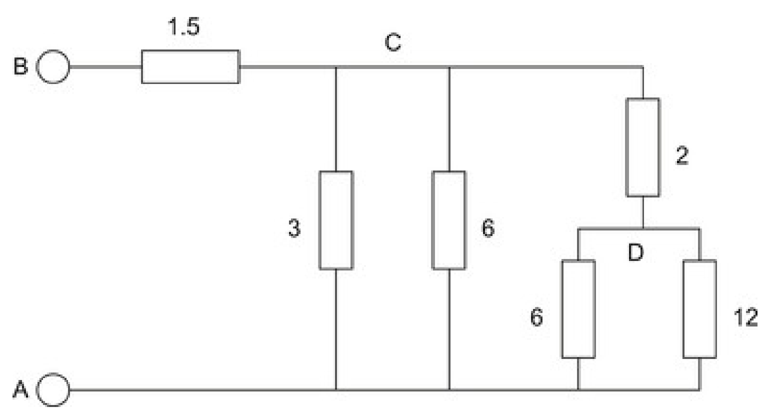


Figure 1.6

Now this may look horrendous, but it is not a problem if we attack it logically. The hardest part of the problem is not wielding the equations or numbers, but where to start.

We want to know the resistance looking into the terminals A and B, but we do not have any rules for finding this directly, so we must look for a point where we can apply our rules. We can apply only one rule at a time, so we look for a combination of components made up *only* of series *or* parallel components.

In this example, we find that between node A and node D there are *only* parallel components. We can calculate the value of an equivalent resistor and substitute it back into the circuit:

$$R_{\text{parallel}} = \frac{\text{product}}{\text{sum}}$$
$$R_{\text{parallel}} = \frac{6 \times 12}{6 + 12}$$
$$R_{\text{parallel}} = 4\ \Omega$$

We redraw the circuit (see Figure 1.7).

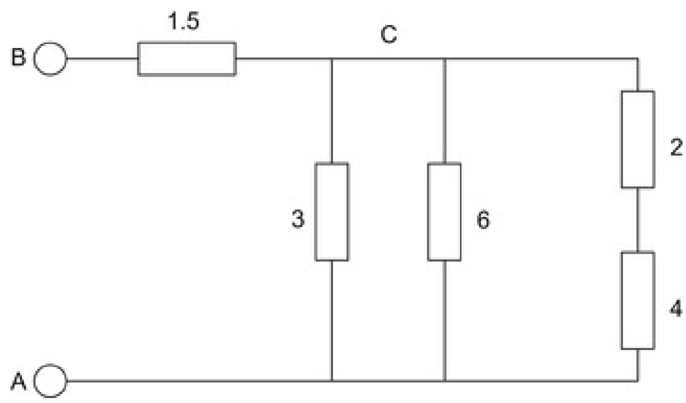


Figure 1.7

Looking again, we find that now the *only* combinations made up of series *or* parallel components are between node A and node C, but we have a choice – either the series combination of the 2 Ω and 4 Ω, or the parallel combination of the 3 Ω and 6 Ω. The one to go for is the series combination. This is because it will result in a single resistor that will then be in parallel with the 3 Ω and 6 Ω resistors. We can cope with the three parallel resistors later:

$$R_{\text{series}} = R_1 + R_2$$
$$R_{\text{series}} = 4 + 2$$
$$R_{\text{series}} = 6\ \Omega$$

We redraw the circuit (see Figure 1.8).

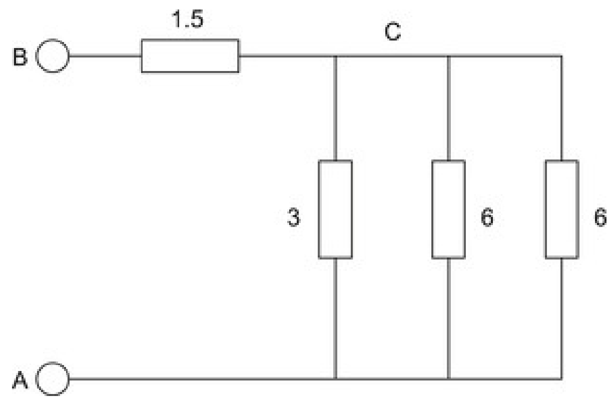


Figure 1.8

We now see that we have three resistors in parallel:

$$\frac{I}{R} = \frac{I}{R_1} + \frac{I}{R_2} + \frac{I}{R_3}$$

$$\frac{I}{R} = \frac{1}{3} + \frac{1}{6} + \frac{1}{6}$$

$$\frac{I}{R} = \frac{2}{3}$$

Hence:

$$R = \frac{3}{2} = 1.5\ \Omega$$

We have reduced the circuit to two 1.5 Ω resistors in series, and so the total resistance is 3 Ω.

This took a little time, but it demonstrated some useful points that will enable you to analyse networks much faster the second time around:

- The critical stage is choosing the starting point.
- The starting point is generally as far away from the terminals as it is possible to be.
- The starting point is made up of a combination of *only* series *or* parallel components.
- Analysis tends to proceed outwards from the starting point towards the terminals.
- Redrawing the circuit helps. You may even need to redraw the original circuit if it does not make sense to you. Redrawing as analysis progresses reduces confusion and errors – do it!

Potential Dividers

Figure 1.9 shows a potential divider. This could be made up of two discrete resistors, or it could be the moving wiper of a volume control. As before, we will suppose that a current *I* flows through the two resistors. We want to know the ratio of the output voltage to the input voltage (see Figure 1.9).

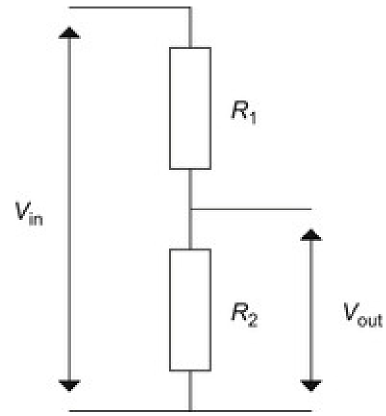


Figure 1.9 Potential divider.

$$V_{out} = IR_2$$
$$V = IR_1 + IR_2$$
$$V = I(R_1 + R_2)$$

Hence:

$$\frac{V_{out}}{V} = \frac{IR_2}{I(R_1 + R_2)}$$

$$\boxed{\frac{V_{out}}{V} = \frac{R_2}{R_1 + R_2}}$$

This is a very important result and, used intelligently, can solve virtually anything.

Equivalent Circuits

We have looked at networks of resistors and calculated equivalent *resistances*. Now we will extend the idea to equivalent *circuits*. This is a tremendously powerful concept for circuit analysis.

It should be noted that this is not the only method, but it is usually the quickest and kills 99% of all known problems. Other methods include Kirchhoff's laws combined with lots of simultaneous equations and the superposition theorem.

These methods may be found in standard texts, but they tend to be cumbersome, so we will not discuss them here.

The Thévenin Equivalent Circuit

When we looked at the potential divider, we were able to calculate the ratio of output voltage to input voltage. If we were now to connect a battery across the input terminals, we could calculate the output voltage. Using our earlier tools, we could also calculate the total resistance looking into the output terminals. As before, we could then redraw the circuit, and the result is known as the *Thévenin equivalent circuit*. If two black boxes were made, one containing the original circuit and the other the Thévenin equivalent circuit, you would not be able to tell from the output terminals which was which. The concept is simple to use and can break down complex networks quickly and efficiently (see Figure 1.10).

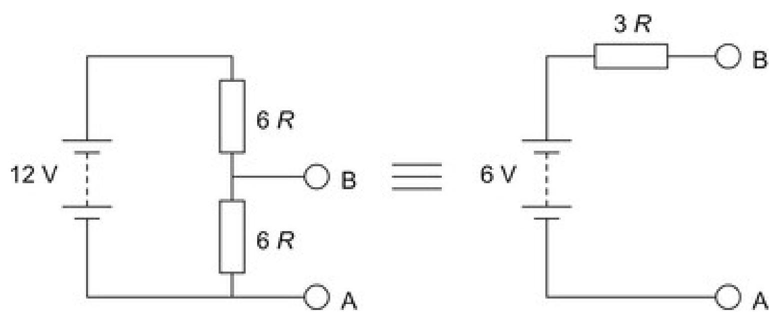


Figure 1.10 A 'black box' network and its Thévenin equivalent circuit.

This is a simple example to demonstrate the concept. First, we find the equivalent resistance, often known as the *output resistance*. Now, in the world of equivalent circuits, batteries are perfect voltage sources; they have zero *internal resistance* and look like a short circuit when we consider their resistance. Therefore, we can ignore the battery, or replace it with a piece of wire whilst we calculate the resistance of the total circuit:

$$R_{\text{output}} = \frac{R_1 R_2}{R_1 + R_2}$$

$$R_{\text{output}} = 3\ \Omega$$

Next, we need to find the output voltage. We will use the potential divider equation:

$$\frac{V_{\text{out}}}{V} = \frac{R_2}{R_1 + R_2}$$

$$\frac{V_{\text{out}}}{V} = \frac{1}{2}$$

So:

$$V_{\text{out}} = \frac{1}{2}\ \text{V}$$

$$V_{\text{out}} = 6\ \text{V}$$

Now for a much more complex example. This will use all the previous techniques and really demonstrate the power of Thévenin (see Figure 1.11).

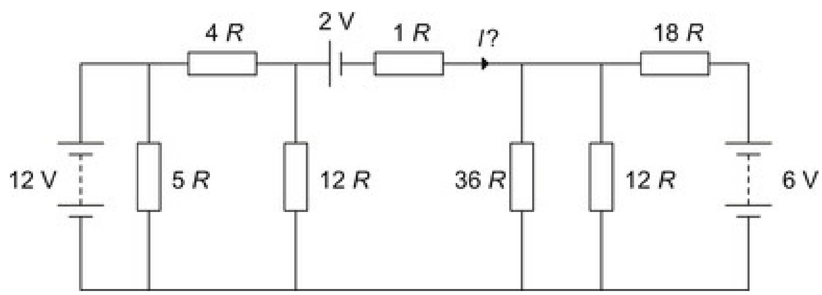


Figure 1.11

For some obscure reason, we want to know the current flowing in the 1 Ω resistor. The first thing to do is to redraw the circuit. Before we do this we can observe that the 5 Ω resistor in parallel with the 12 V battery is irrelevant. Yes, it will draw current from the battery, but it does not affect the operation of the rest of the circuit. A short circuit in parallel with 5 Ω is *still* a short circuit, so we will throw it away (see Figure 1.12).

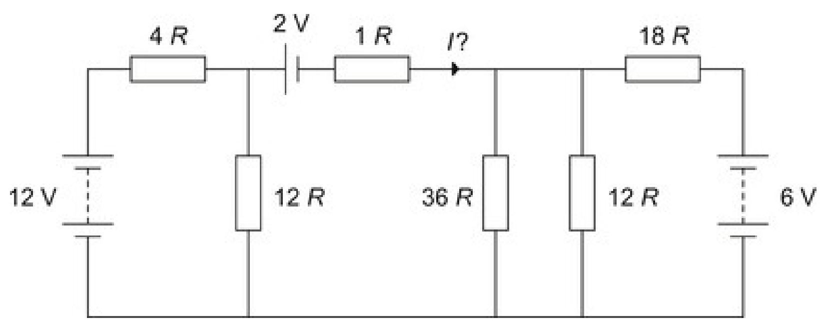


Figure 1.12

Despite our best efforts, this is still a complex circuit, so we need to break it down into modules that we recognise. Looking at the left-hand side, we find a battery with a 4 Ω and 12 Ω resistor which looks suspiciously like the simple problem that we saw earlier, so let us break the circuit there and make an equivalent circuit (see Figure 1.13).

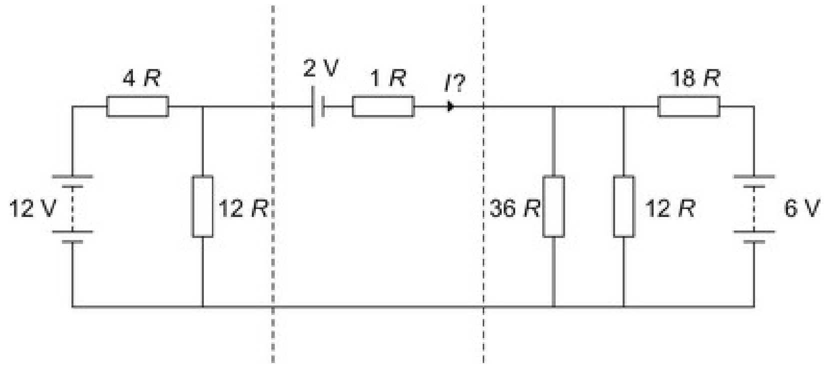


Figure 1.13

Using the potential divider rule:

$$V_{\text{out}} = \frac{R_2}{R_1 + R_2} \cdot V$$
$$V_{\text{out}} = 9 \text{ V}$$

Using ‘product over sum’:

$$R_{\text{out}} = \frac{R_1 R_2}{R_1 + R_2}$$
$$R_{\text{out}} = 3 \ \Omega$$

Looking at the right-hand side, we can perform a similar operation to the right of the dashed line.

First, we find the resistance of the parallel combination of the 36 Ω and 12 Ω resistors, which is 9 Ω. We now have a potential divider, whose output resistance is 6 Ω and the Thévenin voltage is 2 V.

Now we redraw the circuit (see Figure 1.14).

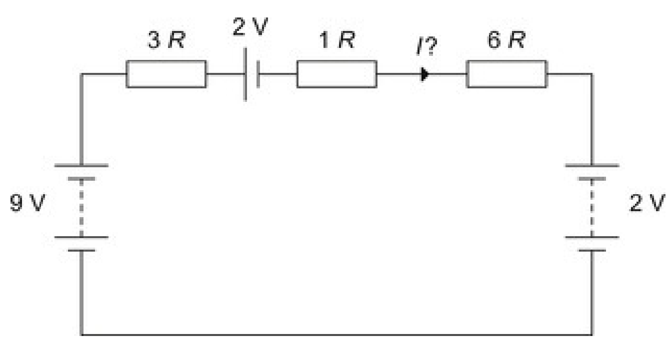


Figure 1.14

We can make a few observations at this point. First, we have three batteries in series, why not combine them into one battery? There is no reason why we should not do this provided that we take note of their polarities. Similarly, we can combine some, or all, of the resistors (see Figure 1.15).

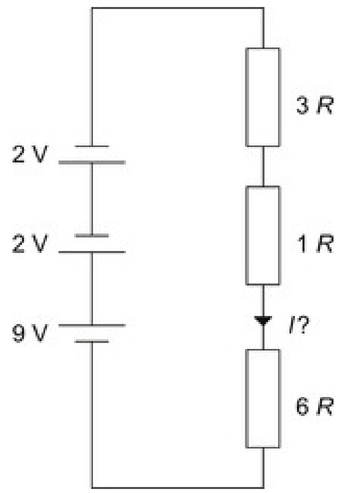


Figure 1.15

The problem now is trivial, and a simple application of Ohm's law will solve it. We have a total resistance of 10 Ω and a 5 V battery, so the current must be 0.5 A.

Useful points to note:

- Look for components that are irrelevant, such as resistors directly across battery terminals.
- Look for potential dividers on the outputs of batteries and 'Thévenise' them. Keep on doing so until you meet the next battery.
- Work from battery terminals outwards.
- Keep calm, and try to work neatly – it will save mistakes later.

Although it is possible to solve most problems using a Thévenin equivalent circuit, sometimes a Norton equivalent is more convenient.

The Norton Equivalent Circuit

The Thévenin equivalent circuit was a perfect voltage source in *series* with a resistance, whereas the *Norton* equivalent circuit is a perfect *current* source in *parallel* with a resistance (see Figure 1.16).

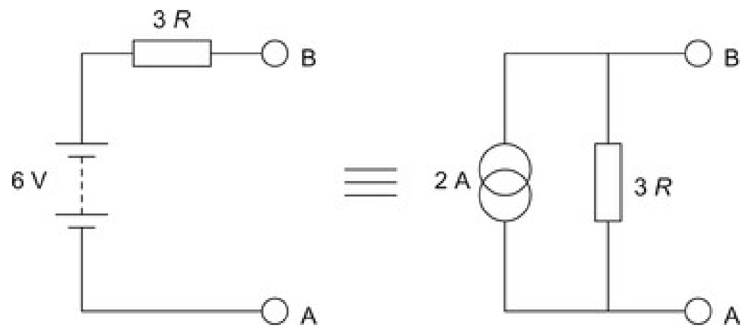


Figure 1.16 The Norton equivalent circuit.

We can easily convert from a Norton source to a Thévenin source, or vice versa, because the resistor has the same value in both cases. We find the value of the current source by short circuiting the output of the Thévenin source and calculating the resulting current – this is the Norton current.

To convert from a Norton source to a Thévenin source, we leave the source open circuit and calculate the voltage developed across the Norton resistor – this is the Thévenin voltage.

For the vast majority of problems, the Thévenin equivalent will be quicker, mostly because we become used to thinking in terms of voltages that can easily be measured by a meter or viewed on an oscilloscope. Occasionally, a problem will arise that is intractable using Thévenin, and converting to a Norton equivalent causes the problem to solve itself. Norton problems usually involve the summation of a number of currents, when the only other solution would be to resort to Kirchhoff and simultaneous equations.

Units and Multipliers

All the calculations up to this point have been arranged to use convenient values of voltage, current and resistance. In the real world, we will not be so fortunate, and to avoid having to use scientific notation, which takes longer to write and is virtually unpronounceable, we will prefix our units with multipliers [1].

Prefix	Abbreviation	Multiplies by
yocto	y	10^{-24}
zepto	z	10^{-21}
atto	a	10^{-18}
femto	f	10^{-15}
pico	p	10^{-12}
nano	n	10^{-9}
micro	μ	10^{-6}
milli	m	10^{-3}
kilo	k	10^{3}

mega	M	10^{6}
giga	G	10^{9}
tera	T	10^{12}
peta	P	10^{15}
exa	E	10^{18}
zetta	Z	10^{21}
yotta	Y	10^{24}

Note that the *case* of the prefix is important; there is a large difference between 1 mΩ and 1 MΩ. Electronics uses a very wide range of values; small-signal pentodes have anode to grid capacitances measured in fF (F=farad, the unit of capacitance), and petabyte data stores are already common. Despite this, for day-to-day electronics use, we only need to use pico to mega.

Electronics engineers commonly abbreviate further, and you will often hear a 22 pF (picofarad) capacitor referred to as 22 ‘puff’, whilst the ‘ohm’ is commonly dropped for resistors, and 470 kΩ (kiloohm) would be pronounced as ‘four-seventy-kay’.

A rather more awkward abbreviation that arose before high-resolution printers became common (early printers could not print ‘µ’), is the abbreviation of µ (micro) to m. This abbreviation persists, particularly in the US text, and you will occasionally see a 10 mF capacitor specified, although the context makes it clear that what is actually meant is 10 µF. For this reason, true 10 mF capacitors are invariably specified as 10,000 µF.

Unless an equation states otherwise, assume that it uses the base physical units, so an equation involving capacitance and time constants would expect you to express capacitance in farads and time in seconds. Thus, 75 µs=75×10^{-6} s, and the value of capacitance determined by an equation might be 2.2×10^{-10} F=220 pF. Very occasionally, it is handier to express an equation using real-world units such as mA or MΩ, in which case the equation or its accompanying text will *always* explain this break from convention.

The Decibel

The human ear spans a vast dynamic range from the near silence heard in an empty recording studio to the deafening noise of a nearby pneumatic drill. If we were to plot this range linearly on a graph, the quieter sounds would hardly be seen, whereas the difference between the noise of the drill and that of a jet engine would be given a disproportionate amount of room on the graph. What we need is a graph that gives an equal weighting to *relative* changes in the level of both quiet and loud sounds. By definition, this implies a logarithmic scale on the graph, but electronics engineers went one better and invented a logarithmic ratio known as the *decibel* (*dB*) that was promptly hijacked by the acoustical engineers. (The fundamental unit is the bel, but this is inconveniently large, so the decibel is more commonly used.)

The dB is *not* an absolute quantity. It is a *ratio*, and it has one formula for use with currents and voltages and another for powers:

$$\mathrm{dB} = 20\log_{10}\left(\frac{V_1}{V_2}\right) = 20\log_{10}\left(\frac{I_1}{I_2}\right) = 10\log_{10}\left(\frac{P_1}{P_2}\right)$$

The reason for this is that $P \propto V^2$ or I^2, and with logarithms, multiplying the logarithm by 2 is the same as squaring the original number. Using a different formula to calculate dBs when using powers ensures that the resulting dBs are equivalent, irrespective of whether they were derived from powers or voltages.

This might seem complicated when all we wanted to do was to describe the difference in two signal levels, but the dB is a very handy unit.

Useful common dB values are:

dB	V_1/V_2	P_1/P_2
0	1	1
3	√2	2
6	2	4
20	10	100

Because dBs are derived from logarithms, they obey all the rules of logarithms, and adding dBs is the same as multiplying the ratios that generated them. Note that dBs can be negative, implying loss or a drop in level.

For example, if we had two cascaded amplifiers, one with a voltage gain of 0.5 and the other with a voltage gain of 10, then by *multiplying* the individual gains, the combined voltage gain would be 5. Alternatively, we could find the gain in dB by saying that one amplifier had −6 dB of gain whilst the other had 20 dB, and *adding* the gains in dB to give a total gain of 14 dB.

When *designing* amplifiers, we will not often use the above example, as absolute voltages are often more convenient, but we frequently need dBs to describe filter and equalisation curves.

Alternating Current (AC)

All the previous techniques have used *direct current* (*DC*), where the current is constant and flows in one direction only. Listening to DC is not very interesting, so we now need to look at *alternating currents* (*AC*).

All of the previous techniques of circuit analysis can be applied equally well to AC signals.

The Sine Wave

The *sine* wave is the simplest possible alternating signal, and its equation is:

$$v = V_{peak}\sin(\omega t + \theta)$$

where

v=the instantaneous value at time t

V_{peak}=the peak value

ω=angular frequency in radians/second (ω=2 πf)

t=the time in seconds

θ=a constant phase angle.

These mathematical concepts are shown on the diagram (see Figure 1.17).

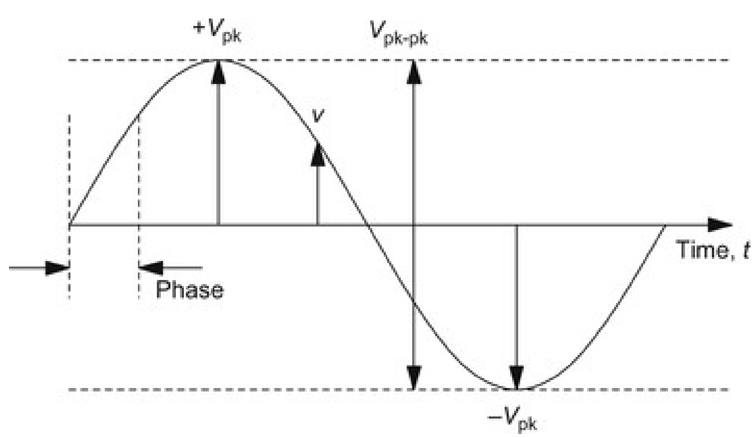


Figure 1.17

Equations involving changing quantities use a convention. Upper case letters denote DC, or constant values, whereas lower case letters denote the instantaneous AC, or changing, value. It is a form of shorthand to avoid having to specify separately that a quantity is AC or DC. It would be nice to say that this convention is rigidly applied, but it is often neglected, and the context of the symbols usually makes it clear whether the quantity is AC or DC.

In electronics, the word ‘peak’ (pk) has a very precise meaning and, when used to describe an AC waveform, it means the voltage from zero volts to the peak voltage reached, either positive or negative. Peak to peak (pk–pk) means the voltage from positive peak to negative peak, and for a symmetrical waveform, $V_{pk\text{–}pk}$=2 V_{pk}.

Although electronics engineers habitually use ω to describe frequency, they do so only because calculus requires that they work in radians. Since ω=2 πf, we can rewrite the equation as:

$$v = V_{\text{peak}} \sin(2\pi ft + \theta)$$

If we now inspect this equation, we see that apart from time *t*, we could vary other constants before we allow time to change and determine the waveform. We can change V_{peak}, and this will change the *amplitude* of the sine wave, or we can change *f*, and this will change the *frequency*. The inverse of frequency is *period*, which is the time taken for one full cycle of the waveform to occur:

$$\text{period } (T) = \frac{1}{f}$$

If we listen to a sound that is a sine wave, and change the amplitude, this will make the sound louder or softer, whereas varying frequency changes the pitch. If we vary θ (phase), it will sound the same if we are listening to it, and unless we have an external reference, the sine wave will look exactly the same viewed on an oscilloscope. Phase becomes significant if we compare one sine wave with another sine wave *of the same frequency* or a harmonic of that frequency. Attempting to compare phase between waveforms of unrelated frequencies is meaningless.

Now that we have described sine waves, we can look at them as they would appear on the screen of an oscilloscope. See Figure 1.18.

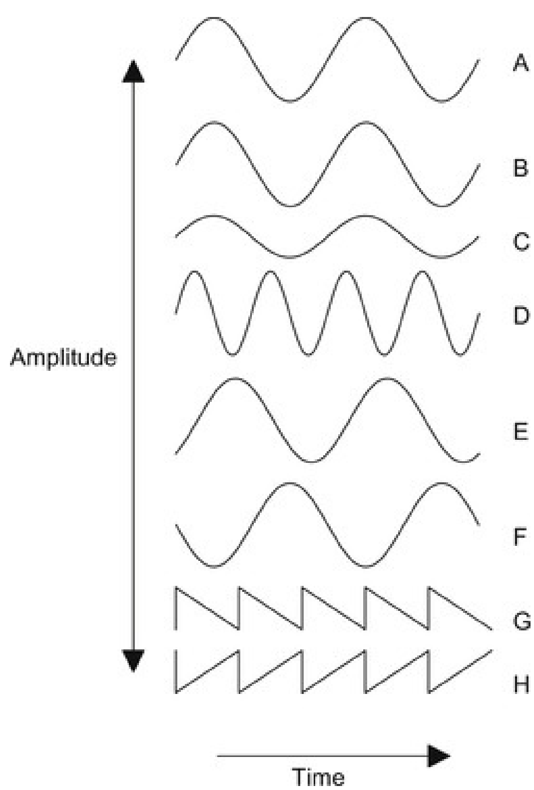


Figure 1.18

Sine waves A and B are identical in amplitude, frequency and phase. Sine wave C has lower amplitude, but frequency and phase are the same. Sine wave D has the same amplitude, but double the frequency. Sine wave E has identical amplitude and frequency to A and B, but the phase θ has been changed.

Sine wave F has had its polarity *inverted*. Although, for a sine wave, we cannot see the difference between a 180° phase change and a polarity inversion, for asymmetric waveforms there is a distinct difference. We should, therefore, be very careful if we say that two waveforms are 180° out of phase with each other that we do not actually mean that one is inverted with respect to the other.

The *sawtooth* waveform G has been inverted to produce the waveform H, and it can be seen that this is completely different from a 180° phase change. (Strictly, the UK term 'phase splitter' is entirely incorrect for this reason, and the US description 'phase inverter' is much better, but falls short of the technically correct description 'polarity inverter' used by nobody.)

The Transformer

When the electric light was introduced as an alternative to the gas mantle, there was a great debate as to whether the distribution system should be AC or DC. The outcome was settled by the enormous advantage of the *transformer*, which could step up, or step down, the voltage of an AC supply. The DC supply could not be manipulated in this way, and evolution took its course.

A transformer is essentially a pair of electrically insulated windings that are magnetically coupled to each other, usually on an iron core. They vary from the size of a fingernail to that of a large house, depending on power rating and operating frequency, with high frequency transformers being smaller. The symbol for a transformer is modified depending on the core material. Solid lines indicate a laminated iron core and dotted lines denote a dust core, whilst an air core has no lines (see Figure 1.19).

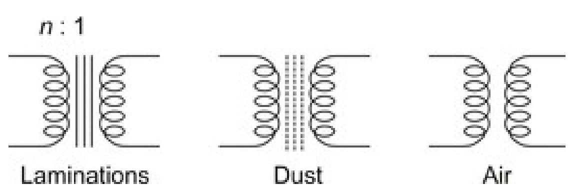


Figure 1.19 Transformer symbols.

The perfect transformer changes one AC voltage to another, more convenient voltage with no losses whatsoever; all of the power at the input is transferred to the output:

$$P = P_{\text{out}}$$

Having made this statement, we can now derive some useful equations:

$$V \cdot I = V_{\text{out}} \cdot I_{\text{out}}$$

Rearranging:

$$\frac{V}{V_{\text{out}}} = \frac{I_{\text{out}}}{I} = n$$

The new constant *n* is very important and is the ratio between the number of turns on the input winding and the number of turns on the output winding of the transformer. Habitually, when we talk about transformers, the input winding is known as the *primary* and the output winding is the *secondary*.

$$n\ (\text{turns ratio}) = \frac{\text{number of primary turns}}{\text{number of secondary turns}}$$

Occasionally, an audio transformer may have a winding known as a *tertiary* winding, which usually refers to a winding used for feedback or monitoring, but it is more usual to refer to multiple primaries and secondaries.

When the perfect transformer steps voltage down, perhaps from 240 V to 12 V, the current ratio is stepped up, and each ampere of primary current is due to 20 A drawn from the secondary. This implies that the resistance of the load on the secondary is different from that seen looking into the primary. If we substitute Ohm’s law into the conservation of power equation:

$$\frac{V^2_{\text{primary}}}{R_{\text{primary}}} = \frac{V^2_{\text{secondary}}}{R_{\text{secondary}}}$$

$$\frac{R_{\text{secondary}}}{R_{\text{primary}}} = \left(\frac{V_{\text{primary}}}{V_{\text{secondary}}}\right)^2 = n^2$$

The transformer changes resistances by the *square* of the turns ratio. This will become very significant when we use audio transformers that must match loudspeakers to output valves.

As an example, an output transformer with a primary to secondary turns ratio of 31.6:1 would allow the output valves to see the 8 Ω loudspeaker as an 8 kΩ load, whereas the loudspeaker sees the Thévenin output resistance of the output valves stepped down by an identical amount.

The concept of looking into a device in one direction, and seeing one thing, whilst looking in the opposite direction, and seeing another, is very powerful, and we will use it frequently when we investigate simple amplifier stages.

Practical transformers are not perfect, and we will investigate their imperfections in greater detail in Chapter 4.

Capacitors, Inductors and Reactance

Previously, when we analysed circuits, they were composed purely of resistors and voltage or current sources.

We now need to introduce two new components: *capacitors* and *inductors*. Capacitors have the symbol *C*, and the unit of capacitance is the *farad* (*F*). 1 F is an extremely large capacitance, and more common values range from a few pF to tens of thousands of μF. Inductors have the symbol *L*, and the unit of inductance is the *henry* (*H*). The henry is quite a large unit, and common values range from a few μH to tens of H. Although the henry, and particularly the farad, is rather large for our very specialised use, its size derives from the fundamental requirement for a coherent system of units; a coherent system allows units (such as the farad) to be derived from base units (such as the ampere) with an absolute minimum of scaling factors.

The simplest capacitor is made of a pair of separated plates, whereas an inductor is a coil of wire, and this physical construction is reflected in their graphical symbols (see Figure 1.20).

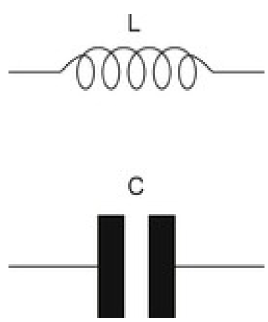


Figure 1.20 Inductor and capacitor symbols.

Resistors had resistance, whereas capacitors and inductors have *reactance*. Reactance is the AC equivalent of resistance – it is still measured in ohms and is given the symbol *X*. We will often have circuits where there is a combination of inductors and capacitors, so it is normal to add a subscript to denote which reactance is which:

$$X_C = \frac{1}{2\pi fC}$$

$$X_L = 2\pi fL$$

Looking at these equations, we find that reactance changes with frequency and with the value of the component. We can plot these relationships on a graph (see Figure 1.21).

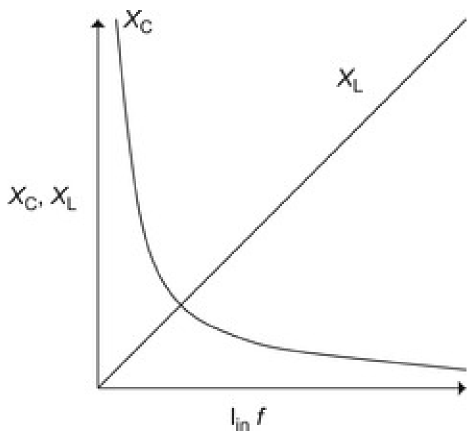


Figure 1.21 Reactance of inductor and capacitor against frequency.

An inductor has a reactance of zero at zero frequency. More intuitively, it is a short circuit at DC. As we increase frequency, its reactance rises.

A capacitor has infinite reactance at zero frequency. It is open circuit at DC. As frequency rises, reactance falls.

A circuit made up of only one capacitor, or one inductor, is not very interesting, and we might want to describe the behaviour of a circuit made up of resistance and reactance, such as a moving coil loudspeaker. See Figure 1.22.

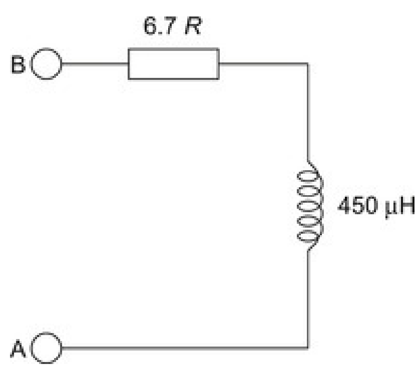


Figure 1.22

We have a combination of resistance and inductive reactance, but at the terminals A and B we see neither a pure reactance, nor a pure resistance, but a combination of the two factors known as *impedance*.

In a traditional electronics book, we would now lurch into the world of vectors, phasors and complex number algebra. Whilst fundamental AC theory is essential for electronics engineers who have to pass examinations, we cannot justify the mental trauma needed to cover the topic in depth, so we will simply pick out useful results that are relevant to our highly specialised field of interest.

Filters

We mentioned that reactance varies with frequency. This property can be used to make a *filter* that allows some frequencies to pass unchecked, whilst others are attenuated (see Figure 1.23).

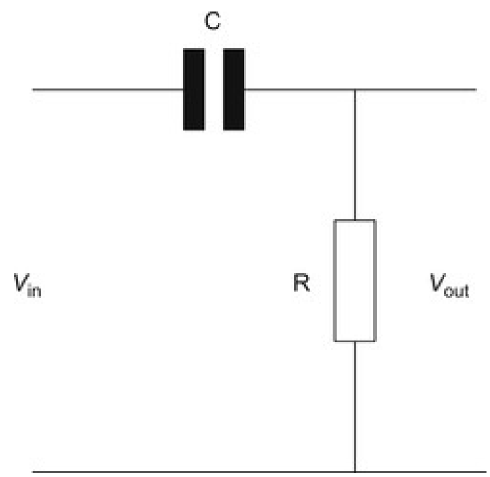


Figure 1.23 *CR* high-pass filter.

All filters are based on potential dividers. In this filter, the upper leg of the potential divider is a capacitor, whereas the lower leg is a resistor. We stated earlier that a capacitor is an open circuit at DC. This filter, therefore, has infinite attenuation at DC – it *blocks* DC. At infinite frequency, the capacitor is a short circuit, and the filter passes the signal with no attenuation, so the filter is known as a *high-pass* filter (see Figure 1.24).

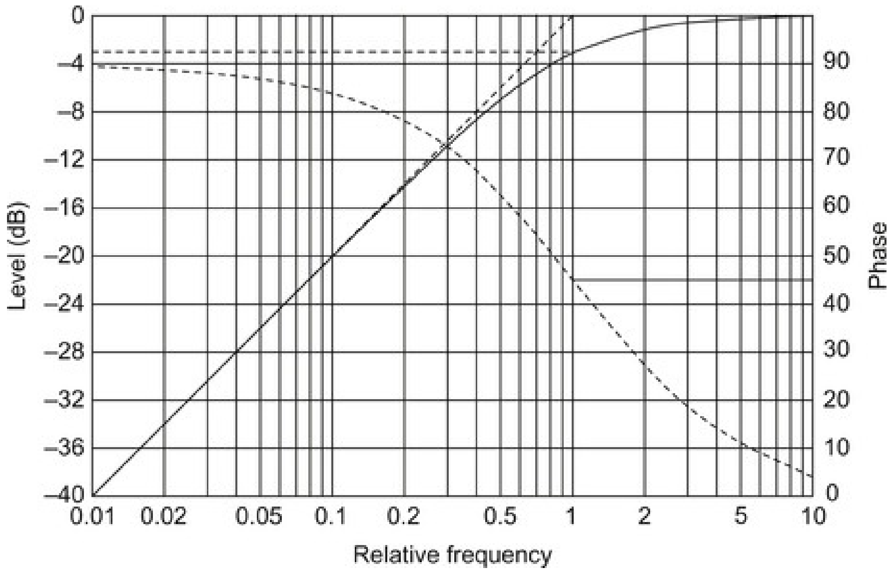


Figure 1.24 Frequency and phase response of the *CR* high-pass filter.

Frequency is plotted on a logarithmic scale to encompass the wide range of values without cramping. When we needed a logarithmic unit for amplitude ratios, we invented the dB, but a logarithmic unit for frequency already existed, so engineers stole the *octave* from the musicians. An octave is simply a halving or doubling of frequency and corresponds to eight ‘white keys’ on a piano keyboard.

The curve has three distinct regions: the *stop-band*, *cut-off* and the *pass-band*.

The stop-band is the region where signals are stopped or attenuated. In this filter, the attenuation is inversely proportional to frequency, and we can see that at a sufficiently low frequency (LF), the shape of the curve in this region becomes a straight line. If we were to measure the slope of this line, we would find that it tends towards 6 dB/octave.

Note that the *phase* of the output signal changes with frequency, with a maximum change of 90° when the curve finally reaches 6 dB/octave.

This slope is very significant, and all filters with only one reactive element have an ultimate slope of 6 dB/octave. As we add more reactive elements, we can achieve a higher slope, so filters are often referred to by their *order*, which is the number of reactive elements contributing to the slope. A third-order filter would have three reactive elements, and its ultimate slope would, therefore, be 18 dB/octave.

Although the curve reaches an ultimate slope, the behaviour at cut-off is of interest, not least because it allows us to say at what frequency the filter begins to take effect. On the diagram, a line was drawn to determine the ultimate slope. If this is extended until it intersects with a similar line drawn continued from the pass-band attenuation, the point of intersection is the filter cut-off frequency. (You will occasionally see idealised filter responses drawn in this way, but this does not imply that the filter response actually changes abruptly from pass-band to stop-band.)

If we now drop a line down to the frequency axis from the cut-off point, it passes through the curve, and the filter response at this point is 3 dB down on the pass-band value. The cut-off frequency is, therefore, also known as $f_{-3\ \mathrm{dB}}$, or the −3 dB point, and at this point the phase curve is at its steepest, with a phase change of 45°.

Second-order filters, and above, have considerable freedom in the way that the transition from pass-band to stop-band is made, and so the *class* of filter is often mentioned in conjunction with names like Bessel, Butterworth and Chebychev, in honour of their originators.

Although we initially investigated a high-pass Capacitor-Resistance (CR) filter, other combinations can be made using one reactive component and a resistor (see Figure 1.25).

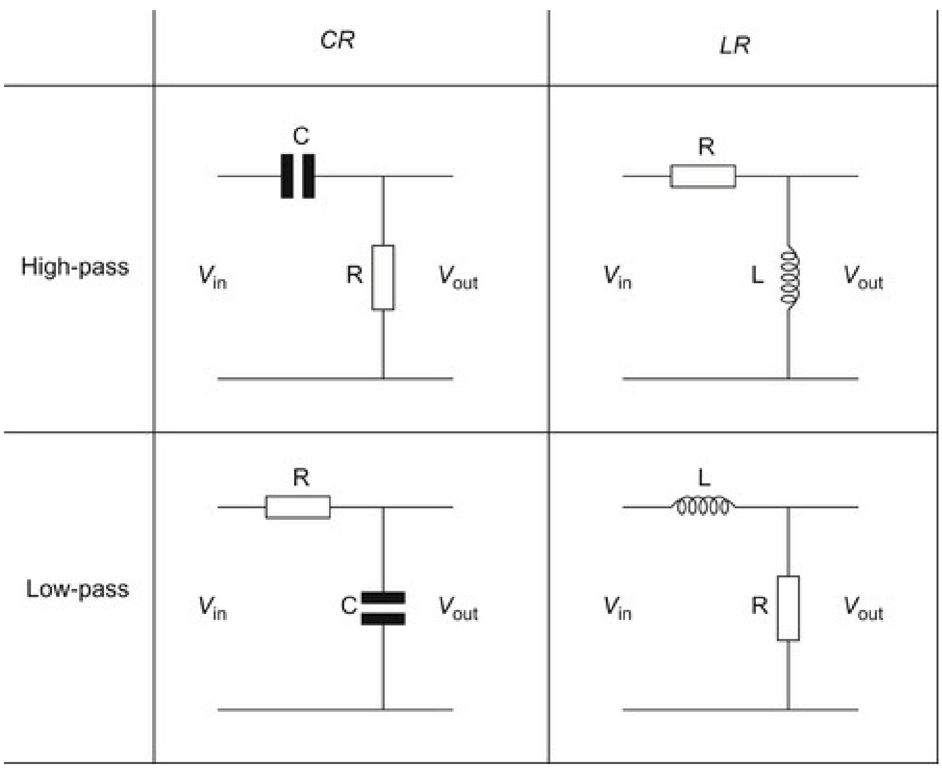


Figure 1.25

We now have a pair of high-pass filters and a pair of *low-pass* filters; the low-pass filters have the same slope, and cut-off frequency can be found from the graph in the same way.

Now that we are familiar with the shape of the curves of these simple filters, referring to them by their cut-off frequency and slope is more convenient (the word ‘ultimate’ is commonly neglected). For these simple filters, the equation for cut-off frequency is the same whether the filter is high-pass or low-pass. For a *CR* filter:

$$f_{-3\,\mathrm{dB}} = \frac{1}{2\pi CR}$$

And for an *LR* filter:

$$f_{-3\,\mathrm{dB}} = \frac{R}{2\pi L}$$

Time Constants

In audio, simple filters or *equalisation* networks are often described in terms of their *time constants*. These have a very specialised meaning that we will touch upon later, but in this context they are simply used as a shorthand form of describing a first-order filter that allows component values to be calculated quickly.

For a *CR* network, the time constant τ (tau) is:

$$\tau = CR$$

For an *LR* network:

$$\tau = \frac{L}{R}$$

Because it is a *time* constant, the unit of τ is seconds, but audio time constants are habitually given in µs. We can easily calculate the cut-off frequency of the filter from its time constant τ:

$$f = \frac{1}{2\pi\tau}$$

Note that τ is quite distinct from period, which is given the symbol T.

Examples of audio time constants:

- FM analogue broadcast High Frequency de-emphasis: UK 50 µs and USA 75 µs
- RIAA vinyl record characteristic: 3,180 µs, 318 µs and 75 µs.

A 75 µs High Frequency de-emphasis circuit needs a low-pass filter, usually *CR*, so a pair of component values whose product equals 75 µs, 1 nF and 75 kΩ would do nicely (see Figure 1.26).

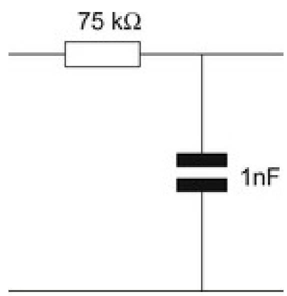


Figure 1.26 A 75 μs de-emphasis network.

Resonance

So far, we have made filters using only one reactive component, but if we make a network using a capacitor *and* an inductor, we find that we have a *resonant* circuit. Resonance occurs everywhere in the natural world, from the sound of a tuning fork to the bucking and twisting of the Tacoma Narrows bridge. (A bridge that finally collapsed during a storm on 7 November 1940 because the wind excited a structural resonance.) A resonant electronic circuit is shown in Figure 1.27.

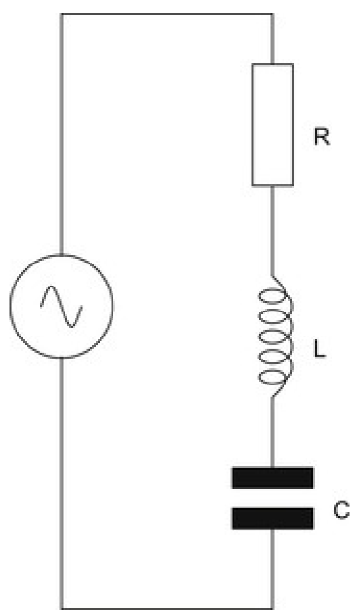


Figure 1.27 Series resonant circuit.

If we were to sweep the frequency of the source, whilst measuring the current drawn, we would find that at the resonant frequency, the current would rise to a maximum determined purely by the resistance of the resistor. The circuit would appear as if the reactive components were not there. We could then plot a graph of current against frequency (see Figure 1.28).

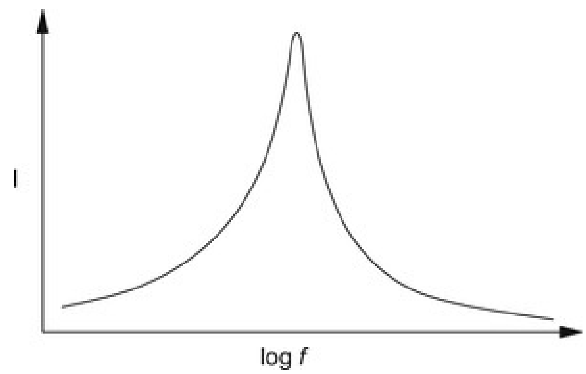


Figure 1.28 Current against frequency for series resonant circuit.

The sharpness and the height of this peak are determined by the *Q* or *magnification factor* of the circuit:

$$Q = \frac{1}{R}\sqrt{\frac{L}{C}}$$

This shows us that a small resistance can cause a high *Q*, and this will be very significant later. The frequency of resonance is:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

Our first resonant circuit was a *series* resonant circuit, but *parallel* resonance, where the *total* current falls to a minimum at resonance, is also possible (see Figure 1.29).

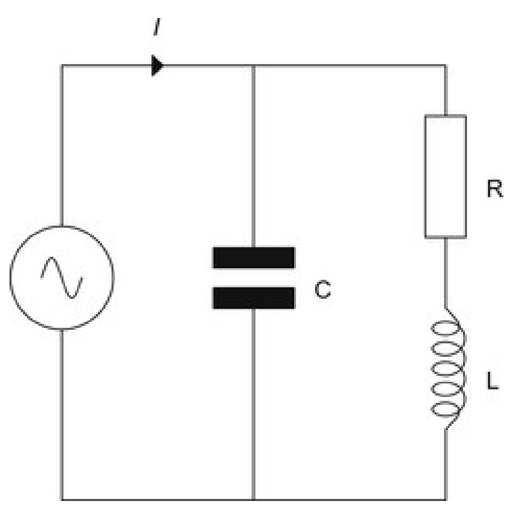


Figure 1.29 Parallel resonant circuit.

If $Q>5$, the above equations are reasonably accurate for parallel resonance. We will not fret about the accuracy of resonant calculations, since we rarely want resonances in audio, and so do our best to remove or damp them.

RMS and Power

We mentioned power earlier, when we investigated the flow of current through a lamp using a 240 V battery. Mains electricity is AC and has recently been respecified in the UK to be 230 VAC +10%−6% at 50±1 Hz, but how do we define that 230 V?

If we had a valve heater filament, it would be most useful if it could operate equally well from AC or DC. As far as the valve is concerned, AC electricity will heat the filament equally well, just so long as we apply the correct voltage.

> *The RMS voltage of any waveform is equivalent to the DC voltage having the same heating effect as the original waveform.*

RMS is short for Root of the Mean of the Squares, which refers to the method of calculating the value. Fortunately, the ratios of V_{RMS} to V_{peak} have been calculated for the common waveforms, and in audio design we are mostly concerned with the sine wave, for which:

$$V_{\mathrm{peak}} = \sqrt{2} \cdot V_{\mathrm{RMS}}$$

All sinusoidal AC voltages are given in V_{RMS} unless specified otherwise, so a heater designed to operate on 6.3 VAC would work equally well connected to 6.3 VDC.

We have only mentioned RMS voltages, but we can equally well have RMS currents, in which case:

$$P = V_{\mathrm{RMS}} \cdot I_{\mathrm{RMS}}$$

> There is no such thing as an RMS watt! Please refer to the definition of RMS.

The Square Wave

Until now, all of our dealings have been with sine waves, which are pure tones. When we listen to music, we do not hear pure tones, instead we hear a *fundamental* with various proportions of *harmonics* whose frequencies are arithmetically related to the frequency of the fundamental. We are able to distinguish between one instrument and another because of the differing proportions of the harmonics and the *transient* at the beginning of each note.

A useful waveform for testing amplifiers quickly would have many harmonics and a transient component. The square wave has precisely these properties because it is composed of a fundamental frequency plus odd harmonics whose amplitudes steadily decrease with frequency (see Figure 1.30).

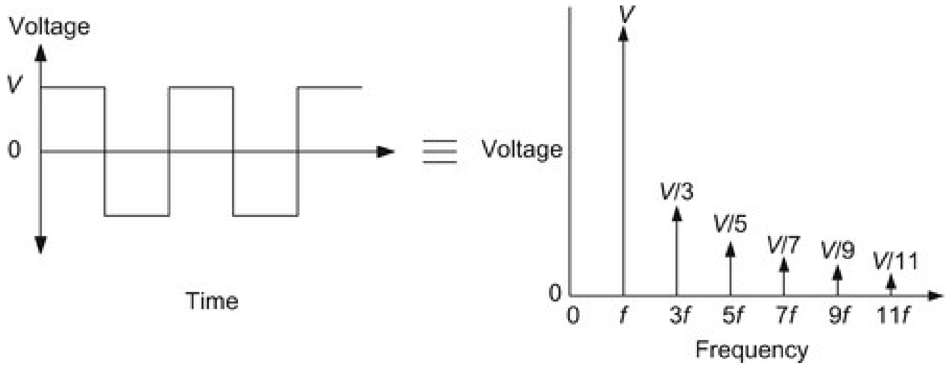


Figure 1.30 Square wave viewed in time and frequency.

A square wave is thus an infinite series of harmonics, all of which must be summed from the fundamental to the infinite harmonic. We can express this argument mathematically as a *Fourier* series, where *f* is the fundamental frequency:

$$\text{square wave} = \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{\sin[(2n-1)\cdot 2\pi ft]}{2n-1}$$

This is a shorthand formula, but understanding the distribution of harmonics is much easier if we express them in the following form:

$$\text{square wave} \propto 1(f) + \frac{1}{3}(3f) + \frac{1}{5}(5f) + \frac{1}{7}(7f) + \frac{1}{9}(9f) + \ldots$$

We can now see that the harmonics die away very gradually and that a 1 kHz square wave has significant harmonics well beyond 20 kHz. What is not explicitly stated by these formulae is that the relative phase of these components is critical. The square wave thus tests not only amplitude response, but also phase response.

Square Waves and Transients

We briefly mentioned earlier that the square wave contained a transient component. One way of viewing a square wave is to treat it as a DC level whose polarity is inverted at regular intervals. At the instant of inversion, the voltage has to change instantaneously from its negative level to its positive level, or vice versa. The abrupt change at the leading edge of the square wave is the transient, and because it occurs so quickly, it must contain a high proportion of high frequency components. Although we already knew that the square wave contained these high frequencies, it is only at the leading edge that they all sum constructively, so any change in high frequency response is seen at this leading edge.

We have considered the square wave in terms of frequency; now we will consider it as a series of transients in *time*, and investigate its effect on the behaviour of *CR* and *LR* networks.

The best way of understanding this topic is with a mixture of intuitive reasoning coupled to a few graphs. Equations *are* available, but we very rarely need to use them.

In electronics, a *step* is an *instantaneous* change in a quantity such as current or voltage, so it is a very useful theoretical concept for exploring the response of circuits to transients. We will start by looking at the voltage across a capacitor when a voltage step is applied via a series resistor (see Figure 1.31).

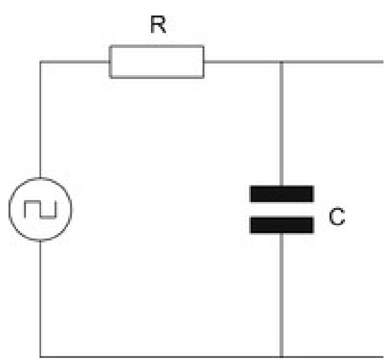


Figure 1.31

The capacitor is initially discharged (V_C=0). The step is applied and switches from 0 V to +V, an instantaneous change of voltage composed mostly of high frequencies. The capacitor has a reactance that is inversely proportional to frequency and, therefore, appears as a short circuit to these high frequencies. If it is a short circuit, we cannot develop a voltage across it. The resistor, therefore, has the full applied voltage across it and passes a current determined by Ohm's law. This current then flows through the capacitor and starts charging it. As the capacitor charges, its voltage rises, until eventually it is fully charged, and no more current flows. If no more current flows into the capacitor (I_C=0), then I_R=0, and so V_R=0. We can plot this argument as a pair of graphs showing capacitor and resistor voltage (see Figure 1.32).

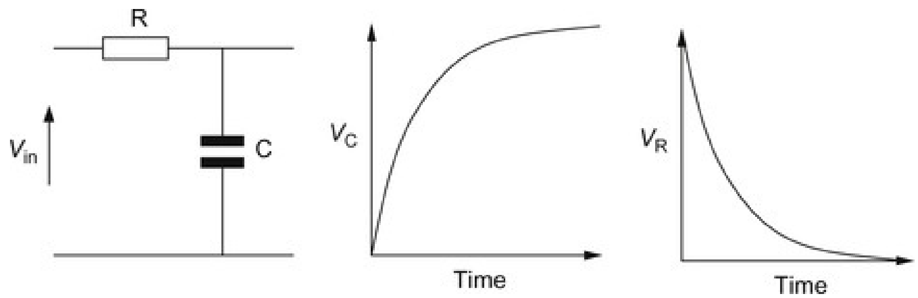


Figure 1.32 Exponential response of *CR* circuit to voltage step.

The first point to note about these two graphs is that the shape of the curve is an *exponential* (this term will be explained further later). The second point is that when we apply a step to a *CR* circuit (and even an *LR* circuit), the current or voltage curve will *always* be one of these curves. Knowing that the curve can only be one of these two possibilities, all we need to be able to do is to choose the appropriate curve.

The transient edge can be considered to be of infinitely High Frequency, and the capacitor is, therefore, a short circuit. Developing a voltage across a short circuit requires infinite current.

> Infinite current is required to instantaneously change the voltage across a capacitor.

An inductor is the *dual* or inverse of a capacitor, and so an inductor has a similar rule.

> An instantaneous change of current through an inductor creates an infinite voltage.

We can now draw graphs for each of the four combinations of *CR* and *LR* circuits when the same step in voltage is applied (see Figure 1.33).

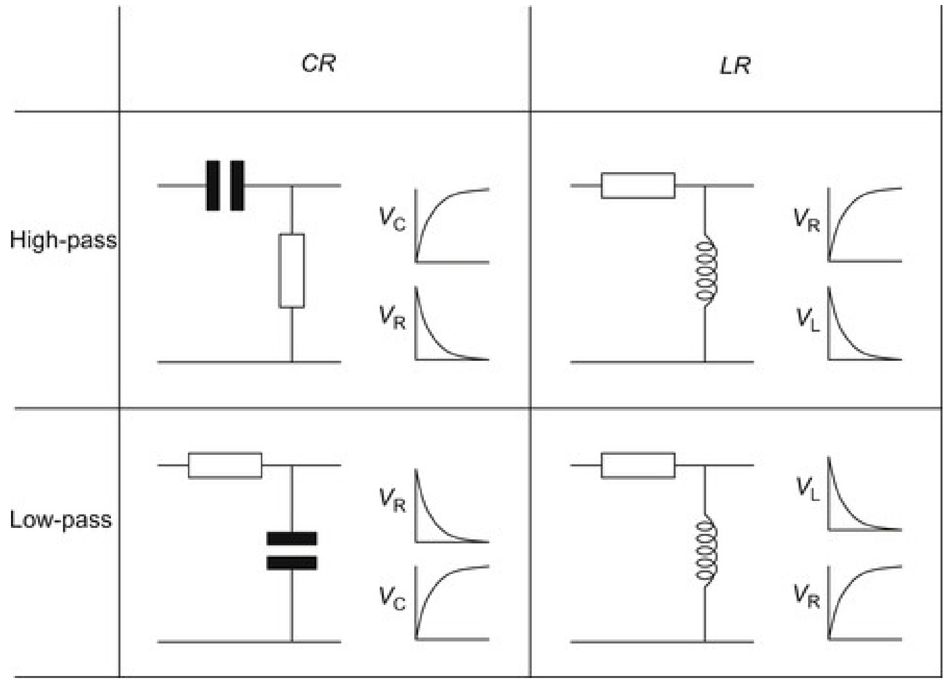


Figure 1.33

Having stated that the shape of the curves in each of the four cases is identical, we can now examine the fundamental curves in a little more detail.

Using the original *CR* circuit as an example, the capacitor will eventually charge to the input voltage, so we can draw a dashed line to represent this voltage. The voltage across the capacitor has an initial slope, and if we continue this slope with another dashed line, we will find that it intersects the first line at a time that corresponds to *CR*, which is the *time constant* that we met earlier. The *CR* time constant is defined as the time taken for the capacitor voltage to reach its final value had the initial rate of charge been maintained (see Figure 1.34).

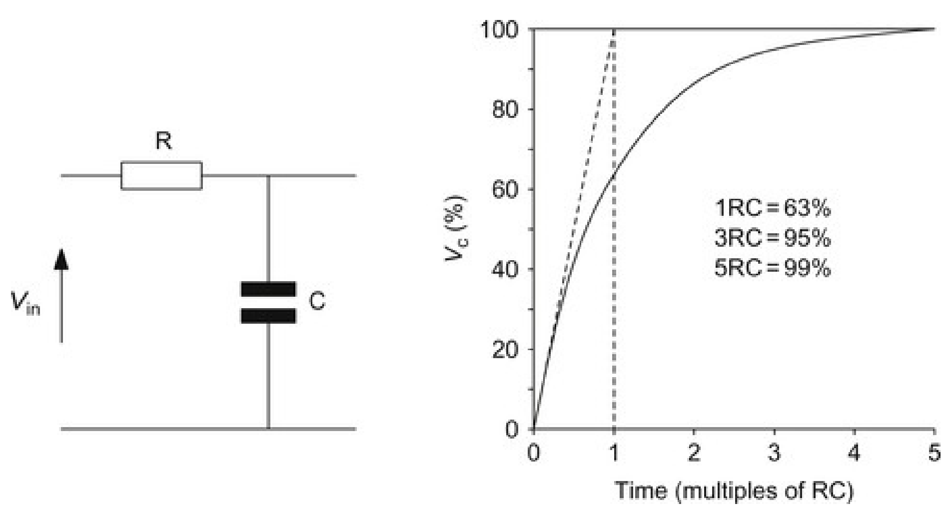


Figure 1.34 τ and its significance to the exponential curve.

The equation for the falling curve is:

$$v = V \cdot e^{t/\tau}$$

The equation for the rising curve is:

$$v = V \cdot (1 - e^{t/\tau})$$

where ‘e’ is the base of natural logarithms and is the key marked ‘e ^{x}’ or ‘exp’ on your scientific calculator. These curves derive their names because they are based on an exponential function.

We could now find what voltage the capacitor actually achieved at various times. Using the equation for the rising curve:

1 τ	63%
3 τ	95%
5 τ	99%

Because the curves are all the same shape, these ratios apply to all four of the *CR* and *LR* combinations. The best way to use the ratios is to first decide which way the curve is heading; the ratios then determine how much of the *change* will be achieved, and in what time.

Note that after 5 τ, the circuit has very nearly reached its final position or steady state. This is a useful point to remember when considering what will happen at switch-on to high voltage semiconductor circuits.

When we considered the response to a single step, the circuit eventually achieved a steady state because there was sufficient time for the capacitor to charge or for the inductor to change its magnetic field. With a square wave, this may no longer be true. As was mentioned before, the square wave is an excellent waveform for testing audio amplifiers, not least because oscillators that can generate both sine and square waves are fairly cheaply available.

If we apply a square wave to an amplifier, we are effectively testing a *CR* circuit made up of a series resistance and a capacitance to ground (often known as a shunt capacitance). We should, therefore, expect to see some rounding of the leading edges, because some of the high frequencies are being attenuated.

If the amplifier is only marginally stable (because it contains an unwanted resonant circuit), the high frequencies at the leading edges of the square wave will excite the resonance, and we may see a damped train of oscillations following each transition.

We can also test low frequency response with a square wave. If the coupling capacitors between stages are small enough to change their charge noticeably within one half-cycle of the square wave, then we will see *tilt* on the top of the square wave. Downward tilt, more commonly called *sag*, indicates low frequency loss, whilst upward tilt indicates low frequency boost. This is a very sensitive test of low frequency response, and if it is known that the circuit being measured includes a *single* high-pass filter, but with a cut-off frequency too low to be measured directly with sine waves, then a square wave may be used to infer the sine wave $f_{-3\ \mathrm{dB}}$ point. The full derivation of the equation that produced the following table is given in the Appendix, but if we apply a square wave of frequency f (Table 1.1).

Table 1.1 Relating Square Wave Sag to Low Frequency Cut-Off Frequency

Sag observed using a square wave of frequency f (%)	Ratio of applied square wave frequency (f) to low frequency cut-off ($f_{-3\ \mathrm{dB}}$)
10	30
5	60
1	300

Most analogue audio oscillators are based on the Wien bridge and, because of amplitude stabilisation problems, rarely produce frequencies lower than ≈10 Hz. 10% and 5% square wave sag may be measured relatively easily on an oscilloscope and can, therefore, be used to infer sine wave performance at frequencies below 10 Hz.

Another useful test is to apply a high-level, High Frequency *sine* wave. If at all levels and frequencies, the output is still a sine wave, then the amplifier is likely to be free of slewing distortion. If the output begins to look like a triangular waveform, this is because one, or more, of the stages within the amplifier is unable to fully charge or and discharge a shunt capacitance sufficiently quickly. The distortion is known as slewing distortion because the waveform is unable to slew correctly from one voltage to another. The solution is usually to increase the anode current of the offending stage, thereby enabling it to charge or discharge the capacitance.

Random Noise

The signals that we have previously considered have been repetitive signals – we could always predict precisely what the voltage level would be at any given time. Besides these *coherent* signals, we shall now consider *noise*.

Noise is all around us, from the sound of waves breaking on a seashore, and the radio noise of stars, to the daily fluctuations of the stock markets. Electrical noise can generally be split into one of two categories: *white* noise, which has constant level with frequency (like white light), and *1/f* noise, whose amplitude is inversely proportional to frequency.

White noise is often known as Johnson or thermal noise and is caused by the random thermal movement of atoms knocking the free electrons within a conductor. Because it is generated by a thermal mechanism, cooling critical devices reduces noise, so the input transistor associated with a High Purity Germanium (HPGe) detector used for gamma radiation spectroscopy is also cooled by the liquid nitrogen (77 K) needed to reduce leakage currents in the crystal, but some radio telescopes need the much colder liquid helium (4 K) for their head amplifiers. All resistors produce white noise and generate a noise voltage:

$$v_{\text{noise}} = \sqrt{4kTBR}$$

where

k=Boltzmann's constant ≈1.381×10^{-23} J/K

T=absolute temperature of the conductor ≈°C+273.16

B=bandwidth of the following measuring device

R=resistance of the conductor.

From this equation, we can see that if we were to cool the conductor to 0 K or −273.16 °C, there would be no noise because this would be absolute zero, at which temperature there is no thermal vibration of the atoms to produce noise.

The *bandwidth* of the measuring system is important too because the noise is proportional to the square root of bandwidth. Bandwidth is the difference between the upper and the lower $f_{-3\ \text{dB}}$ limits of measurement. It is important to realise that in audio work, the noise measurement bandwidth is always that of the human ear (20 Hz to 20 kHz), and although one amplifier might have a wider bandwidth than another, this does not necessarily mean that it will produce any more noise.

In audio, we cannot alter the noise bandwidth or the value of Boltzmann's constant and reducing the temperature is expensive, so our main weapon for reducing noise is to reduce resistance. We will look into this in more detail in Chapter 3 and Chapter 7.

$1/f$ noise is also known as *flicker* noise or *excess* noise, and it is a particularly insidious form of noise because it is not predictable. It could almost be called 'imperfection' noise because it is generally caused by imperfections such as imperfectly 'clean rooms' used for making semiconductors or valves, 'dry' soldered joints, poor metal-to-metal contacts in connectors – the list is endless. Semiconductor manufacturers usually specify the $1/f$ *noise corner*, where the $1/f$ noise becomes dominant over white noise, for their devices, but equivalent data

do not exist for valves.
Because noise is random, or *uncorrelated*, we cannot add noise voltages or currents, but must add noise *powers*, and some initially surprising results emerge. Noise can be considered statistically as a deviation from a mean value. When an opinion poll organisation uses as large a sample as possible to reduce error, it is actually *averaging* the noise to find the mean value.
If we parallel *n* input devices in a low-noise amplifier, the uncorrelated noise sources begin to cancel, but the wanted signal remains at constant level, resulting in an improvement in signal to noise ratio of $\sqrt{n}$ dB. This technique is feasible for semiconductors where it is possible to make 100 matched parallelled transistors on a single chip (LM394, MAT-01, etc.), but we are lucky to find a pair of matched triodes in one envelope, let alone more than that!

Active Devices

We have investigated resistors, capacitors, inductors and transformers, but these were all *passive* components. We will now look at *active* devices, which can *amplify* a signal. All active devices need a power supply because amplification is achieved by the source controlling the flow of energy from a power supply into a load via the active device.
We will conclude this chapter by looking briefly at semiconductors. It might seem odd that we should pay any attention at all to semiconductors, but a modern valve amplifier generally contains rather more semiconductors than valves, so we need some knowledge of these devices to assist the design of the (valve) amplifier.

Conventional Current Flow and Electron Flow

When electricity was first investigated, the electron had not been discovered, and so an arbitrary direction for the flow of electricity was assumed. There was a 50/50 chance of guessing correctly, and the early researchers were unlucky. By the time the mistake was discovered and it was realised that electrons flowed in the opposite direction to the way that electricity had been thought to flow, it was too late to change the convention.
We are, therefore, saddled with a conventional current that flows in the *opposite* direction to that of the electrons. Mostly, this is of little consequence, but when we consider the internal workings of the transistor and the valve, we must bear this distinction in mind.

Silicon Diodes

Semiconductor devices are made by doping regions of crystalline silicon to form areas known as *N-type* or *P-type*. These regions are permanently charged and at their junction this charge forms a potential barrier that must be overcome before forward conduction can occur. Reverse polarity strengthens the potential barrier, so no conduction occurs.

The *diode* is a device that allows current to flow in one direction, but not the other. Its most basic use is, therefore, to *rectify* AC into DC. The arrow-head on the diode denotes the direction of conventional current flow, and R_L is the load resistance (see Figure 1.35).

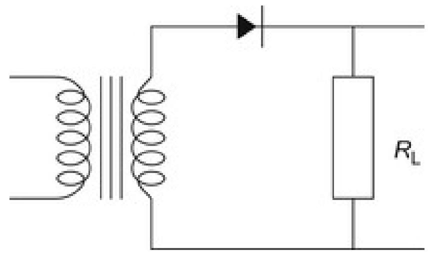


Figure 1.35 Use of a diode to rectify AC.

Practical silicon diodes are not perfect rectifiers and require a forward *bias* voltage before they conduct any appreciable current. At room temperature, this bias voltage is between 0.6 V and 0.7 V (see Figure 1.36).

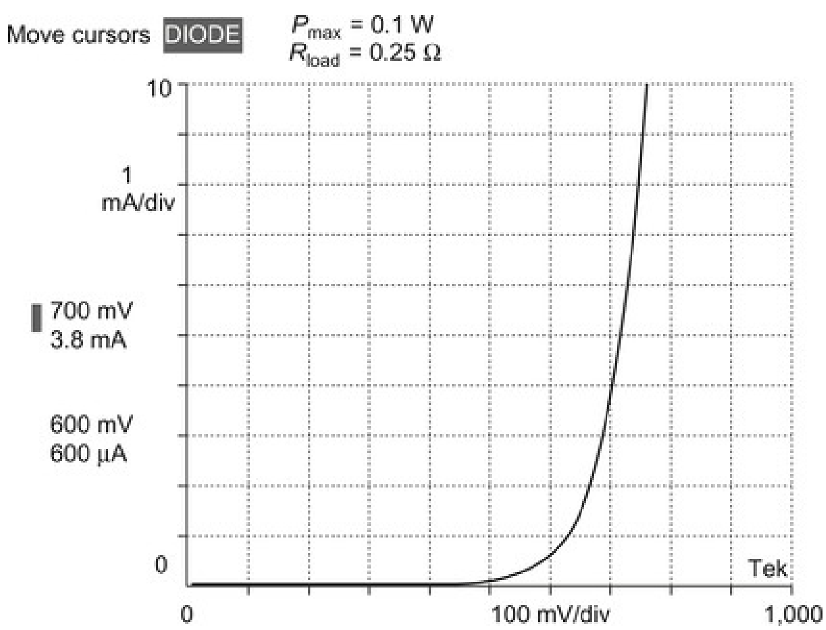


Figure 1.36 Silicon diode current against forward bias voltage.

This forward bias voltage is always present, so the output voltage is always less than the input voltage by the amount of the *diode drop*. Because there is always a voltage drop across the diode, current flow must create heat, and sufficient heat will melt the silicon. All diodes, therefore, have a maximum current rating.

In addition, if the reverse voltage is too high, the diode will break down and conduct; if this reverse current is not limited, the diode will be destroyed.

Unlike valves, the mechanism for conduction through the most common (bipolar junction) type of silicon diode is complex and results in a charge being temporarily stored within the diode. When the diode is switched off by the external voltage, the charge within the diode is quickly discharged and produces a brief current overshoot that can excite external resonances. Fortunately, *Schottky* diodes do not exhibit this phenomenon, and *soft recovery* types are fabricated to minimise it.

Voltage References

The forward voltage drop across a diode junction is determined by the Ebers/Moll equation involving absolute temperature and current, whilst the reverse breakdown voltage is determined by the physical construction of the individual diode. This means that we can use the diode as a rectifier or as a voltage reference.

Voltage references are also characterised by their *slope resistance*, which is the Thévenin resistance of the voltage reference when operated correctly. It does not imply that large currents can be drawn, merely that for small current changes in the linear region of operation, the voltage change will be correspondingly small.

Voltage references based on the forward voltage drop are known as *bandgap* devices, whilst references based on the reverse breakdown voltage are known as *Zener* diodes. All voltage references should pass only a limited current to avoid destruction, and ideally a constant current should be passed.

Zener diodes are commonly available in power ratings up to 75 W, although the most common rating is 400 mW, and voltage ratings are from 2.7 V to 270 V. In reality, true Zener action occurs at ≤5 V, so a 'Zener' diode with a rating ≥5 V actually uses the avalanche effect [2]. This is fortuitous because producing a 6.2 V device requires both effects to be used in proportions that cause the two different mechanisms to cancel the temperature coefficient almost to zero and reduce slope resistance. Diodes such as the 1N82* series deliberately exploit this phenomenon to produce a 6.2 V reference with vanishingly low temperature coefficient (0.0005%/°C for the 1N829A), but note that this performance only obtains if a Zener current of 7.5 ± 0.01 mA is used. Slope resistance rises sharply below 6 V, and more gradually above 6 V, but is typically ≈10 Ω at 5 mA for a 6.2 V Zener.

Bandgap references are often actually a complex integrated circuit and usually have an output voltage of 1.2 V, but internal amplifiers may increase this to 10 V or more. Because they are complex internally, bandgap references tend to be more expensive than Zeners, and we may occasionally need a cheap low-noise reference. Light Emitting Diodes (LEDs) and small-signal diodes are operated forward biassed, so they are quiet and reasonably cheap (Table 1.2).

Table 1.2 Comparison of Forward Drops and Slope Resistances for Various Diodes

Diode type	Typical forward drop at 10 mA (V)	Typical r_{slope} at 10 mA (Ω)
Small-signal silicon diode (1N4148)	0.75	6.0
Infrared LED (950 nm)	1.2	5.4
Cheap red LED	1.7	4.3
Cheap yellow, yellow/green LED	≈2	10
True green LED (525 nm)	3.6	30
Blue LED (426 nm)	3.7	26

The Agilent HLMP6000 red LED is particularly useful as a low-noise voltage reference (see Figure 1.37).

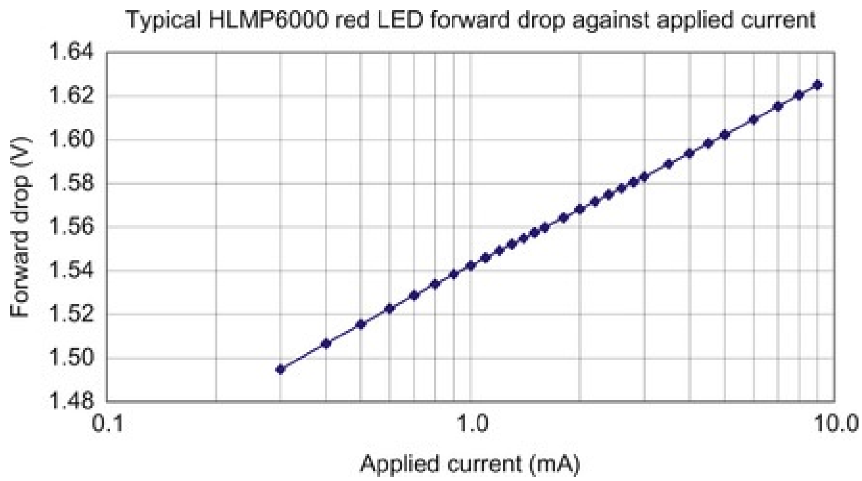


Figure 1.37 HLMP6000 red LED forward drop against applied current.

A straight line on a logarithmic scale implies a logarithmic equation:

$$V = 0.0378 \ln I_{DC(mA)} + 1.5418$$

More significantly, differentiating the equation gives us d V/d I, which we know as slope resistance:

$$r_{slope(\Omega)} = \frac{37.8}{I_{DC(mA)}}$$

Thus, a typical HLMP6000 passing 10 mA has a slope resistance of 3.8 Ω. Note that this is an example of an equation where it was more useful to use the scaled unit (mA) than the base unit (A).

Bandgap references usually incorporate an internal amplifier, so their output resistance is much lower, typically stated as ≈0.2 Ω or less, but this rises with frequency. So if low resistance must be maintained to high frequencies (>1 MHz), the HLMP6000 may be a better choice.

Bipolar Junction Transistors (BJTs)

BJTs are the most common type of transistor; they are available in NPN and PNP types and can be used to amplify a signal. The name transistor is derived from *trans*ferred res *istor*. Before the appellation 'transistor' became commonplace, they were rather charmingly known as crystal triodes (see Figure 1.38).

Figure 1.38 The GET1 – one of the very first transistors.

We can imagine an NPN transistor as a sandwich of two thick slabs of N-type material separated by an extremely thin layer of P-type material. The P-type material is the *base*, whilst one of the N-types is the *emitter* and emits electrons, which are then collected by the other N-type, which is known as the *collector*.

If we simply connect the collector to the positive terminal of a battery and the emitter to the negative, no current will flow because the negatively charged base repels electrons. If we now apply a positive voltage to the base to neutralise this charge, the electrons will no longer be repelled, but because the base is so thin, the attraction of the strongly positively charged collector pulls most of the electrons straight through the base to the collector, and collector current flows.

The base/emitter junction is now a forward biassed diode, so it should come as no surprise to learn that 0.7 V is required across the base/emitter junction to cause the transistor to conduct electrons from emitter to collector. Because the base is so thin and the attraction of the collector is so great, very few electrons emerge from the base as base current, so the ratio of collector current to base leakage current is high. The transistor, therefore, has current gain, which is sometimes known as β (beta) but more commonly as h_{FE} for DC current gain or h_{fe} for AC current gain, although for all practical purposes, $h_{FE}= h_{fe}$, making the AC distinction trivial. (h_{FE}: *h*ybrid model, *f*orward current transfer ratio and *e*mitter as common terminal. Don't you wish you hadn't asked?)

Realistically, h_{FE} is a defect, not a parameter, and should be treated as such. It is not constant between samples and varies as device parameters change (see Figure 1.39).

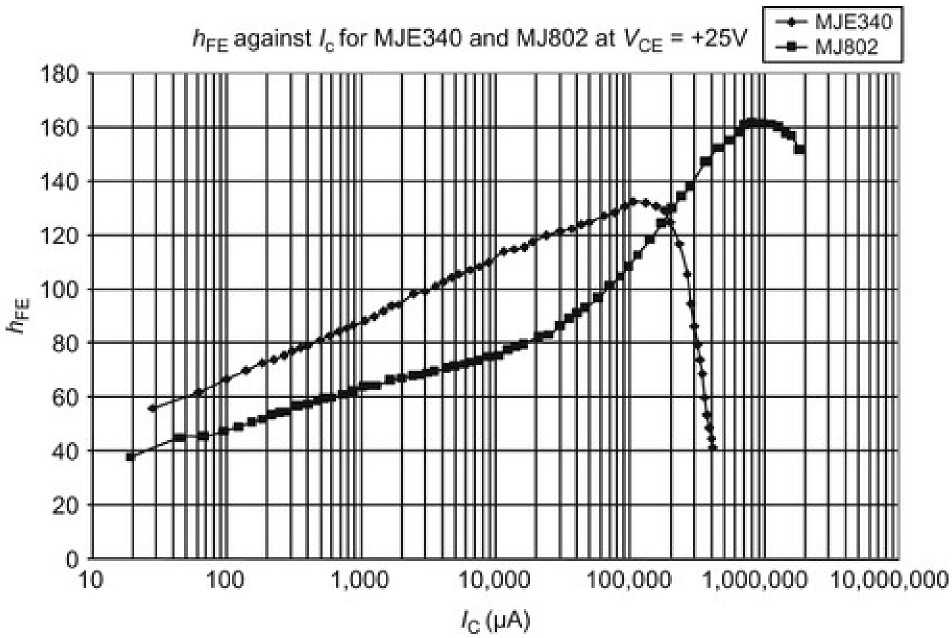


Figure 1.39 Variation of the defect h_{fe}.

As can be seen, there is a general trend for h_{FE} to rise with increasing collector current but fall sharply as $I_{C(max)}$ is approached. The MJ802 is a special audio power transistor that has been engineered to maintain h_{FE} at high currents, and this accounts for the kink at 40 mA.

A far more important and predictable parameter is the *transconductance* (more usefully called *mutual conductance* in valves) g_m, which is the change in collector current caused by a change in base/emitter voltage:

$$g_m = \frac{\Delta I_C}{\Delta V_{BE}}$$

When we look at valves, we will see that we must always measure g_m at the operating point, perhaps from a graph. For transistors, transconductance is defined by the Ebers/Moll [3] equation, and for small currents (less than ≈100 mA), g_m can be estimated for any BJT using:

$$g_{\mathrm{m}} \approx 35 I_{\mathrm{C}}$$

The Common Emitter Amplifier

Now that we have a means of predicting the change in collector current caused by a change in V_{BE}, we could connect a resistor R_L in series with the collector and the supply to convert the current change into a voltage change. By Ohm's law:

$$\Delta V_{\mathrm{CE}} = \Delta I_C \cdot R_{\mathrm{L}}$$

But $\Delta I_C = g_m \cdot \Delta V_{BE}$, so:

$$\Delta V_{\mathrm{CE}} = \Delta V_{\mathrm{BE}} \cdot g_{\mathrm{m}} \cdot R_{\mathrm{L}}$$

$$\frac{\Delta V_{\mathrm{CE}}}{\Delta V_{\mathrm{BE}}} = A_v = g_{\mathrm{m}} \cdot R_{\mathrm{L}}$$

We are now able to find the voltage amplification A_v (also known as gain) for this circuit, which is known as a common emitter amplifier because the emitter is common to both the input and the output circuits or *ports*. Note that the output polarity of the amplifier is *inverted* with respect to the input signal (see Figure 1.40).

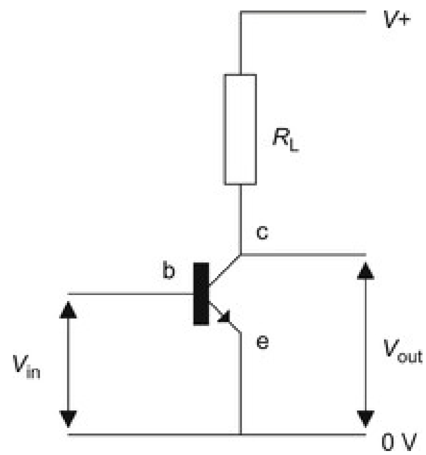


Figure 1.40 Common emitter transistor amplifier.

As the circuit stands, it is not very useful because any input voltage below +0.7 V will not be amplified, so the amplifier creates considerable distortion.

To circumvent this, we assume that the collector voltage is set to half the supply voltage because this allows the collector to swing an equal voltage both positively and negatively. Because we know the voltage across the collector load R_L and its resistance, we can use Ohm's law to determine the current through it – which is the same as the transistor collector current. We could then use the relationship between I_B and I_C to set this optimum collector current. As we saw earlier, the value of h_{FE} for a transistor is not guaranteed and varies widely between devices. For a small-signal transistor, it could range from 50 to more than 400. The way around this is to add a resistor in the emitter and to set a base *voltage* (see Figure 1.41).

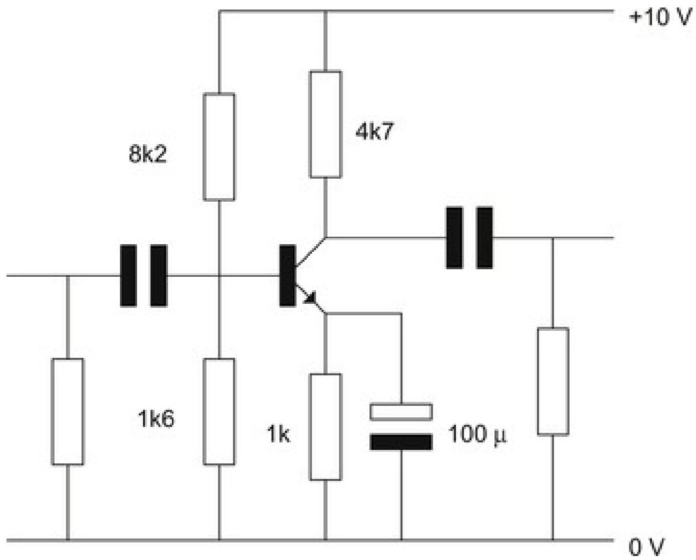


Figure 1.41 Stabilised common emitter amplifier.

The amplifier now has input and output *coupling* capacitors and an emitter *decoupling* capacitor. The entire circuit is known as a stabilised common emitter amplifier and is the basis of most linear transistor circuitry.

Considering DC Conditions

The potential divider chain passes a current that is at least 10 times the expected base current and therefore sets a fixed voltage at the base of the transistor independent of base current. Because of diode drop, the voltage at the emitter is thus 0.7 V lower. The emitter resistor has a *fixed* voltage across it, and it must, therefore, pass a fixed current from the emitter. $I_E = I_C - I_B$, but since I_B is so small, $I_E \approx I_C$, and if we have a fixed emitter current, then collector current is also fixed.

The emitter is decoupled to prevent *negative feedback* from reducing the AC gain of the circuit. We will consider negative feedback later in this chapter.

Briefly, if the emitter resistor was not decoupled, then any change in collector current (which is the same as emitter current) would cause the voltage across the emitter resistor to change with the applied AC. The emitter voltage would change, and v_{be} would effectively be reduced, causing AC gain to fall. The decoupling capacitor is a short circuit to AC and, therefore, prevents this reduction of gain. This principle will be repeated in the next chapter when we look at the cathode bypass capacitor in a valve circuit.

Input and Output Resistances

In a valve amplifier, we will frequently use transistors as part of a bias network or as part of a power supply, and being able to determine input and output resistances is, therefore, important. The following AC resistances looking into the transistor do not take account of any external parallel resistance from the viewed terminal to ground.

The output resistance $1/h_{oe}$ looking into the collector is high, typically tens of kΩ at $I_C \approx 1$ mA, which suggests that the transistor would make a good constant current source (h_{oe}: *h*ybrid model, *o*utput admittance and *e*mitter as common terminal). In theory, if all the collector curves of a bipolar junction transistor are extended negatively, they intersect at a voltage known as the Early [4] voltage. Sadly, the real world is not so tidy, and attempts by the author to determine Early voltages resembled the result of dropping a handful of uncooked spaghetti on the graph paper. Nevertheless, the concept of Early effect is useful because it indicates that $1/h_{oe}$ falls as I_C rises (see Figure 1.42).

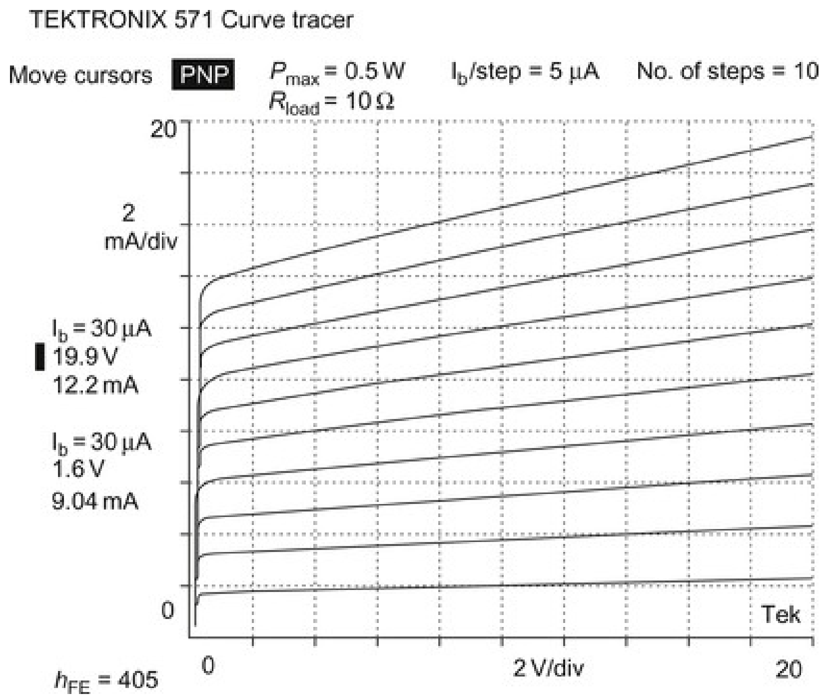


Figure 1.42 Collector curves for BC558B PNP transistor.

The Early effect is caused by the depletion region straddling the reverse-biassed base/collector junction. As V_{CB} rises, the depletion region widens, effectively making the base narrower and less able to capture electrons; this increases h_{fe} and results in increased collector current at higher collector voltages.

The resistance looking into the emitter is low, $r_e=1/g_m$, so ≈20 Ω is typical. If the base is *not* driven by a source of zero resistance ($R_b\neq0$), there is an additional series term, and r_e is found from:

$$r_e = \frac{1}{g_m} + \frac{R_b}{h_{fe}}$$

where R_b is the Thévenin resistance of all the paths to ground *and supply* seen from the base of the transistor. Note that even though we are no longer explicitly including a battery as the supply, the supply is still assumed to have zero output resistance from DC to light frequencies.

Looking into the base, the path to ground is via the base emitter junction in series with the emitter resistor. If the emitter resistor is not decoupled, then the resistance will be:

$$r_b = h_{fe}\left(\frac{1}{g_m} + R_e\right)$$

If the emitter resistor is decoupled, then R_e=0, and the equation reduces to:

$$r_b = \frac{h_{fe}}{g_m} = h_{ie}$$

The AC input resistance owing purely to the base/emitter junction is often known as h_{ie} and is generally quite low, <10 kΩ (h_{ie}: *h*ybrid model, *i*nput resistance and *e*mitter as common terminal).

If we now consider the effect of the external parallel resistances, we see that the input resistance of the amplifier is low, typically <5 kΩ. The output resistance seen at the emitter is low, typically <100 Ω (even if the source resistance is quite high), and the output resistance at the collector is ≈ R_L.

The Emitter Follower

Very occasionally, you will see this amplifier called a *common collector* amplifier, although this phraseology is rare because it does not convey clearly what the circuit does.

If we reduce the collector load to zero and take our output from the emitter, then we have an amplifier with $A_v \approx 1$. The voltage gain *must* be ≈1 because $V_E = V_B - 0.7$ V; the emitter *follows* the base voltage, and the amplifier is *non-inverting*. Although $A_v = 1$, the current gain is much greater, and we can calculate input and output resistances using the equations presented for the common emitter amplifier (see Figure 1.43).

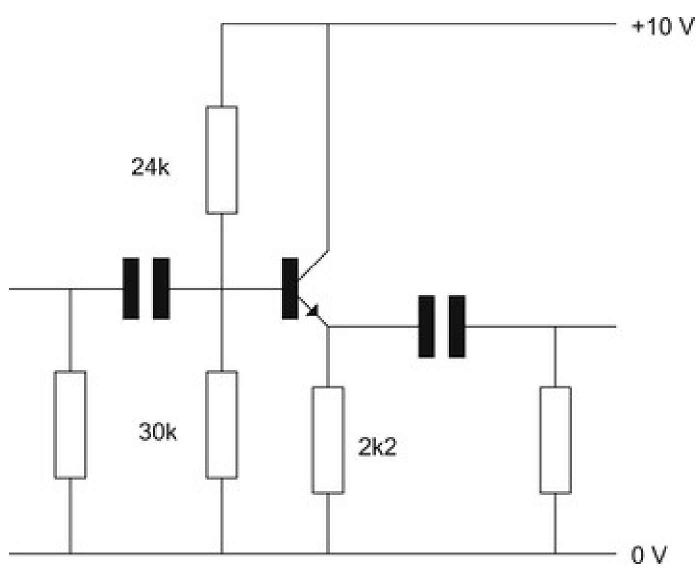


Figure 1.43 Emitter follower.

Because of its low output resistance and moderately high input resistance, the emitter follower is often used as a *buffer* to match high-impedance circuitry to low-resistance loads.

The Darlington Pair

Sometimes, even an emitter follower may not have sufficient current gain, and the solution is to use a *Darlington pair*; this is effectively two transistors in cascade, with one forming the emitter load for the other (see Figure 1.44).

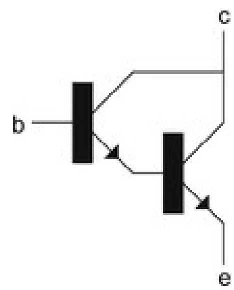


Figure 1.44 Darlington pair.

The two transistors form a composite transistor with V_{BE}=1.4 V and $h_{FE\ (total)} = h_{FE1} \times h_{FE2}$. A Darlington pair can replace a single transistor in any configuration if it seems useful. Common uses are in the output stage of a power amplifier and in linear power supplies. Darlingtons can be bought in a single package, but making your own out of two discrete transistors usually gives better performance and is cheaper in terms of component cost, although putting two parts instead of one on a Printed Circuit Board (PCB) is more expensive.

General Observations on BJTs

We mentioned earlier that the BJT could be considered to be a sandwich with the base separating the collector and the emitter. We can now develop this model and use it to make some useful generalisations, particularly about the defect known as h_{FE}.

As the base becomes thicker, it becomes more and more probable that an electron passing from the emitter to the collector will be captured by the base and flow out of the base as base current; h_{FE} is, therefore, inversely proportional to base thickness.

When a transistor passes a high collector current, its base current must be proportionately high. In order for the base not to melt due to this current, the base must be thickened. The thicker base reduces h_{FE}, and required base current must rise yet further, requiring an even thicker base. h_{FE} is, therefore, inversely proportional to the square of maximum permissible collector current, and high-current transistors have low h_{FE}.

High-voltage transistors need a thick base to allow for the widened depletion region caused by a high collector to base voltage, so high-voltage transistors also have low h_{FE}.

High-current transistors must have a large silicon die area in order for the collector not to melt – this large area increases collector/base capacitance. The significance of capacitance in amplifying devices will be made clear when we consider the Miller effect in Chapter 2, but for the moment we can simply say that high-current transistors will be *slow*.

We have barely scratched the surface of semiconductor devices and circuits. Other semiconductor circuits will be presented as and when they are needed.

Feedback

Feedback is a process whereby we take a fraction of the output of an amplifier and sum it with the input. If, when we sum it with the input, it causes the gain of the amplifier to increase, then it is known as *positive* feedback, and this is the basis of oscillators. If it causes the gain to fall, then it is known as *negative* feedback, and this technique is widely used in audio amplifiers.

The Feedback Equation

The description of feedback was deliberately rather vague because there are many ways that feedback can be applied, and they each have differing effects. Before we can look at these effects, we need a few definitions and a simple equation:

$$A = \frac{A_0}{1 + \beta A_0}$$

where

A=amplification with feedback

A_0=amplification with zero (0) feedback

β=feedback fraction.

This is the general feedback equation that defines how the gain of an amplifier will be modified by the application of feedback. β is the *feedback fraction* and is the proportion of the output that is fed, or looped, back to the input. It is because β is so commonly used in the feedback equation that the apparently clumsy term h_{FE} is more popular for bipolar transistor current gain.

If βA_0 is very large *and positive* (causing negative feedback – a reduction of gain), then $\beta A_0 \approx \beta A_0+1$, and the gain of the amplifier becomes:

$$A = \frac{1}{\beta}$$

A_0 no longer affects A, and the *closed loop* gain of the amplifier is determined solely by the network that provides the feedback signal.

This result is very significant because it implies many things:

- Distortion is produced by variations in gain from one voltage level to another. If *open loop* gain is no longer part of the equation, then small variations in this gain are irrelevant, and the amplifier produces no distortion.
- If the feedback acts to maintain the correct gain *under all circumstances*, then it must change the apparent input and output resistances of the amplifier.
- If the feedback fraction β is set by pure resistors, then the equation for closed loop gain does not contain any term including frequency. Theoretically, the output amplitude is, therefore, independent of frequency.

In the late 1970s, when cheap gain became readily available, designers became very excited by the possibilities and implications of the feedback equation, and set out to exploit it by designing amplifiers that were thought to have very high levels of feedback. In practice, these amplifiers did *not* have high levels of feedback at all frequencies and power levels, and it was the *lack* of feedback to linearise these fundamentally flawed circuits that caused their poor sound quality.

Before we explore the expected benefits of feedback, we should, therefore, examine how the feedback equation could break down.

Practical Limitations of the Feedback Equation

The feedback equation implies improved performance provided that $\beta A_0 >> 1$. If, for any reason, the open loop gain of the amplifier is less than infinite, then βA_0 will *not* be much greater than 1, and the approximation will no longer be true.

Practical amplifiers always have finite gain; moreover, this gain falls with frequency. A practical amplifier will always distort the input signal, and because the distortion-reducing ability of negative feedback falls with frequency, the closed loop distortion *must* rise with frequency.

Crossover distortion in Class B amplifiers can be considered to be a severe reduction of gain as the amplifier traverses the switching point of the transistors or valves. Because of the drastically reduced open loop gain in this region, negative feedback is not very effective at reducing crossover distortion.

Although not explicitly stated by the feedback equation, the phase of the feedback signal is crucially important. If the phase should change by 180°, then the feedback will no longer be negative, but positive, and our amplifier may turn into an oscillator.

We will explore the practical limitations of feedback in Chapter 6, but we should realise that, as with any weapon wielded carelessly, it is possible to shoot oneself in the foot.

Feedback Terminology and Input and Output Impedances

We can make a number of general statements about feedback that enable us to quickly predict effects at either input or output:

• The way in which the feedback is *derived* affects the output impedance, whereas the way that it is *applied* affects input impedance.

• If we make a *parallel*, or *shunt*, connection, then we are dealing with a *voltage*, but if we make a *series* connection, we are dealing with a *current*.

• Feedback confined to one stage is known as *local* feedback, whereas a feedback loop enclosing a number of stages is known as *global*.

• We may have more than one global feedback loop, with one loop enclosed by another, in which case the loops are said to be *nested*.

• Once we have defined how the feedback is connected, we can state that for *negative* feedback, voltage feedback *reduces* impedances, but current feedback *increases* impedances.

• All the preceding statements apply to *negative* feedback and are reversed if the *feedback* becomes positive (thus an increase becomes a decrease, and vice versa).

As an example, we can now combine these statements to describe the global negative feedback loop of a typical power amplifier as being parallel-derived, series-applied, but we could equally well describe it as being voltage-derived, current-applied. Almost all power amplifiers use this particular feedback strategy because it ensures:

• Low output impedance (needed to drive electromagnetic loudspeakers as the loudspeaker designer intended)

• High input impedance (to avoid excessive loading of pre-amplifiers)

• Non-inverting gain (output is the same polarity as the input).

Noting that positive feedback reverses effects, positive voltage feedback would increase impedances, whilst positive current feedback would reduce impedances. Using a combination of positive and negative feedbacks, it is possible to make a power amplifier with zero, or even negative, output impedance, and this idea resurfaces at roughly 10-year intervals.

Unfortunately, amplifiers with negative output impedance are liable to oscillate because they reduce the total series damping resistance of the external (invariably resonant) load to zero. Nevertheless, some early valve power amplifiers had the facility to adjust output impedance through zero to a negative value, in an effort to improve the bass performance of the accompanying loudspeaker. As mentioned, this idea resurfaces periodically, but it is really only of use for the dedicated amplifiers in a loudspeaker system with an active crossover. (Active, or low-level, crossovers are used *before* the power amplifiers so that each drive unit has a dedicated power amplifier. Although this scheme may seem profligate with expensive power amplifiers, it has much to recommend the use of active, or low-level, crossovers.)

Although we have described the various modes of feedback and their trend on impedances, we now need to quantify the effect. Impedances are changed by the ratio of the *feedback factor*:

$$\text{feedback factor} = (1 + \beta A_0)$$

This is also the factor by which the gain of the amplifier has been reduced and is often expressed in dBs. An amplifier with 20 dB of global feedback has had its total gain reduced by a factor of 20 dB, and if this was a conventional power amplifier with shunt-derived feedback, its output impedance would therefore be reduced by a factor of 10.

The Operational Amplifier

Conventionally, when we think of computers, we think of *digital* computers, but once upon a time there were also *analogue* computers, which were hardwired to model complex differential equations, such as those required for calculating the ballistics of anti-aircraft shells. These analogue computers used a basic building block that became known as the *operational amplifier*. It was a small device for its time (smaller than a brick) and used valve circuits that operated from ±300 V supplies. With suitable external components, these operational amplifiers could be made to perform the mathematical operations of inversion, summation, multiplication, integration and differentiation.

Thankfully, the valve analogue computer is no longer with us, but the term ‘operational amplifier’, usually shortened to *op-amp*, is still with us, even if it now refers to eight-legged silicon beetles.

Op-amps attempt to define the performance of the final amplifier *purely* by feedback, and to do this successfully the op-amp must have enormous gain; 120 dB gain at DC is not uncommon, although AC gain invariably falls with frequency.

In the following discussions, we always make two fundamental assumptions about op-amps:

- Gain is infinite.
- Input resistance is infinite.

These assumptions specify ideal op-amps. Real-world op-amps have limitations and will *not* achieve this ideal at all frequencies, voltage levels, *etc.* Provided that we remember this important caveat *at all times*, we will not run into trouble.

It is habitual in op-amp circuit diagrams to omit the (typically ±15 V) power supply lines to the op-amp to aid clarity; nevertheless, the op-amp still needs power!

The Inverter and Virtual Earth Adder

The op-amp inverter has parallel-derived, parallel-applied, negative feedback, and since op-amp gain is infinite, the point of both derivation and application must have zero resistance; the amplifier has zero output resistance (see Figure 1.45).

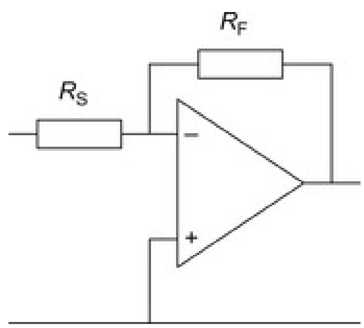


Figure 1.45 Inverting amplifier.

The inverting input of the op-amp also has zero resistance to earth (0 V) because of the feedback. Since the gain of the op-amp is infinite, if the non-inverting input is at earth potential, then the inverting input must also be at earth potential. In this configuration, the inverting input is, therefore, known as a *virtual earth*.

Although it is a virtual earth by virtue of feedback, the inverting input of the op-amp itself has infinite resistance, and *no signal current flows into the op-amp*. Input signal current from R_S can, therefore, only flow to ground via R_F and the zero output resistance of the op-amp. The signal currents in R_S and R_F are, therefore, equal, and using Ohm's law:

$$\frac{V_S}{R_S} = \frac{V_F}{R_F}$$

Since the inverting input is a virtual earth, $V_S = V_{in}$, $V_F = -V_{out}$ (the op-amp inverts) and the voltage gain of the amplifier is:

$$A_v = \frac{V_{out}}{V} = -\frac{R_F}{R_S}$$

Note that this amplifier can achieve $A_v<1$ and can *attenuate* the input signal. This *is* useful because it provides an attenuated output from an almost zero source resistance, whereas a potential divider would have significant output resistance. The minus sign reminds us that the amplifier inverts polarity.

Because the inverting node of the amplifier is a virtual earth, input resistance is equal to R_S.

When we analysed this amplifier, we considered the input signal current. There is no reason why this input current should come from only one source via one resistor, and we can sum currents at the inverting node in accordance with Kirchhoff's law (see Figure 1.46).

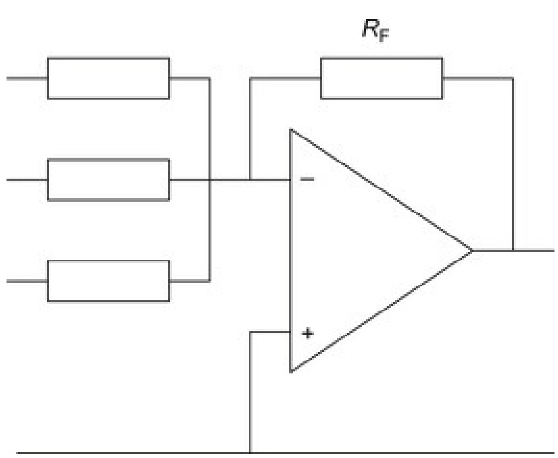


Figure 1.46 Virtual earth adder.

This circuit is known as the *virtual earth adders* and is most useful in bias servo circuits where we may need to add a number of correction signals. Voltage gain for *each* input may be determined using the inverter equation. When multiple inputs are driven, it is often best to determine the output voltage by summing the signal currents using Kirchhoff's current law, finding the resultant current in R_F and using Ohm's law to determine the output voltage. This is not as tedious as it sounds.

The Non-Inverting Amplifier and Voltage Follower

Frequently, we need a non-inverting amplifier (see Figure 1.47).

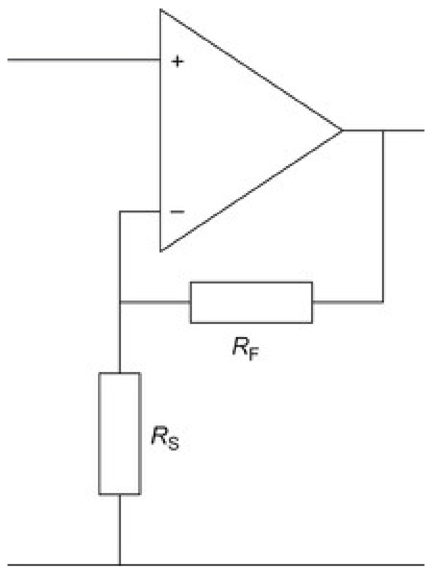


Figure 1.47 Non-inverting amplifier.

In this configuration, we still have R_F and R_S, but the amplifier has been turned upside down, and the far end of R_S is now connected directly to ground. Like the inverter we saw earlier, the amplifier has parallel-derived negative feedback, so its output resistance remains zero, but the feedback is now series-applied, making input resistance infinite. R_F and R_S now form a potential divider across the output of the amplifier, and the voltage at the inverting input is:

$$V_{(\text{inverting input})} = \frac{R_S}{R_S + R_F} \cdot V_{\text{out}}$$

Since the op-amp has infinite gain, the voltage at the inverting input is equal to that at the non-inverting input, which is V_{in}, so the gain of the amplifier is:

$$\frac{V_{\text{out}}}{V} = \frac{R_S + R_F}{R_S}$$

If we now reduce the value of R_F to zero and discard R_S, the gain reduces to 1, and the amplifier is known as the *voltage follower* (see Figure 1.48).

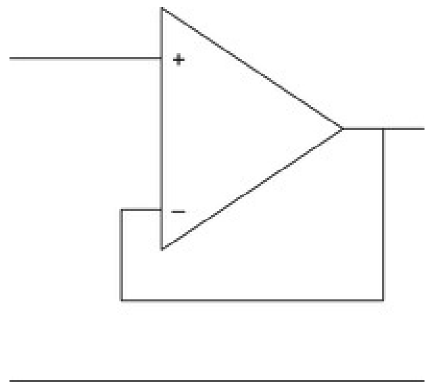


Figure 1.48 Voltage follower.

The voltage follower is an excellent *buffer* stage for isolating high-impedance circuits from low-impedance loads, and because it applies more negative feedback it is far superior to the single-transistor emitter follower that we saw earlier. (A very useful analytical technique is to treat an emitter follower as a common collector amplifier with 100% feedback and use the feedback equation to determine input and output resistances.) If we need more current than the op-amp can provide, we can add an emitter follower, or even a Darlington pair, to the output and enclose this within the feedback loop (see Figure 1.49).

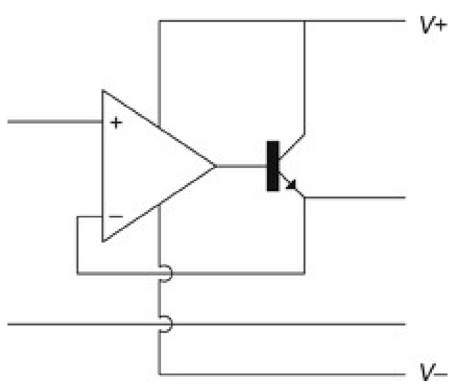


Figure 1.49 Addition of emitter follower to increase output current.

The Integrator

This circuit is essentially an inverter with R_F parallelled by C_F, but if R_F is omitted, it becomes an *integrator* and accumulates *charge*. The circuit may be considered to be a low-pass filter, in which case the cut-off frequency of the filter is:

$$f_{-3\ \mathrm{dB}} = \frac{1}{2\pi C_F R_F}$$

We can now see that if $R_F=\infty$, the cut-off frequency=0, and the amplifier has infinite gain at DC. Any DC offsets will gradually accumulate sufficient charge on the capacitor to cause the output of the op-amp to rise to maximum output voltage ($\pm V_{supply}$), and the op-amp is then said to be *saturated*. For this reason, practical integrators are usually enclosed by a servo control loop. The servo error signal charges C_F, but the loop aims to minimise the error, and so the voltage across C_F tends to zero.

Again, we could sum several input currents if we wished, and one possible use of this circuit would be for monitoring the voltages across the cathode resistors of several output valves because the integrating function would remove the audio signal and provide a DC output voltage proportional to the *total* output valve DC current.

The Charge Amplifier

Noting that the voltage gain of the inverting amplifier was the ratio of resistances, there is no reason why those resistances should not be replaced by reactances. Although an amplifier with gain defined by inductances is neither practical nor useful, replacing the resistors with capacitors *is* useful and is known as a *charge amplifier*. Charge amplifiers are commonly associated with transducers such as condenser microphones and radiation detectors that change or produce a charge between the plates of a capacitor. Such sources thus look like a Norton source (pure current source in parallel with capacitance) and can be converted into a Thévenin source having a pure voltage source in series with the source capacitance. Connecting such a source directly to the inverting input of the amplifier produces a circuit very similar to the previously analysed resistive inverting amplifier (see Figure 1.50).

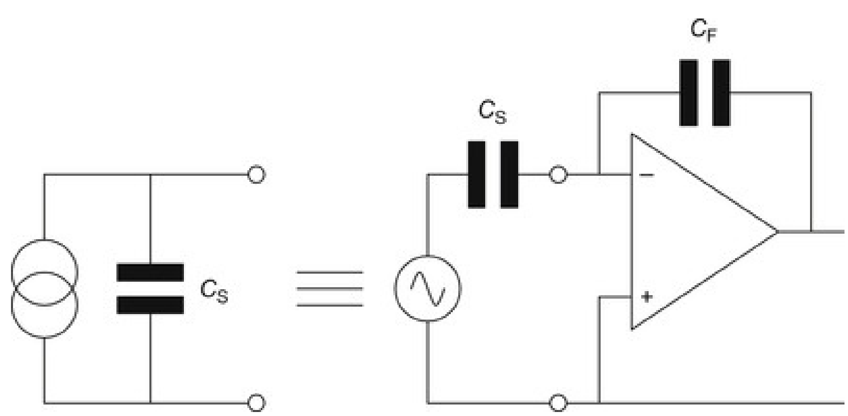


Figure 1.50 Charge amplifier.

The significance of this arrangement is that the voltage gain is the ratio of two reactances and is *independent* of frequency. It is initially surprising that an amplifier employing capacitors should respond to DC but is easily proved by the assumption that the op-amp draws zero input current. If the op-amp does not draw input current, then any DC input current to the amplifier must pass through the feedback capacitor, causing the output to respond. Charge amplifiers are typically associated with small source capacitances (30 pF for a large-capsule condenser microphone) and require small feedback capacitances (2–3 pF is typical) to provide useful gain.

Like the integrator, practical charge amplifiers require a means of discharging the feedback capacitor to prevent amplifier saturation, and this is commonly achieved by a parallel discharge resistor (see Figure 1.51).

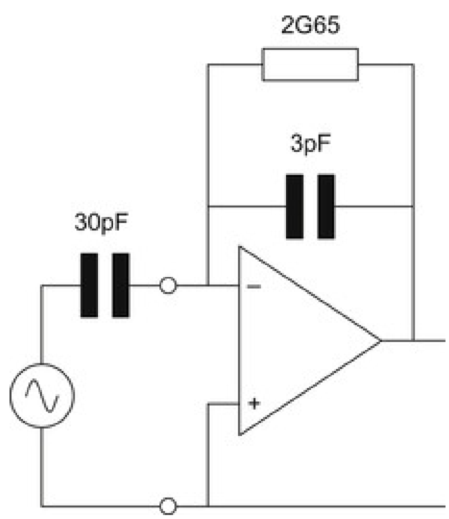


Figure 1.51 Practical charge amplifier.

Unfortunately, the effect of adding the discharge resistor is to turn the amplifier into a differentiator. In audio terms, we have added a high-pass filter and would like to place its $f_{-3\ \mathrm{dB}}$ frequency as low as possible, certainly 20 Hz.

$$R_{\mathrm{F}} = \frac{1}{2\pi f_{-3\ \mathrm{dB}} C_{\mathrm{F}}} = \frac{1}{2\pi \times 20 \times 3 \times 10^{-12}} = 2.65\ \mathrm{G}\Omega$$

As the equation shows, charge amplifiers invariably require very large discharge resistances. Nevertheless, such resistors are readily available, even if they do require the wearing of clean cotton gloves when handling to avoid adding surface leakage currents.

Interestingly, because the charge amplifier responds to charge (or over the short term, currents), it is the noise *current* produced by the discharge resistor that is important, not its voltage. Remembering that the thermal noise voltage of a resistor is:

$$v_{\text{noise}} = \sqrt{4kTBR}$$

substituting into Ohm's law gives:

$$i_{\text{noise}} = \sqrt{\frac{4kTB}{R}}$$

Thus, for lowest noise, the discharge resistor of a charge amplifier must be as large as possible.

DC Offsets

We briefly mentioned DC offsets when considering the integrator. Op-amps are not magical; they contain real transistors, and base, or leakage, current, known as *bias current*, will flow out of the inputs. The bias currents will not be perfectly matched, and the imbalance is known as *offset current*. The input transistors will not be perfectly matched for voltage either, so there will be an *offset voltage* between the inputs. These imperfections are detailed in the manufacturer's datasheets and should be investigated during circuit design.

References

[1] Kaye, GWC; Laby, TH, *Tables of physical and chemical constants*. 16 th ed. (1995)Longman.

[2] Parker, G, *Introductory semiconductor device physics*. (1994)Prentice Hall, UK, UK.

[3] Ebers, JJ; Moll, JL, Large-signal behavior of junction transistors, *Proc IRE* **December** (1954) 1761–1772.

[4] Early, JM, Effects of space-charge layer widening in junction transistors, *Proc IRE* **November** (1952) 1401–1406.

Recommended Further Reading

- Gray PR, Hurst PJ, SH, Meyer RG. Analysis and design of analog integrated circuits. 5 th ed. John Wiley. Do not be put off by the title – this is a superb book that goes into the detail of (semiconductor) circuit building blocks at individual transistor level.
- Jung WGG, editor. Op amp applications. Analog devices, US; 2002. As would be expected, this is circuit/system design from an IC manufacturer's perspective and includes plenty of hints and tips about how to best use real-world (as opposed to idealised) op-amps.

Chapter 2. Basic Building Blocks

In this chapter, we will look mainly at the triode valve, how to choose operating conditions and what effect these choices have on the AC performance of the stage. The analysis will use a combination of graphical and algebraic techniques, which has the advantage of being quick to use, and the results of the theory agree well with practice. This last point might seem to be an obvious requirement, but it is one that is sometimes overlooked.

The Common Cathode Triode Amplifier

The most common use of a valve is amplification. Therefore, we need to know how to configure and bias the valve so that it can amplify in a linear manner and minimise distortion. We will begin by investigating the *anode characteristics* of an ECC83/12AX7 (see Figure 2.1).

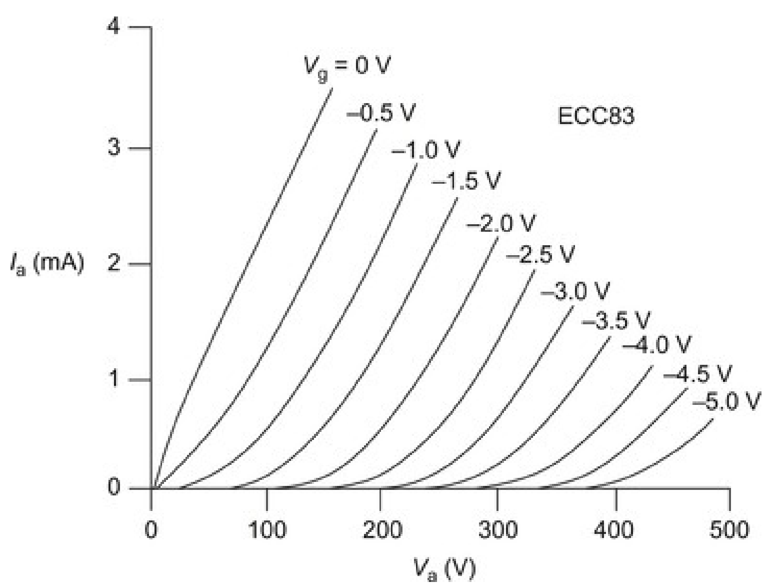


Figure 2.1 Triode anode characteristics.

The anode characteristics are the most useful set of curves for a valve, and the plot shows anode current I_a against anode voltage V_a, for differing values of grid to cathode voltage (V_{gk}). The first point to note is that valves operate at high voltages (typically a factor of 10 greater than transistor circuits) and quite low currents. The second point is that if there is no *bias* voltage (V_{gk}=0), then a large anode current flows. This is known as the *space-charge-limited* condition and means that the flow of current is limited only by the number of electrons that can

be released from the cathode. In contrast to the bipolar junction transistor, we have to turn the triode off, rather than on, in order to bias it correctly.
The basic amplifier stage has an *anode load* resistor R_L, connected between the anode and the *HT* supply (this is the historical phraseology, and stands for high tension) (see Figure 2.2).

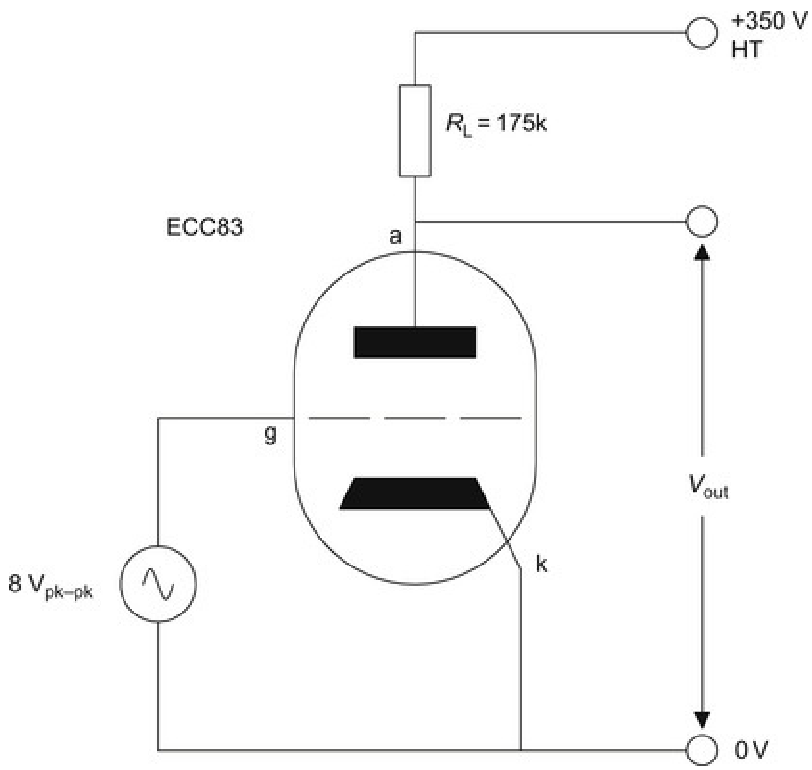


Figure 2.2 Common cathode amplifier.

The HT supply is assumed to have zero output resistance at all frequencies from DC to light (you may wish to consider whether this is in fact the case in a practical amplifier). By applying our input voltage between the grid and the cathode, we modulate V_{gk}, and thereby control anode conditions. This is why this grid is often known as the *control grid* in multi-grid valves such as tetrodes and pentodes.
We will now use the technique of *loadlines* to link the amplifier circuit to the anode characteristics, and to extract useful information from them.
Using Ohm's law, it is apparent that if there is no current flowing through the resistor (and therefore the valve), there must be no voltage across the resistor. If there is no voltage across the resistor, then all of the HT must be across the valve, so we could mark that as a point on the graph of the anode characteristics (V_a=HT=350 V; I_a= I_R=0). Similarly, we can argue that if there is no voltage

across the valve, then the HT must all be across the resistor; we can calculate the current through the resistor, and therefore the valve. In this case, R_L=175 kΩ and HT=350 V, so the anode current I_a=2 mA, and we can plot this point too.

Because Ohm's law is an equation that describes a straight line, if we know two points we have completely defined that straight line. This means that we can now draw a straight line between our two plotted points, as shown (see Figure 2.3).

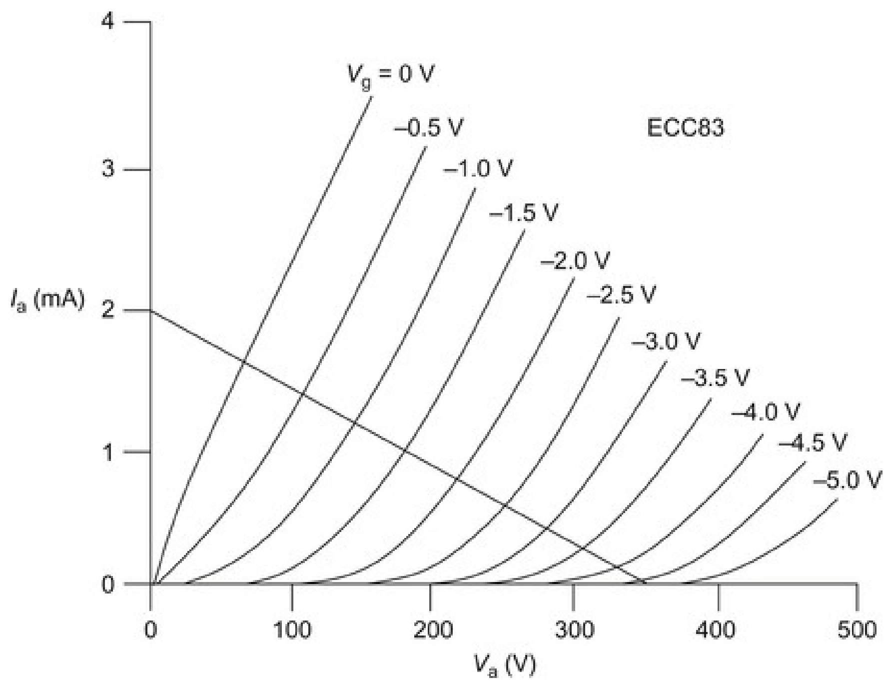


Figure 2.3 The loadline.

This is our *loadline*. This is perhaps the single most useful piece of analysis that can be performed on a valve stage. We have defined the anode current for any anode voltage, using an HT of 350 V and anode load of 175 kΩ. If we want to change our anode load, or HT, we must recalculate and redraw our loadline.

If we look along the loadline, we see that it is intersected at various points by the I_a/ V_a curves for differing values of V_{gk}. What this means is that those differing values of V_{gk} will cause predictable changes in anode voltage, and we could, therefore, calculate the *gain* of the stage.

Let us suppose that we apply an 8 V $_{pk–pk}$ sine wave to the input of the stage. If we start from 0 V, and look to find where the 0 V grid bias line intersects the loadline, this occurs at V_a=72 V. We then let the sine wave swing negative to −4 V, and see that it results in V_a=332 V. For an applied voltage of −4 V on the input, we produced a positive change of voltage at the anode of 260 V. The amplifier *inverts*. Since gain is defined as the ratio of output voltage to input voltage, we have just produced an amplifier with a gain of −65 (the minus sign merely reminds us that this is an inverting amplifier).

Unfortunately, it isn't very linear. If we now allow the input sine wave to continue rising past 0 V, we soon find that the anode voltage is unable to fall any further, and so the output signal no longer looks like the input signal.

We must choose a *bias* or *operating point* at which we will set the *quiescent* (no signal) conditions such that the stage can accommodate both negative and positive excursions of the signal without gross distortion.

Limitations on Choice of the Operating Point

Not only did the previous circuit distort, but the anode DC voltage was superimposed on the output signal, so we add a capacitor and a resistor at the output to block the DC (see Figure 2.4).

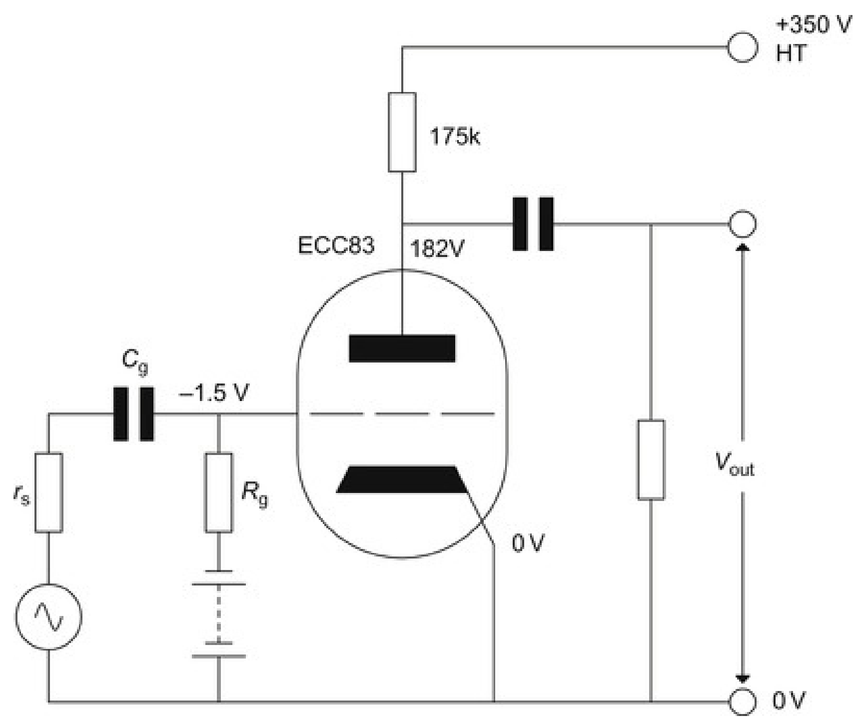


Figure 2.4 Grid bias using battery.

The valve is biassed by superimposing a bias voltage onto the grid via R_g, which prevents the battery from short circuiting the oscillator. C_g is the *coupling capacitor* that prevents the oscillator from shorting the battery, and r_s is the output resistance of the oscillator.

Returning to the loadline, we find that as V_a rises, the grid curves become progressively bunched together, which indicates non-linearity, and is particularly severe when V_a is close to the HT voltage. This region is known as *cut-off* (because the flow of current is being cut off). Operation near cut-off is not advisable if good linearity is required, although we will meet this mode of operation later when looking at some power stages.

Moving in the opposite direction along the loadline, we turn the valve on harder and harder, until finally there is no voltage across it. This is extreme, however, and we will encounter the problem of *positive grid current* long before that. As we make the grid less and less negative, there comes a point when the electrons leaving the cathode are no longer repelled and controlled by the grid, but are actually attracted to the grid and flow out through the grid to ground. This causes the input resistance of the valve, which could previously be regarded as infinite, to fall sufficiently low that it begins to load the oscillator's (non-zero) output resistance. Because this attenuation only happens on the positive *peaks* of the input waveform, this causes distortion of the *input* signal, even though the valve accurately amplifies the grid voltage. The onset of grid current varies with valve type (but is generally around −1 V) and is usually specified on the valve datasheet. For instance, Mullard specifies $V_{gk(max)}$ (I_g=+0.3 μA) as −0.9 V for the ECC83 used in our example.

If we have a voltage across the valve, and a current flowing through it, we must be dissipating power within that valve, and there will be a limit beyond which we are in danger of melting the internal structure of the valve. This is known as the *maximum anode dissipation* and is given on the datasheet as being 1 W for the ECC83. For power valves, the curve that corresponds to this is often drawn on the anode characteristic curves, but, if we wish, we can easily add it ourselves. All we need to do is to calculate the current drawn for 1 W at 0 V, 50 V, 100 V, 150 V and so on. We plot these results on the graph, and draw a curve through the points to form a hyperbola.

The valve datasheet also specifies two more, interlinked, restrictions on the choice of bias point: maximum V_a and maximum $V_{a(b)}$. Maximum V_a is the maximum DC voltage at which the anode may be continuously operated, whereas $V_{a(b)}$ is the maximum voltage to which the anode may be allowed to swing under signal or cold conditions, and is effectively the maximum allowable HT voltage for that valve. Ignoring these limits usually results in premature destruction of the valve to the accompaniment of blue flashes and bangs as residual gas in the valve is ionised and breaks down. This in itself may not cause irreversible damage, but if a path is formed between the anode and the control grid, then a large anode current will flow, and this may damage the valve permanently. You have been warned.

The final limitation is the maximum allowable cathode current $I_{k(max)}$. Usually, one of the other limitations comes into effect first, but input stages may minimise V_a and maximise I_a in order to maximise g_m, and minimise noise, so $I_{k(max)}$ should be checked if possible. (Neither Mullard nor Brimar specified $I_{k(max)}$ for the ECC83.)

We can now draw these limitations onto the anode characteristics, and choose our operating point from within the clear area (see Figure 2.5).

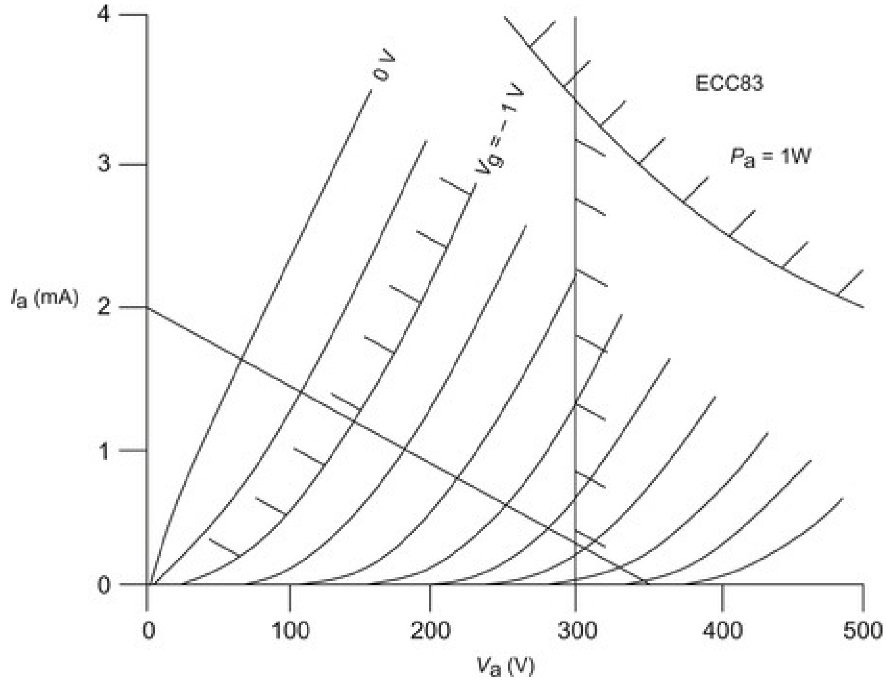


Figure 2.5 Determination of safe operating area.

Conditions at the Operating Point

Although the choice of operating point has now been considerably restricted, we can still optimise various aspects of performance.

In general, there are two main (and usually conflicting) factors: maximum voltage swing and linearity. If we want to bias for maximum voltage swing, then we would set the bias point at V_a=225 V, to allow the anode to swing up to 300 V and down to 150 V; this would be done by setting the grid bias to −2.1 V.

However, it might be that we are more interested in linearity than in maximum voltage swing.

Triodes produce mainly second harmonic distortion, which is generated by the amplifier having unequal gain on the positive half-cycle of the waveform compared to the negative half-cycle, and the distortion is directly proportional to amplitude. To maximise linearity, we should look for an operating point where the distance to the first grid line on either side of the operating point is, as nearly as possible, equal. In this case, we might bias the anode voltage to 182 V by applying −1.5 V to the grid.

Supposing we have chosen the linearity approach, we will now want to determine the *dynamic* or *AC* conditions of the stage to see if they satisfy our needs.

The first, and most obvious, parameter to determine is the voltage amplification (A_v), or gain, of the stage. We do this by looking an equal distance on either side of the operating point to the first intersection with a grid line, noting the anode voltage. Referring to Figure 2.5, if we move from the operating point to the right, we meet the 2 V grid line, which intersects at a voltage of 220 V; similarly, the 1 V grid line intersects at 148 V.

$$\text{Amplification } (A_v) = \frac{\text{change in anode voltage}}{\text{change in grid voltage}}$$

$$A_v = \frac{\Delta V_a}{\Delta V_g} = \frac{220 - 148}{1 - 2} = -72$$

The minus sign reminds us that the amplifier is inverting, but you will usually find this dropped, in the interests of clarity, since most stages invert, and the absolute polarity of any particular stage is often of little consequence.

The next important factor is the maximum undistorted voltage swing. Again, we look symmetrically on either side of the operating point, but this time we look for the first limiting value. In this instance, we look to the left and find that at 148 V we are approaching positive grid current. This would not matter if our source had zero output resistance, but this is unlikely to be the case, and so we must regard this as a limit. If we look to the right, we find that there is no practical limit until V_a=HT. Unfortunately, whilst this means that the valve can swing a large voltage positively, it cannot swing as far negatively. It is the *first* limit to be reached that is important. We can now see that the maximum undistorted peak-to-peak swing at the output is double that of the distance from the bias point to the first limit. In this example this corresponds to 72 $V_{pk\text{–}pk}$, but remembering that AC signals are specified as the RMS value of a sine wave, we should divide this figure by a factor of $2\sqrt{2}$, which results in a value of 25 V_{RMS} as the maximum undistorted sine wave output, which is perhaps not so impressive.

It may be that this value of maximum output swing is insufficient, so we would go back and reselect our operating point. If we are still unable to achieve a satisfactory value, then we may need to choose a different value of R_L, HT or both. It will now be apparent that designing valve stages requires a pencil, a clear ruler, an eraser and plenty of copies of anode characteristics curves.

Assuming that the stage looks promising so far, the next important parameter is the output resistance. A triode can be modelled as a voltage source coupled through a series resistance, known as the *anode resistance* r_a. Remember that because this is an AC, or dynamic, parameter, it is given a lower case letter, and is quite distinct from R_L, the anode load. This dynamic anode resistance is then in parallel with the anode load to form the output resistance r_{out} (see Figure 2.6).

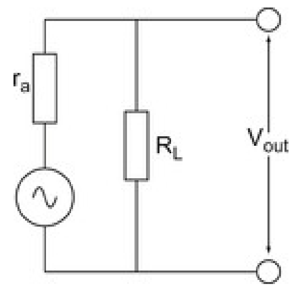


Figure 2.6 Thévenin equivalent of triode anode circuit.

It should be noted that the value of gain predicted for the stage by the loadline *already* includes the attenuation caused by the potential divider formed by r_a and R_L.

To find r_a, we return to the anode characteristics and draw a tangent to the curve where it touches the operating point. What we are aiming to do is to measure the gradient of the curve at that point. This is not as difficult as it sounds. A true tangent will touch, or intersect, the curve at only *one* point, and only at the correct point. A good quality transparent ruler is ideal for this purpose. Having positioned the ruler correctly, we draw a line that reaches the edges of the graduations on the datasheet, and read off the values at these points. The purpose of this is to make the resulting triangle, from which we take our figures, as large as possible in order to minimise errors (see Figure 2.7).

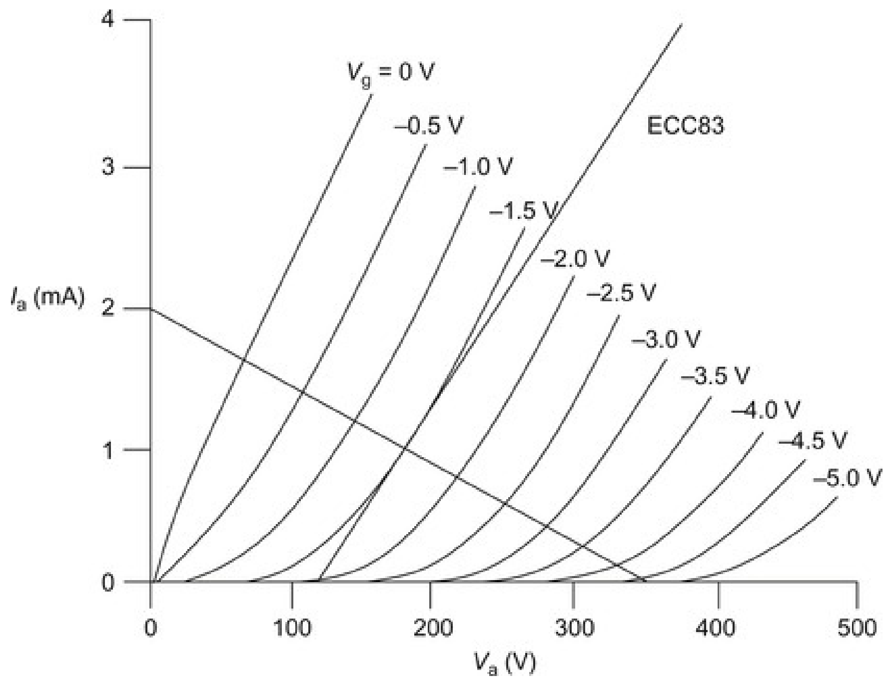


Figure 2.7 Determination of dynamic anode resistance r_a.

The anode resistance r_a can now be calculated from:

$$r_a = \frac{\Delta V_a}{\Delta I_a} = \frac{V_2 - V_1}{I_2 - I_1} = \frac{382 - 121}{4 - 0} = 65\ \mathrm{k}\Omega$$

You will note that the units of mA were used directly in the equation, resulting in an answer in kΩ; this is a conventional practice and saves time. The output resistance r_{out} is simply r_a in parallel with R_L, which results in a value of 47 kΩ. This is quite a high value of output resistance, and is a consequence of using a high μ (mu) valve, as high μ (mu) valves tend to also have a high value of r_a in operation.

Dynamic, or AC, Parameters

So far, we have analysed the behaviour of the valve graphically, but this is not the only method. There are three AC parameters that define completely the characteristics of a valve, *provided that they are evaluated at the operating point*. The importance of this last point is sometimes overlooked.

These parameters are:

μ (mu)=amplification factor (no units)

g_m=mutual conductance (usually mA/V)

r_a=anode resistance (kΩ, Ω).

The amplification factor is defined by:

The amplification factor (μ) of a valve is the ratio of the change in anode voltage Δ V_a to the change in grid voltage Δ V_g, with anode current held constant.

$$\mu = \frac{\Delta V_a}{\Delta V_g}$$

In a more digestible form, it is the maximum possible voltage amplification of the valve, and can only be achieved if $R_L = \infty$. In practice, we rarely achieve a gain as high as this.

Valves are frequently classified by μ as follows:

Low μ:	<8	(6080=2, 12B4A=6.5)
Medium μ:	8–30	(Type 76=13.8, ECC82=18, 6SN7=20)
High μ:	>30	(ECC81=65, 6SL7=70, ECC83=100, ECC807=150, WE416=250, PD500=1050).

We can measure μ at the operating point by drawing a horizontal line through the operating point, which is equivalent to $R_L = \infty$, and calculating the gain as before, by noting the intersections with the grid curves (see Figure 2.8).

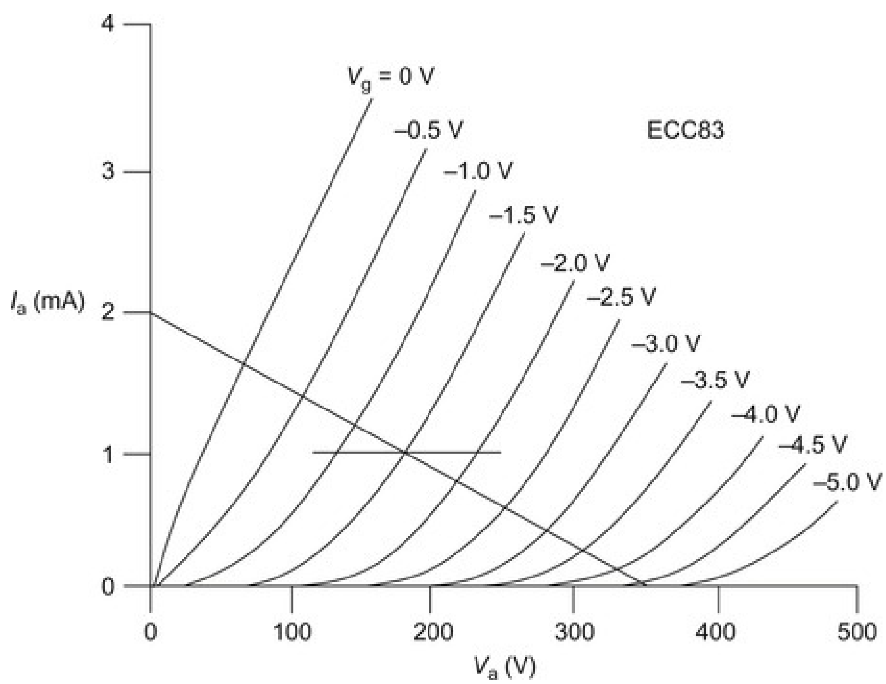


Figure 2.8 Determination of μ.

$$\mu = \frac{233 - 133}{2 - 1} = 100$$

Note that it is usual to ignore the signs of the individual voltages measured in equations like this.

Rather than using the loadline, we can use a formula to determine the voltage gain A_v of the amplifier stage:

$$A_v = \mu\left(\frac{R_L}{R_L + r_a}\right) = 100\left(\frac{175}{175 + 65}\right) = 73$$

which is in good agreement with the value predicted by the loadline (A_v=72). You will find that μ is one of the more stable valve parameters, and varies little with anode current (a fact that will be exploited later). However, this method is not ideal, since the accuracy of the final answer is dependent on how accurately you can draw tangents. It is, however, thoroughly recommended as a check on the general accuracy of your predictions.

The second valve parameter, mutual conductance, is defined by:

> *The mutual conductance g_m of a valve is the ratio of the change in anode current ΔI_a to the change in grid voltage ΔV_g, with anode voltage held constant.*
>
> $$g_m = \frac{\Delta I_a}{\Delta V_g}$$

Finally, we can define r_a:

> *The anode resistance r_a of a valve is the ratio of the change in anode voltage ΔV_a to the change in anode current ΔI_a, with grid voltage held constant.*
>
> $$r_a = \frac{\Delta V_a}{\Delta I_a}$$

There is a very useful equation which links these three parameters together:

$$g_m = \frac{\mu}{r_a}$$

Obviously, we can rearrange this equation as necessary to find the third parameter if we know the other two, but it is specified this way round, because although we can always predict μ and r_a reasonably accurately from the anode characteristics, we cannot directly predict an accurate value for g_m (yet traditional valve testers very rarely measure any parameter other than g_m) (see Figure 2.9).

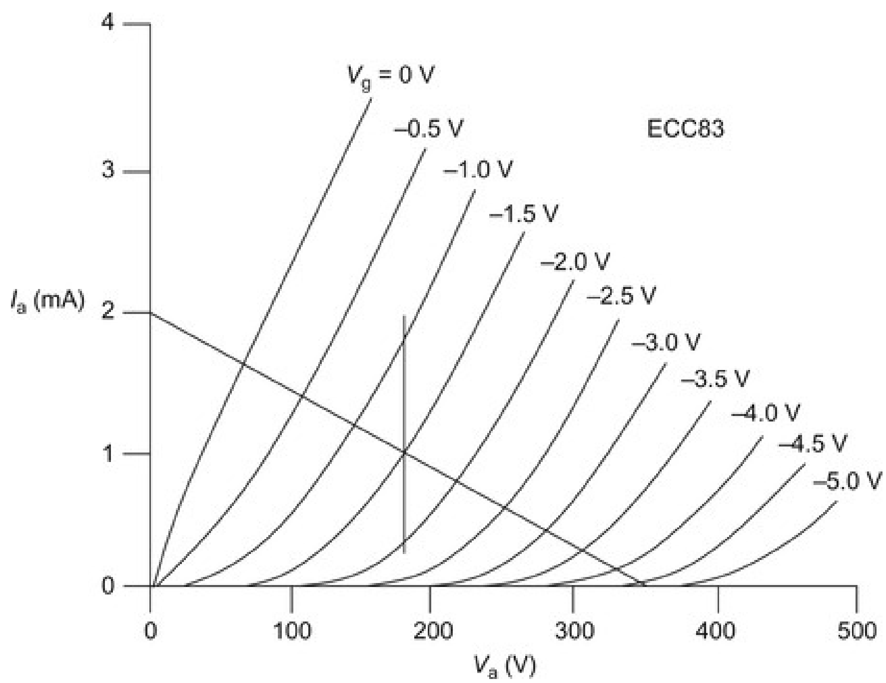


Figure 2.9 Determination of g_m.

In theory, to find g_m, we simply draw a vertical line through the operating point (hold V_a constant), and measure the change in anode current. However, it is immediately apparent that the change from 1.5 V to 1 V is considerably greater than the change from 1.5 V to 2 V, and taking the *average* value from 1 V to 2 V does not give an accurate figure of g_m at the operating point.

$$g_m = \frac{1.72 - 0.3}{1 - 2} = 1.42 \text{ mA/V}$$

Using the equation, however, with accurate values of μ and r_a (they must be because they agreed well with the loadline prediction of gain), we find:

$$g_m = \frac{\mu}{r_a} = \frac{100}{65} = 1.54 \text{ mA/V}$$

which means that the previous value was almost 8% low. We will return to use g_m later.

A parameter that is very occasionally mentioned is *perveance*, which is the ratio of the space-charge-limited anode current to the three-halves power of anode voltage (Child's law):

$$G = \frac{i_{a\,(\text{space-charge-limited})}}{V_a^{3/2}}$$

The practical significance of perveance is that a high-perveance valve requires less anode voltage for a given anode current. Additionally, high-perveance valves such as the 5687 can swing their anodes closer to 0 V, increasing voltage swing and efficiency.

Cathode Bias

Now that we have chosen our operating point and evaluated the dynamic characteristics of our amplifier stage, we need to look at practical ways of implementing the stage. Whilst we could bias the stage using a battery, it is inconvenient to disassemble the amplifier just to change a battery. However, lithium batteries having a shelf life of 10 years are now readily available, so battery replacement could perhaps be less frequent than valve replacement.

Another way of providing *grid bias* would be to have a subsidiary negative power supply and use potential dividers to determine the bias to individual valves. This is frequently done on power stages, but could cause noise and stability problems with small-signal stages.

An alternative method is to insert a *cathode bias* resistor between the cathode and ground, and connect the grid to ground via a *grid-leak* resistor. Conveniently, the grid is now at 0 V, so we no longer need an input coupling capacitor (see Figure 2.10).

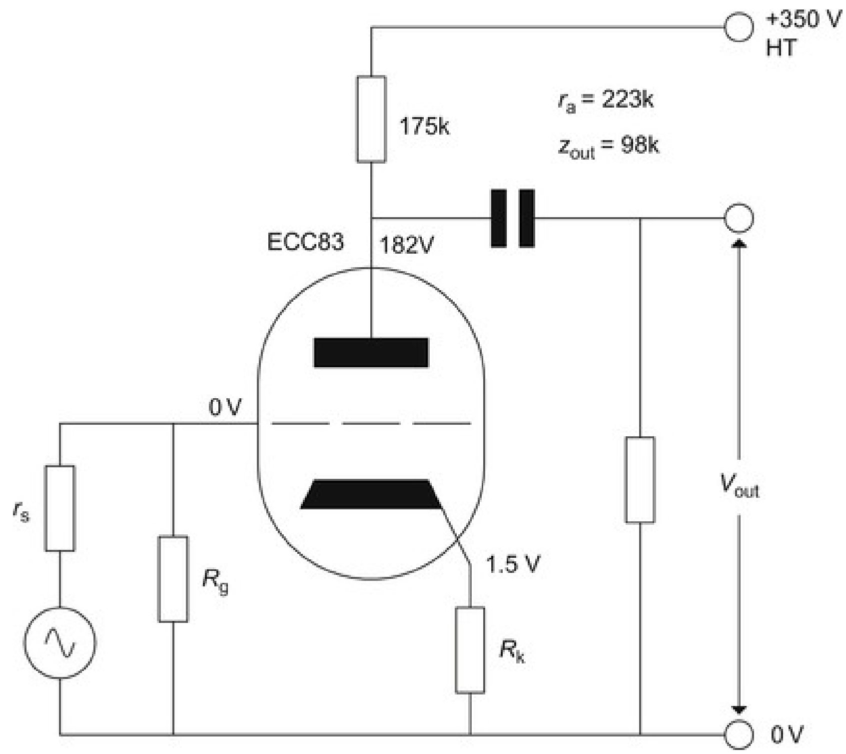


Figure 2.10 Cathode bias.

To understand the operation of this stage, we will assume a perfect valve that does not pass grid current even if $V_{gk}=0$.

Initially, there is no current flowing through the valve. If this is the case, there will be no voltage drop across the cathode bias resistor, and the cathode will be at 0 V. The grid is tied to 0 V, so V_{gk} must be 0 V. This will cause the valve to conduct heavily, but as it does so, the anode current (which in a triode is equal to the cathode current) flows through the cathode bias resistor, causing a voltage drop across it. This voltage drop causes the cathode voltage to rise, V_{gk} falls, and an equilibrium anode current is reached.

We know our operating point, therefore we know anode and, hence, cathode current. We know what value of V_{gk} we need. If the grid is at 0 V, then the cathode must be at + V_{gk}. If we know the voltage across, and the current through, an unknown resistor, then it is a simple matter to apply Ohm's law and find the value of that resistor. In our example we chose to place our operating point at 182 V. We could read off the anode current directly, but it is more accurate in this instance to calculate the current using Ohm's law. (This is because we can read off the value of V_a with greater accuracy.)

$$I_a = \frac{\mathrm{HT} - V_a}{R_L} = \frac{350 - 182}{175} = 0.96\ \mathrm{mA}$$

We know that the cathode voltage is 1.5 V, so the cathode bias resistor will be:

$$R_k = \frac{1.5}{0.96} = 1.56\ \text{k}\Omega$$

Again, note that the equation directly used mA, resulting in a resistance in kΩ.

The Effect on AC Conditions of an Unbypassed Cathode Bias Resistor

---

Although the cathode bias resistor *stabilised* and set the DC conditions of the stage, it did so by means of negative feedback, so we should expect it to affect the AC conditions such as gain and output resistance. We can use the universal feedback equation to determine the effect it will have.

$$A_{fbk} = \frac{A_0}{1 + \beta \cdot A_0}$$

The feedback fraction β in this case is the ratio R_k / R_L, so:

$$A_{fbk} = \frac{72}{1 + (1.56/175)72} = 44$$

The gain has been considerably reduced. The feedback is series-derived, series-applied, so it raises the input and output resistances. Since the input resistance of a valve is virtually infinite anyway, this won't be affected, but the anode resistance r_a will be raised.

Although the feedback equation is very handy for quickly determining the new gain, it is not quite so easily used for finding the new r_a.

Looking down through the anode, the only path to ground is the cathode, via the anode resistance r_a. Since, in this direction, resistances are *multiplied* by (μ+1), we see an effective anode resistance of:

$$r'_a = r_a + (\mu + 1) \cdot R_k = 65 + (100 + 1)1.56 = 223\ \text{k}\Omega$$

The value of r_a rises from 65 k to 223 k. In parallel with R_L, this gives a new output resistance of 98 k, as opposed to 47 k. Incidentally, there is no reason why we should not calculate the new r_a*first* and use that new value in the standard gain formula to determine the new gain:

$$A_v = m\left(\frac{r'_a}{R_L + r'_a}\right) = 100\left(\frac{175}{175 + 223}\right) = 44$$

It is most important to appreciate that the feedback affected only the valve's internal r_a. The anode load resistor R_L was external to the feedback, and therefore not affected.

Having evaluated the new values of gain and output resistance, we may decide that they are no longer satisfactory. We could either choose a new value of R_L, and try a new operating point, or we might even choose a new valve. However, there is another avenue open to us.

The Cathode Decoupling Capacitor

The addition of the cathode bias resistor caused negative feedback, and reduced gain. This may not always be desirable, so we will now consider how to prevent this feedback.

Because the output signal is derived from changing I_a through R_L, and I_a also flows through R_k, we must also develop a signal voltage across R_k. The signal voltage across R_k is in phase with the input signal, but because the valve responds to changes in V_{gk}, which is the *difference* between V_g and V_k, we have effectively reduced the available driving voltage to the valve.

To restore full gain, we must suppress the feedback voltage produced at the cathode with a *decoupling* or *bypass* capacitor. The capacitor should be of sufficiently low reactance that it is a short circuit at all AC frequencies of interest. In conjunction with the output resistance at the cathode, this forms a local low-pass filter (see Figure 2.11).

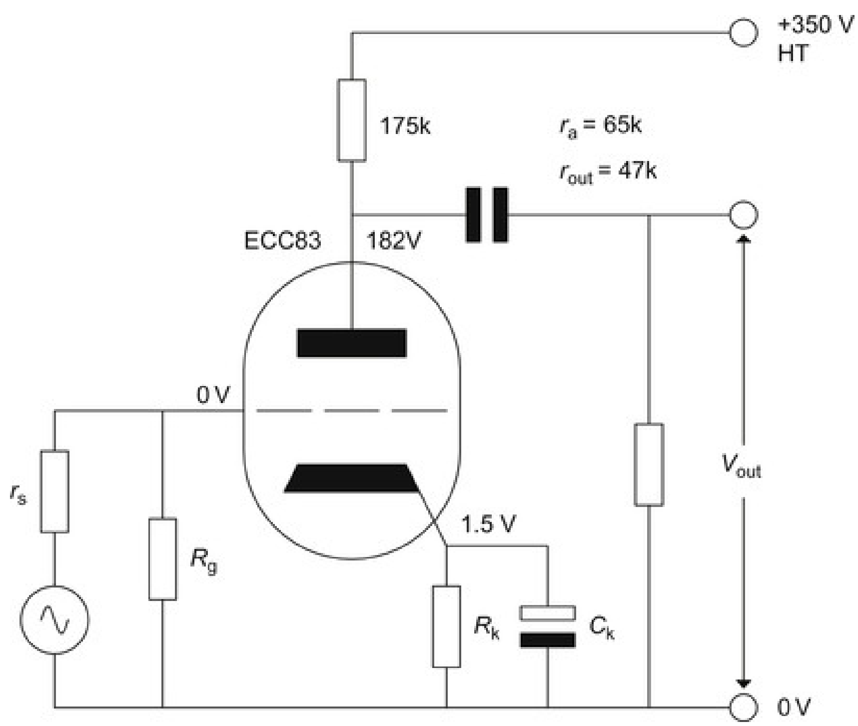


Figure 2.11 Cathode decoupling.

We now need to know what resistance the capacitor 'sees' from its positive terminal to ground. Clearly, it sees the resistor R_k, but it also sees the cathode resistance of the valve. The resistance looking into the cathode is:

$$r_k = \frac{R_L + r_a}{\mu + 1}$$

We can see the HT supply (AC ground) through the series resistance of r_a and R_L, but this is divided by the factor (μ+1) because of the internal voltage gain of the stage. If we now put some numbers into the equation, we have:

$$r_k = \frac{175 + 65}{100 + 1} = 2.38\ \text{k}\Omega$$

In parallel with the 1.56 kΩ cathode bias resistor this gives a total resistance r'_k at the cathode of 946 Ω.

In audio, we usually consider frequencies down to 20 Hz (although a 32 ft organ stop will produce 16 Hz), and large loudspeakers can reproduce them. There will be a number of stages to the amplifier, each with filters, so the effect is cumulative. The filter will be made with electrolytic capacitors, which are not known for their initial tolerance or stability of value, so the filter frequency should be much lower than any other filter frequency in the amplifier in case it changes. It has also been argued that a good low frequency response is required not merely to maintain correct amplitude response, but to ensure that the effects on phase and transient response (which extend in-band to a factor of 10 times the filter cut-off frequency) are kept to a minimum. Bearing all these factors in mind, it is usual to design for a cut-off frequency of 1 Hz, so:

$$C_k = \frac{1}{2\pi f r'_k} = \frac{1}{2 \times 3.14 \times 1 \times 946} = 1.7 \times 10^{-4}\ \text{F} = 170\ \mu\text{F}$$

Note that r'_k had to be entered using the base unit of ohms and that the equation produces a value for C in farads. The nearest standard value to 170 μF is 220 μF, and this is what we would use. You will note that quite a large value of capacitance is needed; valves with a lower r_k are not uncommon, and require a correspondingly larger value of capacitor.

Choice of Value of Grid-Leak Resistor

Although we have shown the grid-leak resistor in place previously, we have not assigned it a value. Historically, it has generally been 1 MΩ for small-signal stages, but somewhat lower for power stages. It is in our interests to make the grid-leak resistor as large as possible, for two reasons:

- The grid-leak resistor forms a potential divider in conjunction with the output resistance of the preceding stage and therefore causes a loss of gain. This loss is generally quite small, but it accumulates, so that at the output of a four-stage amplifier, the gain could be significantly less than predicted if this loss is not taken into account.
- A large value of grid-leak resistor allows the inter-stage coupling capacitor to be as small as possible for a given low frequency cut-off.

Once again, if we consult the valve datasheet, we find that there is a limit on the maximum value of grid-leak resistor. Usually, two values are given, one for cathode bias and one for grid bias; the value for grid bias is invariably lower (2.2 MΩ versus 22 MΩ for the ECC83). The reason that the grid bias value is so much lower is that there is *no* stabilisation of operating conditions in this mode. We set grid voltage, and the anode current is solely determined by the characteristics of that particular valve.

The clue to the interaction between the valve and the resistor is in the name ‘grid-leak’. In practice, there is always a very small leakage current flowing from the grid to ground, partly because there will always be some contamination of the grid with the oxide coating used to form the nearby cathode emissive surface, but also because of gas current.

Gas current occurs because there is always residual gas in the valve. Brownian motion ensures that individual gas molecules are distributed evenly within the valve, so some must be in the electron path. When a high velocity electron strikes a gas molecule, it may have sufficient energy to displace an electron from the molecule’s outer shell. The resulting two electrons then continue their path to the anode, but the gas molecule is now a positively charged ion (because it has lost an electron) and is, therefore, repelled by the anode, so it travels towards the grid/cathode. Statistically, as more electrons flow from cathode to anode, random collisions between electrons and gas molecules become more likely, so positive ion current towards the grid increases with anode current. As each ion strikes the grid, an electron immediately flows up through the grid-leak resistor to discharge it.

The flow of ion-discharging electrons is a significant part of grid leakage current and produces a potential across the grid-leak resistor, making the grid slightly positive. V_{gk} is, therefore, reduced, and if the value of grid-leak resistor is sufficiently high, this fall in V_{gk} becomes significant, and anode current rises. The increase in anode current increases anode dissipation, releasing yet more gas from the hot structures, further increasing ion current, so V_{gk} falls further, the cathode emits more electrons, and the process becomes self-sustaining until the valve is destroyed.

However, although grid ion current reduces V_{gk}, which increases anode current, the increased anode current has a counterbalancing effect because it *increases* V_{gk} because of the voltage drop across the cathode bias resistor R_k.

Mullard [1] published a method for determining the maximum permissible value for the grid-leak resistor under actual operating conditions. To determine the maximum permissible grid-leak resistance of our ECC83 amplifier, we need to know R_k and g_m (R_k=1.56 kΩ; g_m=1.54 mA/V). First, the effective DC cathode resistance of the circuit is found using:

$$R_{k(\text{effective})} \geq R_k + \frac{1}{g_m} = 1.56 + \frac{1}{1.54} = 2.21\ \text{k}\Omega$$

Knowing $R_{k(\text{effective})}$ and g_m, we refer to the graph to find the ratio by which the maximum fixed-bias grid-leak resistance may be multiplied (see Figure 2.12).

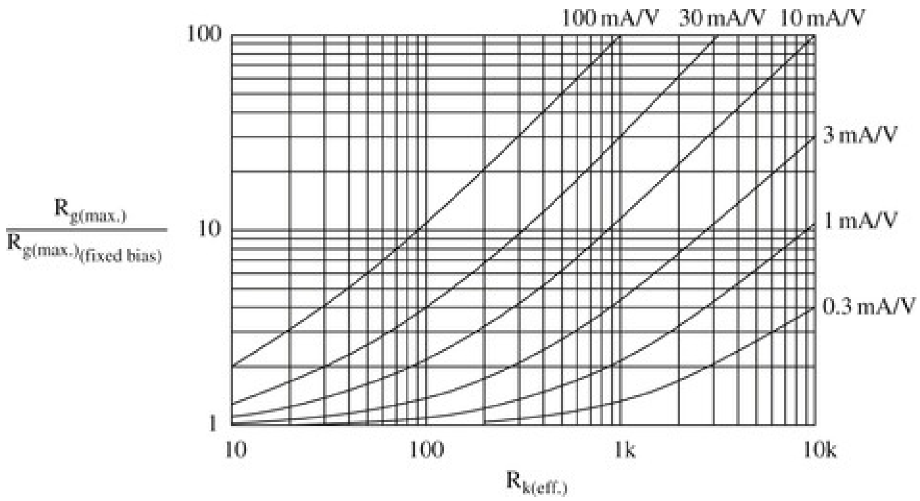


Figure 2.12 Maximum value of grid-leak resistor.
(After Mullard [1]).

Interpolation of the curves suggests that this particular circuit may use a grid-leak resistance, a factor of four times greater than the maximum fixed-bias grid-leak resistance (2.2 MΩ)=8.8 MΩ.

Nevertheless, the author has seen even larger grid-leak resistances in some designs.

Choice of Value of Output Coupling Capacitor

This is actually something of a misnomer, since it actually protects the input of the next stage from the anode voltage of the first stage, but because the input of a valve stage is usually DC coupled, the coupling capacitor becomes associated with the preceding stage.

The first, and most obvious, point to observe is that the capacitor should be able to withstand the anode voltage applied to it. What is not so obvious is that it should also be able to withstand the maximum *likely* HT voltage. Modern amplifiers are frequently built using silicon rectifiers for the HT. This means that at the instant of switch-on, the cathodes of the valves may be cold, causing zero anode current. Because the HT is unloaded, it rises to its maximum possible value, and this voltage appears directly across the coupling capacitors. If they fail catastrophically, then as the valves begin to warm up, the large positive bias on their grids causes them to conduct heavily. The valves may then be destroyed. Using higher voltage capacitors may be slightly more expensive, but it is much cheaper than having to replace an expensive valve (or loudspeakers).

The only other way around this is to ensure that the HT is *never* present before the heaters are warm. Usually this means leaving the heaters on permanently, which may be practical, and beneficial, for pre-amplifiers, but we would not wish to leave power amplifier heaters on permanently. A delay is needed, and valve rectifiers are the traditional solution (see Chapter 5).

The other choice is the value of capacitance of the capacitor. Since we will use either plastic or paper capacitors, which are stable in value, we do not mind if they define the low frequency cut-off of the amplifier. However, all the other arguments used for the cathode decoupling capacitor still apply, and so a choice of 2 Hz for cut-off frequency is not unreasonable. Incidentally, the traditional values of 1 MΩ grid-leak and 0.1 μF form a filter whose −3 dB point is 1.6 Hz. Some modern designs use much larger values, and we will consider the rationale for this later.

Miller Capacitance

So far, we have looked at the external, wanted, components of our amplifier stage. We will now turn to an unwanted component: *Miller capacitance*.

There will always be some capacitance between the anode and the control grid. In a tetrode or pentode it is still there, but greatly reduced. This capacitance is reflected into the grid circuit and forms a low-pass filter in conjunction with the output resistance of the preceding stage (see Figure 2.13).

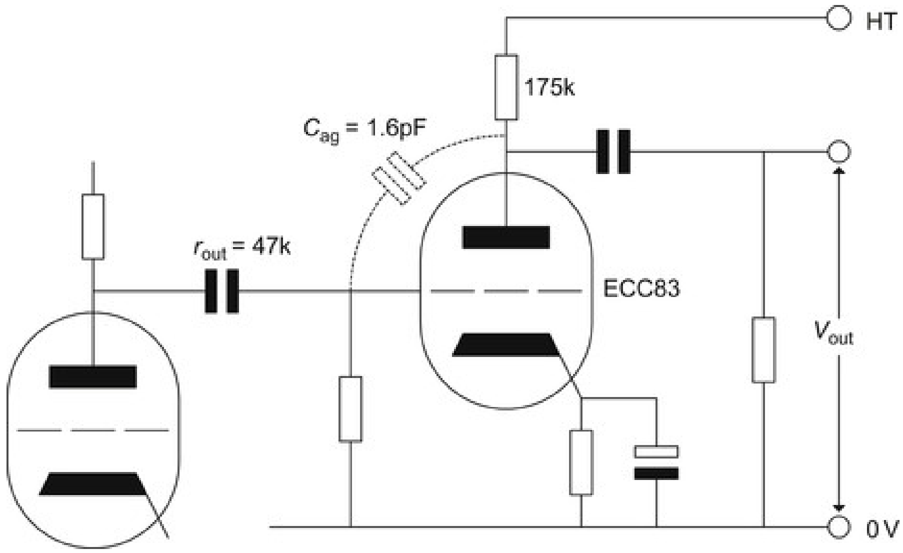


Figure 2.13 Miller capacitance.

We now have two identical stages of the type that we have just designed, connected in *cascade* to form a two-stage amplifier.

Miller capacitance acts like this: When the second valve amplifies the signal, its changing anode voltage is forced to charge and discharge the anode to grid capacitance C_{ag}. That charging current cannot flow into the grid (because the grid is high resistance), so it must be sunk or sourced by the preceding stage. Now, suppose that this capacitance requires a current i to charge it to 1 V. We apply a 1 V step to the input of the amplifier, and the anode moves negatively by 1 V×the gain of the amplifier, in this case by −72 V. The *total* voltage change across the capacitor is therefore $(A+1)V=73$ V.

The total current required to be sunk or sourced by the preceding stage is now $(A+1)\times i$ or 73 i. We could now reflect this capacitance into the grid by saying that exactly the same current would flow from the source if there was a capacitance between the grid and ground that was $(A+1)C_{ag}$; hence, the Miller equation:

$$C_{\text{Miller}} = (A+1)\cdot C_{ag}$$

It is clear that a quite small value of anode to grid capacitance can have an alarming effect on the high frequency response of an amplifier combination. In our particular case, we find that we have a Miller capacitance of 115 pF ($C_{ag}=1.6$ pF for ECC83/12AX7). In combination with the output resistance of the previous stage, this gives a high frequency $f_{-3\ \text{dB}}$ point of 29 kHz. Stray capacitance will reduce this frequency even further.

There are various ways to combat this problem.

- Reduce the output resistance of the preceding stage.
- Reduce significantly C_{ag} by screening the grid support rods from the anode.
- Dramatically reduce C_{ag} by screening the entire grid from the anode (tetrode or pentode).
- Reduce the gain to the offending anode (cascode or cathode follower).

This problem of High Frequency response is so important that we will investigate all four of these methods of improving the performance of the common cathode triode stage.

Reducing Output Resistance of the Previous Stage

Choosing an E88CC/6922 and operating it correctly reduces the output resistance to a typical value of 10 kΩ. If we also change the second stage to E88CC/6922, the Miller capacitance is then lower, typically 50 pF (due to the gain falling to 30), giving an $f_{-3\ \mathrm{dB}}$ point of approximately 300 kHz. However, we have reduced the combined gain of the two-stage amplifier from 5,184 (72^2) to 900 (30^2).
As an alternative, we could place a cathode follower (which we will investigate later in this chapter) between the two stages. A cathode follower easily achieves r_{out}≈1 kΩ, so even driving 115 pF of Miller capacitance gives an $f_{-3\ \mathrm{dB}}$ point of 1.4 MHz.

Guided-Grid, or Beam, Triodes

In the effort to obtain high μ and g_m simultaneously, the spacing between anode and grid must be reduced, forcing C_{ag} to rise, and the Miller effect causes the stage to have high input capacitance.
Although we might think of the grid as being a mesh of fine wires between the cathode and the anode, it has to be supported by rigid vertical metal rods which must be of a much greater diameter than the grid wires in order to determine the cathode/grid spacing precisely. As an example, the grid wire of the 417 A triode is specified [2] to be 7.4 μm in diameter and wound with a pitch of 0.065 mm per turn. On dissection, 80 grid wires were counted under a travelling microscope. The width of the anode near to the grid is ≈3 mm, so the total grid wire area is 80×3×0.0074=1.78 mm ^{2}. The support rods were measured to be 0.875 mm diameter, the length of the anode near to the rods is 5 mm, and there are two rods, so the total rod area is 2×0.875×5=8.75 mm ^{2}. The grid support

rods have a surface area five times that of the grid wires, and consequently five times the capacitance to a flat anode.

Valves for use at high frequencies seek to minimise capacitance between the anode and the support rods, hence the bath-tub anode on the 417A, which brings the anode close to the grid wires and avoids the rods, but this still means that a substantial proportion of C_{ag} is due to a structure that has no effect on the passage of electrons, and could, therefore, be screened from the anode with impunity. Logically, these should be called screened-grid valves, but the term had already been used for tetrodes, so the rather less satisfactory terms *guided-grid* triode or *beam* triode were used (see Figure 2.14).

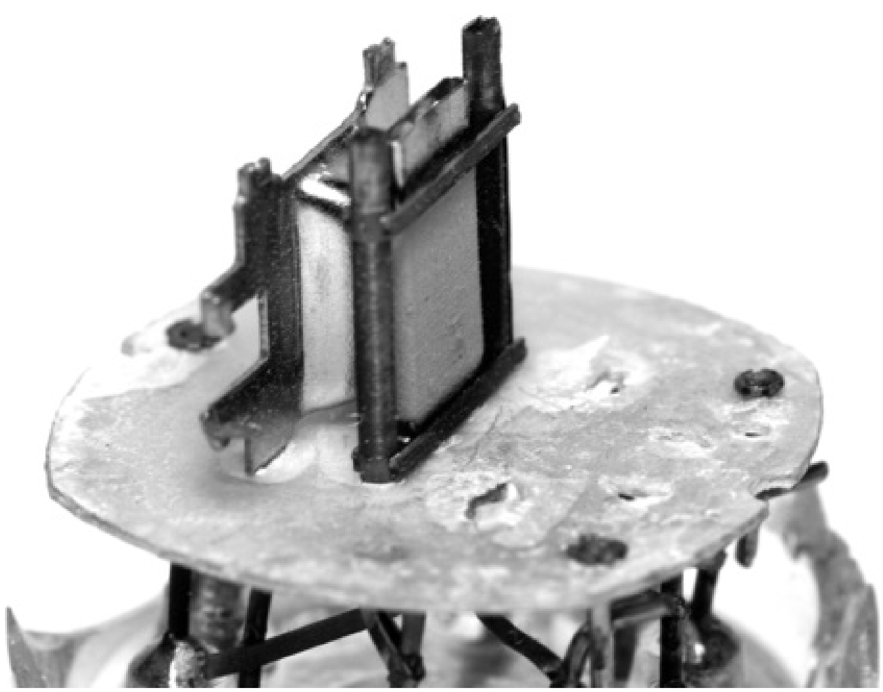


Figure 2.14 Dissected 417A: Note the size and shape of the remaining anode section relative to the active area of the grid.

Valves such as the PC97, PC900 and 6GK5 have internal support rod screens and bath-tub anodes, causing C_{ag} to fall to <0.5 pF – a very worthwhile improvement. Sadly, most of these Ultra High Frequency (UHF) valves were designed to be variable- μ valves, to allow Automatic Gain Control (AGC), and we will see later that this causes distortion (see Figure 2.15).

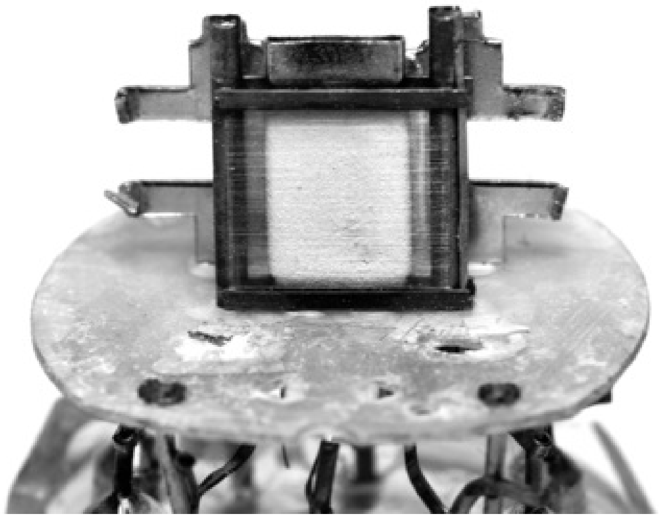


Figure 2.15 Dissected 417A: Note the relative size of the control grid support structure compared to the grid wires.

The Tetrode

As described by the patent, the tetrode [3] was invented to overcome the reduction in gain caused by the electric field of the anode interacting with the electric field of the grid. An auxiliary, or *screen grid*, g_2, is placed between the anode and the grid to screen the changing anode potential from the grid. To maintain electron flow to the anode, g_2 is connected to a positive potential slightly lower than that of the anode so that electrons are attracted to g_2, but most pass through the (coarse) mesh to be captured by the anode as anode current (see Figure 2.16).

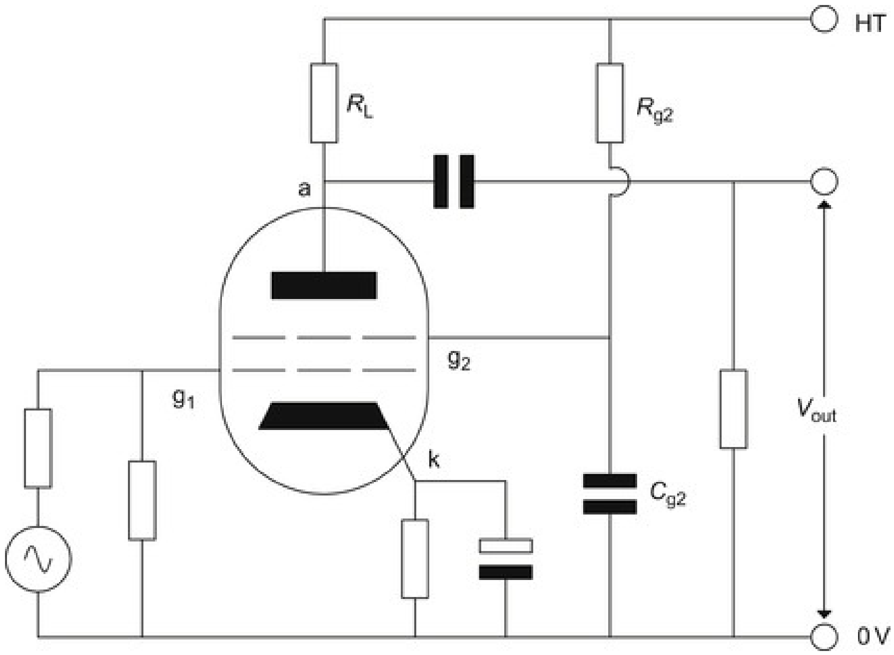


Figure 2.16 The tetrode.

Although originally devised to increase voltage gain, the far more important effect of adding g_2 is that it screens the anode from the grid at AC and greatly reduces Miller capacitance, allowing useful amplification at much higher frequencies. It will come as no surprise to learn that this tinkering with the internal structure to increase gain changes the anode characteristics of the valve (see Figure 2.17).

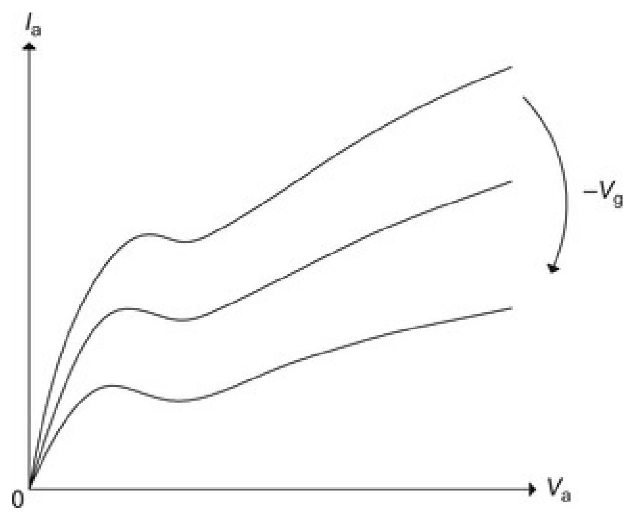


Figure 2.17 Anode characteristics of the tetrode.

The kink in the curves is caused by *secondary emission*. At very low anode voltages, electrons are emitted by the cathode in the normal way and collected by the anode. At slightly higher anode voltages, an electron may hit the anode with such velocity that instead of merely being absorbed by the anode, it dislodges *two* low velocity electrons, which are easily attracted to the higher potential of the screen grid. The anode has effectively emitted one electron, and anode current has fallen. As anode voltage rises still further, although electrons are dislodged from the anode, they swiftly return to the anode because the screen grid is at a low potential relative to the anode, and not so attractive, so anode current rises once more.

Not only does the kink in the anode curves cause distortion of the signal, but it also implies negative anode resistance, which can cause stability problems, so the pure tetrode was soon superseded.

The Beam Tetrode and the Pentode

These two valves sought to keep the advantages of the tetrode (high gain and low C_{ag}) without the disadvantage of the kinked anode characteristic. The pentode works by placing a very coarse grid, the *suppressor grid* g_3, connected to the cathode, between g_2 and the anode, in order to screen g_2 from the anode. The result of this is that the high velocity electrons emitted from the cathode pass straight through the suppressor grid, but the low velocity secondary electrons emitted from the anode are screened from g_2, and return to the anode. Because secondary emission electrons from the anode are no longer attracted to g_2, the kink in the anode characteristics of the tetrode is avoided.

Operation of the beam tetrode is different from that of the pentode in order to avoid infringement of the 1928 Philips pentode patent [4]. Instead of the electrons leaving the cathode from all points of the compass and flowing to the anode, the electrons are directed into two narrow beams of high electron density by the *beam forming plates*, which are connected to the cathode. Each beam is further focussed and divided into thin horizontal sheets because the g_1 and g_2 windings are vertically aligned, which increases electron density still further. Electrons attempting to leave the anode by secondary emission are now repelled by the incoming flood of electrons and are quickly returned to the anode. Because the dynamics of this space are very similar to the space-charge-stabilised emission from the cathode, it is known as a *virtual cathode*. Interestingly, the *pentode* patent hints at a virtual cathode as a means of suppression of secondary emission.

Some electrons may succeed in leaving the anode and travelling a limited distance, and to avoid them reaching g_2, the anode-to-g_2 distance is rather greater than in the pentode, which is why the anode of the beam tetrode (Kinkless Tetrode) KT66 is larger than that of the pentode EL34, despite their very similar ratings.

The necessary alignment of the g_1 and g_2 windings in a beam tetrode focusses the streams of electrons such that they mostly pass between the wires of g_2, thus reducing g_2 current compared to the pentode, which improves efficiency in a power valve, although there is no reason why a pentode should not adopt the same strategy. In practice, when we use a beam tetrode or a pentode, we see so little difference in their electrical characteristics that we can treat them both as pentodes. (Thorn-AEI classified the PCL82 as a triode/beam tetrode, yet Mullard classified it as a triode/pentode.)

The beam tetrode offers some interesting possibilities. For instance, if we had *two* anodes and individually connected beam anodes, we could modulate the voltage between the beam anodes to control the ratio of current split between the two anodes. The 6AR8 is such a valve, and it was designed mainly for colour decoding of video in televisions, but its characteristics were also exploited by audio tuners in the RF mixer stage and in the stereo decoder.

The Significance of the Pentode Curves

If we investigate the anode characteristics of the EF86 small-signal pentode for $V_{g_2} = 100\ \mathrm{V}$, we see that the anode curves are nearly horizontal (see Figure 2.15). We can make some useful observations from these curves.

Firstly, pentode characteristics are very similar to transistor characteristics and indicate an anode resistance that is sufficiently high that for most practical purposes, it may be taken to be infinite. The output resistance of a pentode stage is therefore $\approx R_L$.

Secondly, the anode is able to swing much closer to 0 V than the triode, and so we can obtain a greater peak-to-peak output voltage. This has significant implications for efficiency, and makes the pentode a good choice for high voltage stages.

Thirdly, the shape of the V_a and I_a curves for the pentode (and transistor) is exponential so that:

$$I_a \propto (1 - e^{-kV_a})$$

This relationship not only results in the pentode producing significant odd harmonic distortion, but also the harmonics extend further up the spectrum than for a triode. As an example, an E55L pentode was biassed to I_a=50 mA with a

4.7 kΩ anode load from a 410 V supply. The stage clipped at ≈73 V_{RMS}, so it was tested for distortion at an output of ≈50 V_{RMS}, whereupon the stage produced 1.3% Total Harmonic Distortion (THD), but note that the distortion spectrum contains significant harmonics up to the 12th (see Figure 2.18).

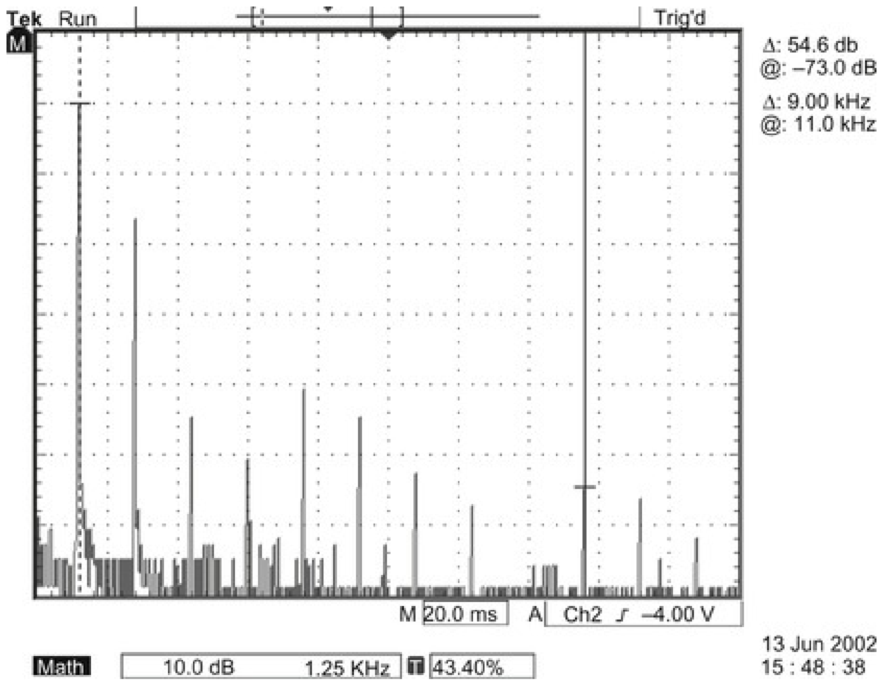


Figure 2.18 E55L pentode distortion spectrum.

By contrast, the shape of the triode anode curve is a power law:

$$I_a \propto V_a^{3/2}$$

This equation can be approximated using a binomial series, and although it contains both odd (x^3, x^5, …) and even (x^2, x^4, …) terms, indicating odd and even harmonics, the terms die away very rapidly (the author has not normally needed to look beyond the sixth harmonic when testing triodes). We can, therefore, expect the triode to produce predominantly second harmonic distortion.

The type of distortion produced is significant because the ear is far more tolerant of even harmonic distortion than of odd, not only because the ear itself produces even harmonic distortion, but also because the higher odd harmonics are no longer musically related to the fundamental and sound discordant. The measured distortion of a pentode amplifier, therefore, needs to be much lower than that of of a triode amplifier because the subjective effect is so much greater, and such amplifiers generally use plenty of negative feedback.

Using the EF86 Small-Signal Pentode

We can now consider how we would use the EF86. R_L is chosen in the normal way, in conjunction with loadlines and the 210 V HT; in this example, R_L=47 kΩ and the operating point is at 108 V (see Figure 2.19).

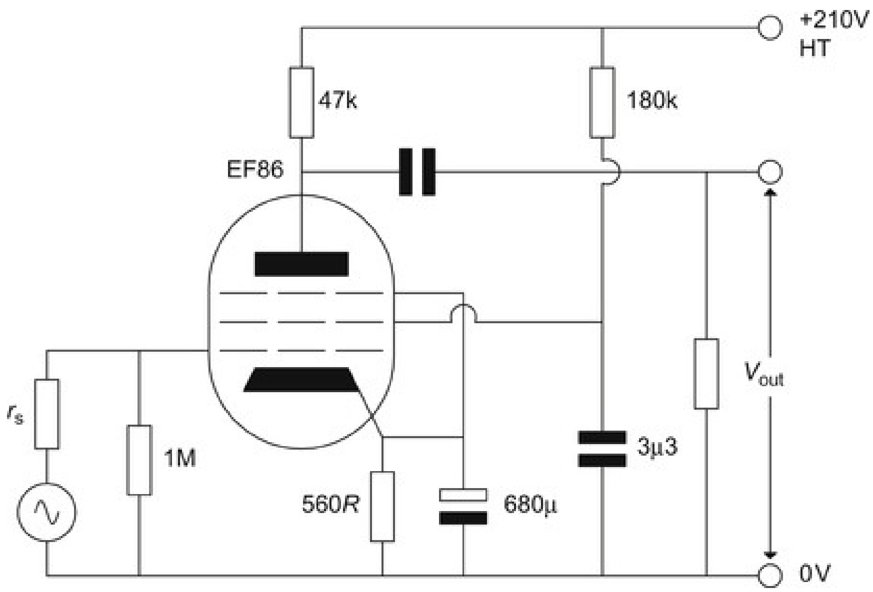


Figure 2.19 A small-signal pentode amplifier.

When we come to calculate the gain, we find that the anode characteristic begins to curve as we reach its intersection with the loadline. It is perfectly valid to treat the anode curve as a straight line, and to project this line onto our loadline in order to find the *small-signal* gain, thus giving a gain in this example of 90 (see Figure 2.20).

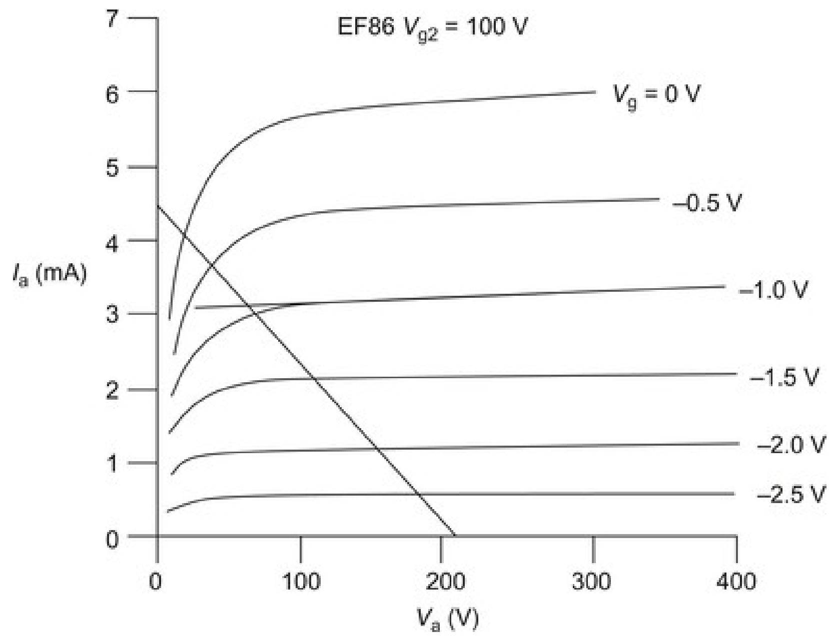


Figure 2.20 Anode characteristics and determination of gain of the pentode.

R_{g_2} is chosen either by a detailed perusal of the full datasheets, or by observing that g $_2$ current is generally a fixed proportion of anode current. For the EF86, this proportion is ≈1:4. Therefore, if the anode voltage and the g $_2$ voltage are to be similar, the g $_2$ resistor should be equal to 4 R_L, and 180 kΩ is, therefore, appropriate. The proportionate method is quicker, but for power valves we must resort to the datasheet.

Although termed a grid, g $_2$ behaves as an anode in that it receives electrons, and it must, therefore, have an 'anode' resistance. We need to know this resistance in order to calculate the value of capacitor required to hold g $_2$ at AC ground potential. Unfortunately, the datasheets for pentodes do not always give $\mu_{g_1-g_2}$, $g_{m(g_2)}$ or r_{g_2}, but these can be deduced from triode-connected valve data (g $_2$ connected to anode):

$$\mu_{g_1-g_2} \approx \mu_{\text{triode}}$$

Remembering that g_m describes the controlling effect of V_{gk} on I_k, once the electrons have left the control grid/cathode region, their numbers are fixed, and the density of the g $_2$ mesh simply determines how the cathode current is split between anode and g $_2$. Thus:

$$g_{m(g_2)} = \frac{I_{g_2}}{I_a} \cdot g_{m(a)}$$

$$r_{g_2} = \frac{I_a + I_{g_2}}{I_{g_2}} \cdot r_{a(\text{triode})}$$

Using the *triode* curves for the EF86, at V_a=108 V, V_g=1.5 V and r_a≈14 kΩ, $r_{g_2} \approx 70\ \text{k}\Omega$. This 70 kΩ is in parallel with R_{g_2}(180 kΩ), giving a final resistance of ≈50 kΩ, and so for a 1 Hz cut-off, C_{g_2}=3.3 μF.
For a pentode, $I_k \neq I_a$, and we must sum I_a (2.17 mA) and I_{g_2} (0.54 mA) to find I_k (2.71 mA), before we can calculate R_k. V_{gk}=1.5 V, so the cathode bias resistor must be 560 Ω.
Evaluating g_m from the anode characteristics, by holding anode voltage constant, and measuring the change in anode current for grid voltage produce a value of about 1.95 mA/V. For the pentode, the cathode resistance r_k=1/ g_m, and as this is in parallel with the 560 Ω R_k, we would need a 680 μF decoupling capacitor for a 1 Hz cut-off.
We can also use this value of g_m in an alternative method of calculating the gain, which is given by the following equation:

$$A_v = g_m \cdot R_L$$
$$A_v = 1.95 \times 47$$
$$A_v = 92$$

The loadline gave a gain of 90, so the agreement is good. Note that this equation does not work for triodes because it assumes infinite r_a.
C_{ag} for the EF86 is given as <50 mpF, which is a rather quaint way of writing 50 fF (femtofarads, 10^{-15} F). You might wish to consider how Mullard measured a value of capacitance this small in 1955. *Clue*: You probably would not measure it directly.

$$C_{\text{Miller}} = (90+1) \times 50 \times 10^{-15} = 4.6\ \text{pF}$$

This is a dramatically reduced value compared to the triode, but because it is so small, we must now consider stray capacitances that were previously insignificant.
Since the control grid g_1 is near to the cathode, it must have significant capacitance to the cathode, which, since we have bypassed it with a capacitor, is at ground potential. In the datasheet, a value for C_{in} is given, which is the capacitance from the grid to all other electrodes *except* the anode and is, therefore, the value of stray capacitance within the valve. For the EF86, C_{in} is 3.8 pF, which gives a total input capacitance (due to the valve) of 8.4 pF. Realistically, we ought to add a few pF for wiring capacitance, so a value of 11.5 pF would be a reasonable total figure.
The ECC83 triode gave a value of 115 pF, so in this respect the pentode is 10

times better. In summary, the pentode has greater gain, greater output voltage swing and dramatically reduced input capacitance compared to a triode. So why don't we use them all the time?

We have already seen the undesirable distribution of harmonics in pentode distortion, but the real killer for small-signal pentodes is noise.

Mullard described the EF86 as a 'low-noise pentode', and in a very strict sense this is true because it *is* low noise by pentode standards. By triode standards, it really isn't very good because of *partition noise*.

This is the *additional* noise, compared to the triode, that is generated by the electron stream splitting to pass either to the anode, or to g_2. This additional noise is related to the ratio of anode to screen grid current and to the mutual conductance of the *screen* grid; typically, this makes a given pentode 6–14 dB noisier than the pentode connected as a triode. (Connected as a triode, the EF86 is actually quite a good triode.) Even worse, partition noise has a 1/ *f* frequency distribution, which means that its amplitude rises as frequency falls, which has been found to be particularly irritating to the ear.

The Cascode

What we would like is a valve, or a *compound device*, that gives all the advantages of the pentode with none of its disadvantages – this compound device, known as the *cascode*[5], was invented for use as a high-gain error amplifier in power supply regulators but quickly became very popular as a pentode substitute (see Figure 2.21).

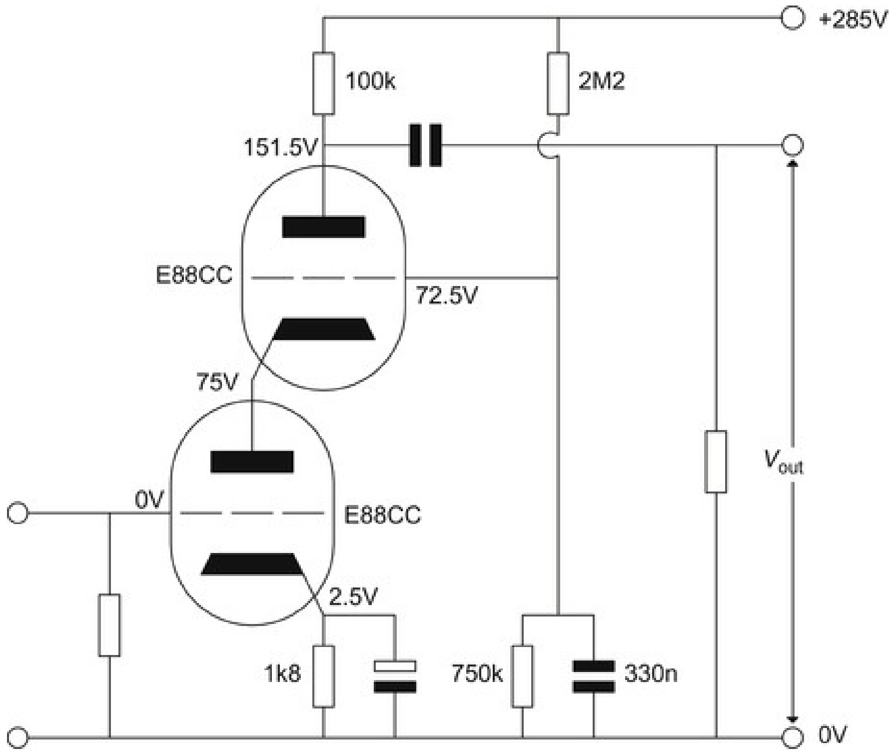


Figure 2.21 The cascode.

The cascode bears considerable similarity to the pentode in that there is an arrangement of components (R_1, R_2 and C_1) that looks very much like a screen grid bias supply, and, indeed, this is what it is. The device has a very high r_a, approximately equal to the r_a of the lower valve, multiplied by (μ+1) of the upper valve.

Operation is as follows: The upper valve has an anode load R_L, as usual, but instead of modulating V_{gk} by varying the grid voltage, and holding the cathode constant, we vary the cathode voltage, but hold the grid constant. The upper grid is biassed to whatever voltage we feel is necessary for linear operation of the upper valve and is held at AC ground by the capacitor. This is significant because it means that the cathode is screened from the anode by the grid, so Miller capacitance is not a problem. Because we are changing the *cathode* rather than the grid voltage, this part of the stage is non-inverting.

Although the upper valve has a grid in the way of the electron stream, it does not draw current, so partition noise does not occur.

The lower valve operates as a normal common cathode stage, except that it has as its anode load the cathode of the upper valve. Because the dynamic resistance looking into the cathode is low, the gain of the lower valve to its anode is low, so its Miller capacitance is also low. Another way to view the cascode is to consider that both the cascode and the pentode seek to screen the changing voltage across R_L from the sensitive input circuit, and thereby reduce C_{in}. The pentode does this by adding an internal screen between the input grid and the anode and directly reduces C_{ag}, whereas the cascode grounds the grid of the upper valve (which then acts as a screen) and drives the upper cathode from the lower valve.

Because the lower valve has a low value of load resistance, it would generate considerable distortion if it were allowed to swing very many volts. Fortunately, most of the gain is provided by the upper stage, and so distortion of the lower stage should not be a significant problem.

An important point to note with cascodes is that the only general purpose valve that was *designed* to work well in a cascode is the ECC88/6DJ8 or E88CC/6922 (special quality version). Try other valves, by all means, but do not expect the performance to be as good.

We will now see how to design a cascode. It is usual to operate the lower anode at about 75 V, so if we have a 285 V HT, this leaves 210 V across the upper valve.

We can choose an anode load for the upper valve and draw a loadline in the usual way. In this case R_L=100 kΩ and V_g=−2.5 V, causing V_a=−76.5 V, which gives a particularly linear operating point. The anode current is, therefore, 1.34 mA (see Figure 2.22).

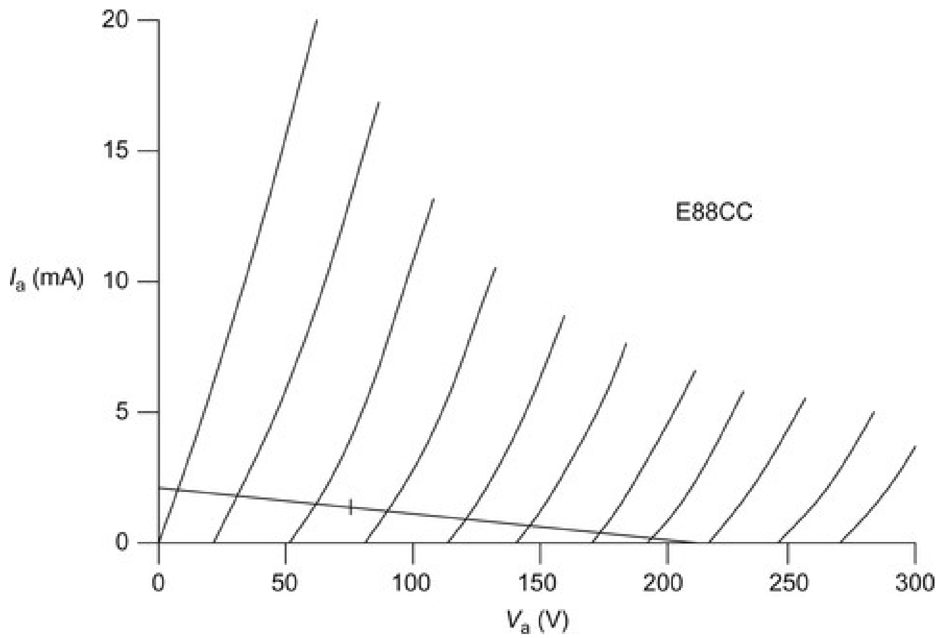


Figure 2.22 Choice of operating point of the upper valve of a cascode.

If the anode of the lower valve is to be operated at 75 V and the upper valve has a V_{gk} of −2.5 V, then the grid of the upper valve must be at 72.5 V. Since the grid of the upper valve does not draw any current, its voltage is set by the potential divider, and completely determines the conditions of the upper stage, which is working in grid or *fixed*-bias mode. We still have to be careful not to exceed the maximum permissible grid-leak resistance of the upper valve, which for an E88CC/6922 is 1 MΩ, but the Thévenin resistance of the potential divider is 560 kΩ, so we are well within limits. (We have assumed that the DC resistance of the power supply is zero in making this calculation.) We only need a 0.33 μF capacitor to make the grid a short circuit to ground as far as AC is concerned ($f_{-3\ \mathrm{dB}}$=1 Hz) compared to 3.3 μF for the EF86 g_2 capacitor.

Attempting to investigate the lower stage using anode characteristic curves is not very helpful. Instead, we will use the *mutual characteristics* of anode current against *grid* voltage (see Figure 2.23).

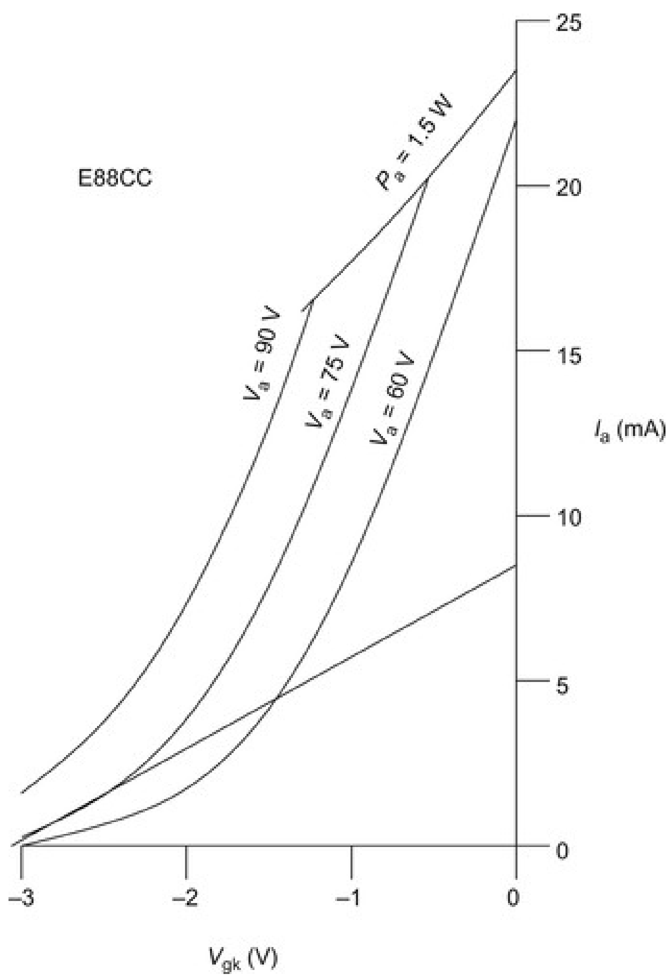


Figure 2.23 Triode *mutual* characteristics.

We know that V_a of the lower valve is 75 V, so we can look along the curve for V_a=75 V until we come to the point where I_a=1.34 mA (upper and lower anode currents are equal); this is the operating point of the lower valve and gives a V_{gk} of about 2.6 V. Plotting the point V_a=75 V and I_a=1.34 mA on the anode characteristics gives V_{gk}=2.4 V, so the agreement is not too bad. From this, we could take an average value of 2.5 V, and calculate the value of R_k at 1.8 kΩ.

Because the cascode is made up of one stage that is non-inverting and one that inverts, the output is inverted with respect to the input. The gain of a cascode, where V_1 is the lower valve and V_2 is the upper valve, with equal anode currents is:

$$A_v = \frac{1}{(1/g_{m_1} \cdot R_L) + (r_{a_2} + R_L/R_L)(1/\mu_1(\mu_2 + 1))}$$

This unwieldy equation is commonly approximated to $g_{m_1} \cdot R_L$. From the equation, we see that we need to find g_m for the lower valve. This is easily done using the mutual characteristics, by measuring the gradient at the operating point.

$$g_m = \frac{\Delta I_a}{\Delta V_g} = \frac{8.35}{3.08} = 2.7\ \text{mA/V}$$

We need r_a for the upper valve, but we are not sitting conveniently on a grid line, so we must interpolate (guess). We could do this either by taking an average of the values either side of the operating point (if they are symmetrical about the operating point), or we could use a French curve to draw a new grid curve where we need it (quite a good method). In this instance, we will take an average value (see Figure 2.24).

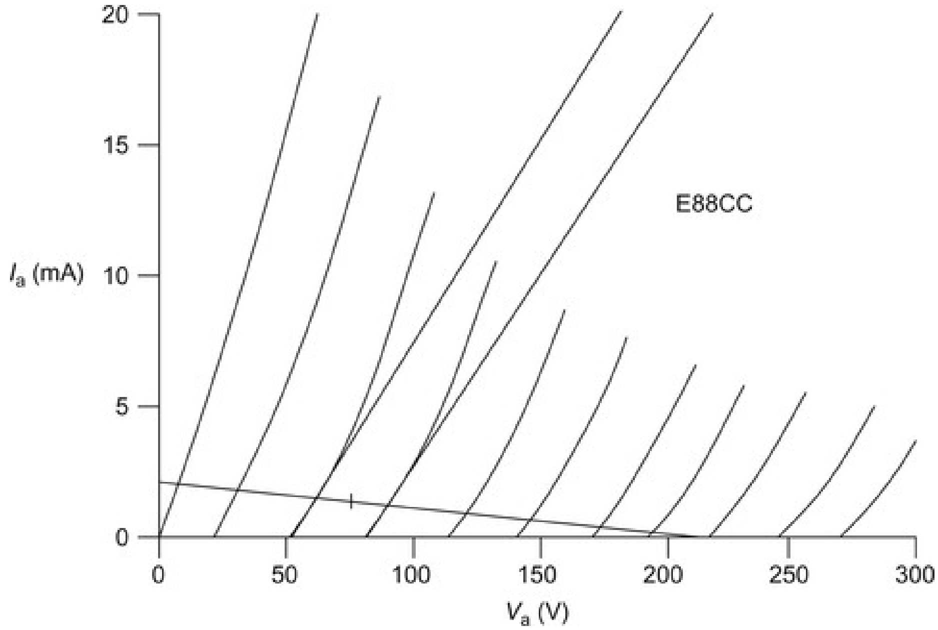


Figure 2.24 Averaging two values of r_a to find intermediate value.

$$r_a(v_g = -2V) = 5.5\ \text{k}\Omega$$
$$r_a(V_g = -3V) = 6.45\ \text{k}\Omega$$

Therefore, we will say that at V_g=2.5 V, r_a≈6 kΩ.

At the operating point of both valves, μ=32.5.

Putting all these values into the equation yields a gain of 214. Using $g_{m_1} \cdot R_L g$ would have given a gain of 270, which is 2 dB high; nevertheless, the approximation is useful because it quickly tells you whether it is worth pursuing the design any further.

We can now use this value of total gain to calculate the gain of the lower stage (if we wish). This is a useful exercise because it allows us to find the voltage swing on the lower anode. From this, we can check linearity (which might be problematic) and Miller capacitance. We can read the gain of the upper stage from the loadline, which gives us a gain of 30, so the gain of the lower stage must be 7.1.

Since C_{ag}=1.4 pF for E88CC, the Miller capacitance must be:

$$C_{\text{Miller}} = (7.1 + 1) \times 1.4\ \text{pF} = 11.3\ \text{pF}$$

As with the pentode, we should add the strays, 3.3 pF for the internal (valve) strays and 3 pF for external strays, to give a total value of 18 pF. This is not quite as good as the pentode that we saw earlier, but if the pentode had used a 100 kΩ anode load, its gain and Miller capacitance would have doubled and the answers would then have been comparable.

The value of cathode bias resistor is calculated in the normal way for a triode, but the load resistance seen by the lower anode is so low as to be negligible, so looking into the lower cathode, we find:

$$r_k = \frac{r_a + R_L}{\mu + 1} \geq \frac{r_a}{\mu} \geq \frac{1}{g_m}$$

which is the same as for the pentode. In our example, we assumed $r_a \approx 6$ kΩ and $\mu = 32.5$, so $r_k \approx 185$ Ω, and this is in parallel with $R_k = 1.8$ kΩ, so for our usual 1 Hz cut-off, we require 950 μF – the nearest standard value of 1,000 μF will be fine.

We need not use equal values of anode current in the upper and lower valves. Adding a resistor from the HT to the lower anode allows additional current to flow into the lower valve. This is useful because increased lower anode current increases gain (by increasing g_{m_1}) and improves linearity (see Figure 2.25).

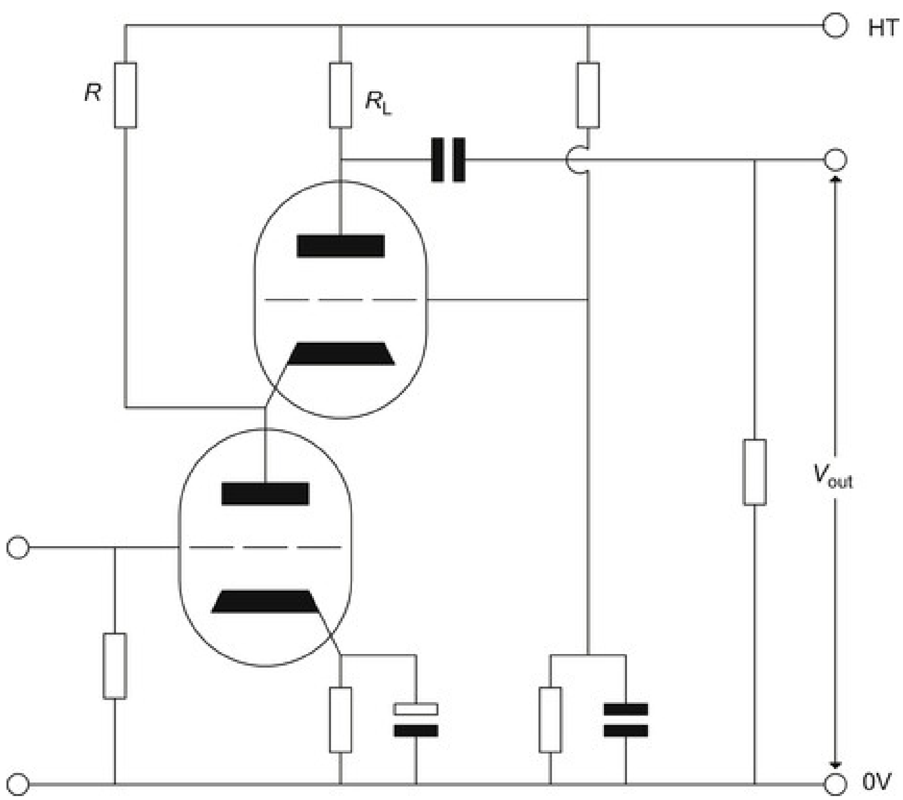


Figure 2.25 Increasing I_a of the lower valve in a cascode.

As an extreme example, we might need a low noise, low distortion cascode using half of a 6SN7 dual triode as the upper valve, so we might set its anode current to 8 mA (good linearity at this current). However, if the lower valve was a triode-strapped E810F, passing 45 mA, an additional 37 mA would be required. If V_a=100 V for the E810F and the HT=400 V, then:

$$R = \frac{V}{I} = \frac{400 - 100}{37} = 8.1\ \text{k}\Omega$$

Rather than applying fixed bias to the upper valve, self-bias can be implemented by adding a conventional cathode bias network to the upper valve, plus the normal grid-leak resistor, but bypassing the grid to ground with a capacitor (see Figure 2.26).

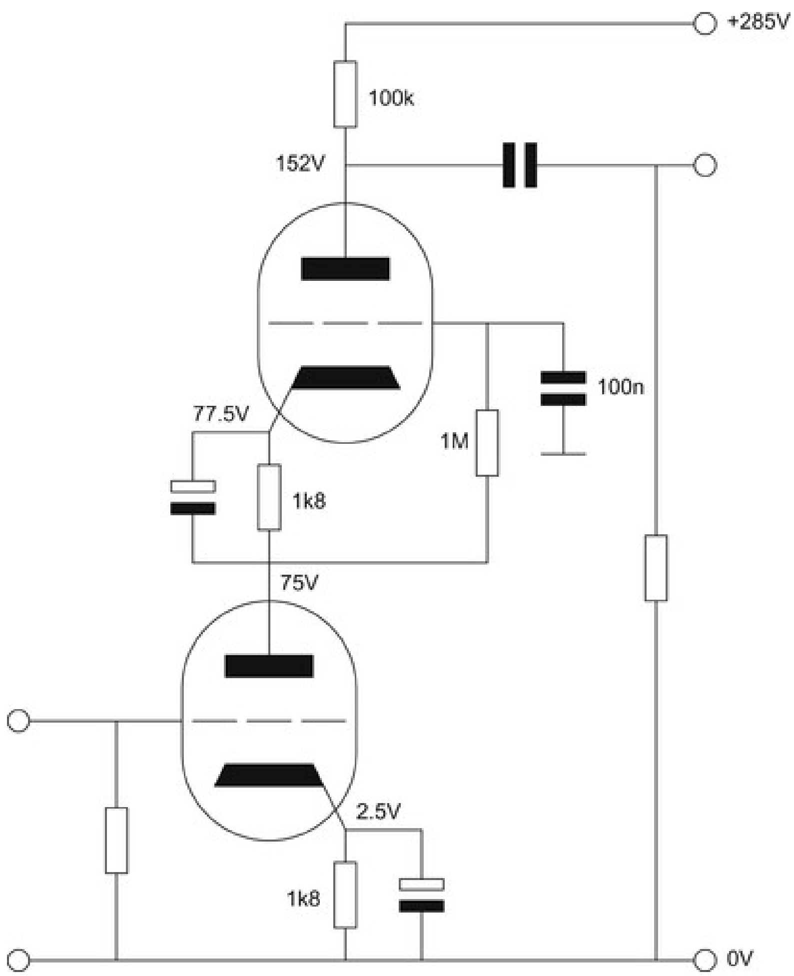


Figure 2.26 Self-biassed cascode.

The self-biassed cascode reduces sensitivity to power supply noise at the expense of less tightly defined DC operating conditions compared to the fixed-bias variant.
Cascodes have such a high output resistance that they are almost invariably DC coupled to a cathode follower (whose detailed design we will investigate very shortly), and a very elegant modification [6] is to derive the upper grid's

required bias voltage from a tapping in the cathode follower's load resistance (see Figure 2.27).

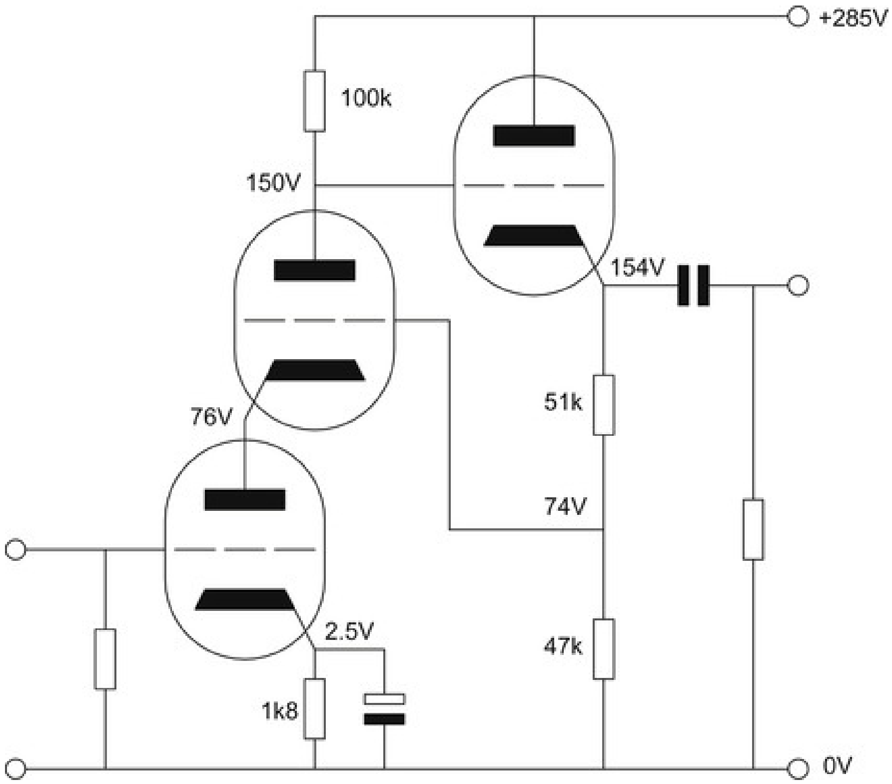


Figure 2.27 Cascode biassed from cathode follower.

The advantage of this connection is that not only does it supply the required DC bias with an absolute minimum of components (and associated time constants), but it also applies negative feedback to the upper valve all the way down to DC, reducing its distortion and further lowering the cathode follower's output resistance. As pointed out in the patent, the feedback reduces the gain of the upper valve and, therefore, increases the resistance seen looking into its cathode, which increases the gain of the lower valve, making noise in the upper valve less significant. However, the price paid for this improvement is that increasing the lower valve's gain inevitably increases its Miller capacitance.

All topologies that involve operating cathodes at voltages significantly above ground have problems because of heater/cathode leakage currents and the maximum allowable heater to cathode voltage V_{hk} (see Chapter 4). It is not uncommon for the cathode of a valve to be unbypassed and, therefore, have signal voltages on it, but in the cascode, the gain to the upper valve's cathode is low, and we are using the device because of its good noise performance, so it is likely that the signal voltage on that cathode is very small, perhaps only a few millivolts. Leakage currents via the heater/cathode insulation become worse as V_{hk} rises, so the combination of V_{hk}=75 V and a small-signal voltage means that the effects can be significant. The author once made a circuit using valves that were rated at $V_{hk(max)}$=150 V, operated the valves at V_{hk}=120 V and suffered low-frequency noise, which was only cured by sitting the relevant heaters on a 150 V DC supply. There is an understandable reluctance to do this because it means that we need two or more heater supplies, one connected to ground as normal and the other connected to an elevated voltage. We will return to this practical problem later.

The Charge Amplifier

The pentode and cascode sought to minimise C_{ag} by screening in order to avoid High Frequency losses. But a more subtle approach when dealing with an awkward parasitic component is to see if it can be persuaded to perform a different function. The charge amplifier was introduced in Chapter 1 as an inverting op-amp configuration, but a triode can be configured to become a charge amplifier (see Figure 2.28).

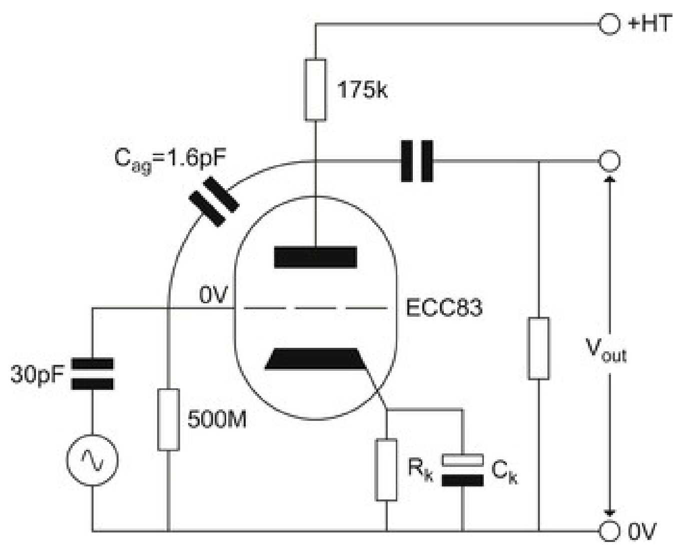


Figure 2.28 Triode charge amplifier.

The significance of using a triode as a charge amplifier is that C_{ag} becomes an intentional feedback component, making gain independent of frequency. Because the amplifier is a feedback amplifier, we must first determine the gain of the common cathode amplifier within the feedback loop, and then apply the feedback equation to determine the final gain. In this example, the open-loop gain A_0 is 73. Remembering that capacitive reactance is inversely proportional to frequency, we invert the equation for determining the feedback fraction:

$$\beta = \frac{C_{ag}}{C} = \frac{1.6}{30} = 0.0533$$

Gain after feedback becomes:

$$A = \frac{A_0}{1+\beta A_0} = \frac{73}{1+0.0533\times 73} \simeq 15$$

This may not seem very exciting, but the circuit is particularly useful as the input amplifier following a condenser microphone. 30 pF is a typical large capsule capacitance, and it is this capacitance that is used as C_{input} – so a built microphone charge amplifier looks for all the world like an ordinary common cathode amplifier. Noting that the feedback is shunt-derived and shunt-applied, it acts to reduce impedances at the point of derivation and application, so it reduces anode impedance (making the amplifier better able to drive cables) and grid impedance (making it less susceptible to leakage currents).

Stray capacitance to ground attenuates the signal from the capsule, directly degrading the signal-to-noise ratio of the microphone, but because this amplifier causes the grid to become a virtual earth, the effect of stray capacitance from grid to earth is proportionately reduced by the feedback factor, minimising its effect on signal-to-noise ratio.

Large capsule microphones are popular for saxophones and singers, both of which are quite loud and may need attenuation to avoid overloading either the capsule amplifier or the mixing desk. A switched capacitor from anode to grid increases feedback, reducing gain without degrading noise, and this is how the common '−10 dB' switch works. Some designers may permanently incorporate an additional anode to grid capacitor to optimise the amplifier's gain and feedback, so please don't incense Tim de Paravacini by removing them and destroying that delicate design balance.

The condenser microphone is the ideal audio application for the charge amplifier, yet there is a very small fly in the ointment. The grid still needs a grid-leak resistor to define DC conditions, but to avoid it causing low-frequency loss it must be very large, and 500 MΩ is not uncommon. We saw earlier that grid current limits the maximum permissible value of grid-leak resistor to avoid anode current runaway, so valves intended for use as microphone charge amplifier must be selected for low grid current. Fortunately, this tends to be a natural consequence of the inevitable selection for low grid current noise.

The Cathode Follower

The circuits that we have considered up until now have been concerned exclusively with providing voltage gain. Sometimes we need a *buffer* stage that provides high input and low output resistance. The cathode follower [7] has a voltage gain of slightly less than 1, a low output resistance, typically ≤1 kΩ, a high input resistance (≈500 MΩ in condenser microphones) and is non-inverting. We will consider the fixed-bias version of the cathode follower first (see Figure 2.29).

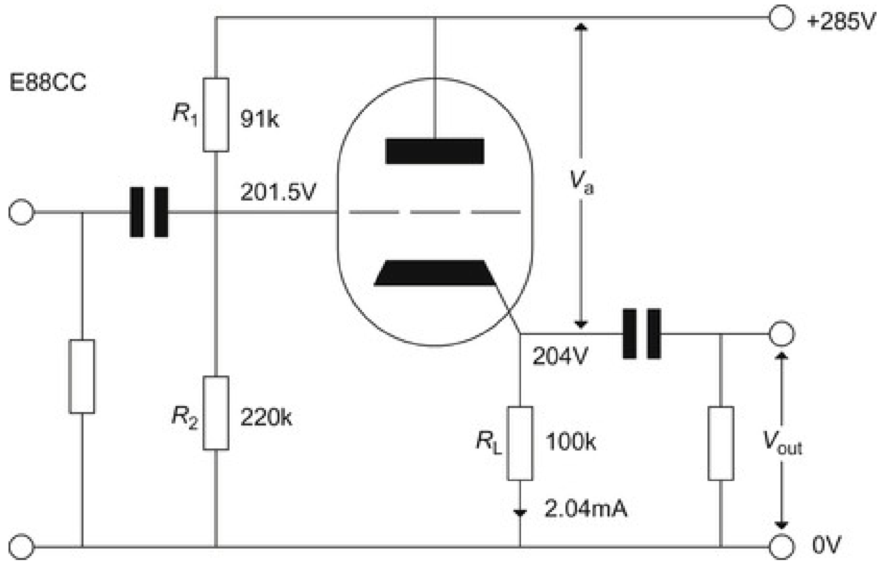


Figure 2.29 Fixed-bias cathode follower.

Compared to the common cathode amplifier, we have changed the position of the load resistor so that the output is now taken from the cathode, but the circuit can still be analysed in the same way as before, using loadlines (see Figure 2.30).

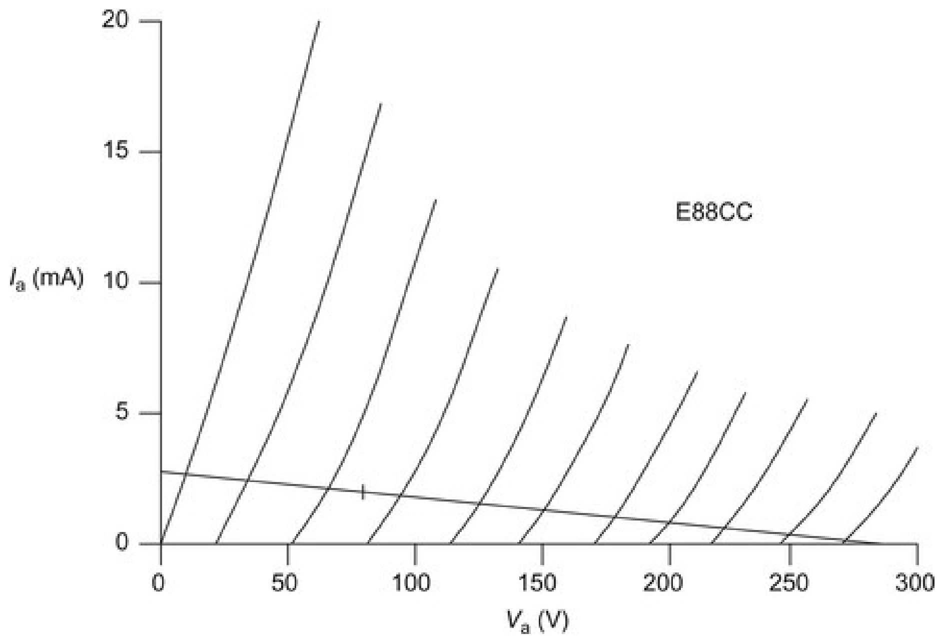


Figure 2.30 Operating point of fixed-bias cathode follower.

R_L=100 kΩ, and so we draw the appropriate loadline, V_g=−2.5 V, with V_a=−81 V, because of the excellent linearity in this region. Remembering that V_a is actually the *anode-to-cathode* voltage, the cathode is now at 285−81 V=204 V, and because V_{gk}=−2.5 V, the grid must be at 201.5 V to bias the valve to this condition. This voltage is set by the potential dividers R_1 and R_2.

The cathode follower is simply a special case of the common cathode amplifier with 100% negative feedback (parallel-derived, series-applied). To find the gain after feedback, we use our normal technique of measuring the gain from the loadline (A_v=28.5), and apply the feedback equation:

$$A_{\text{fbk}} = \frac{A_0}{1 + \beta \cdot A_0}$$

Since we have 100% feedback, β=1, and the gain of our example becomes 28.5/29.5=0.97.

Alternatively, we can combine the common cathode amplifier's gain equation with the feedback equation to calculate cathode follower gain directly:

$$A = \frac{\mu}{\mu + 1} \cdot \frac{R_k}{R_k + r_a}$$

This equation shows explicitly that the gain of an ideal cathode follower tends towards unity and that this is best achieved if μ and R_k are as large as possible.

We saw earlier that the AC resistance at the cathode was:

$$r_k = \frac{R_L + r_a}{\mu + 1}$$

But for a cathode follower, R_L from the anode to the HT is 0, so this equation is frequently approximated to $1/g_m$. From the anode characteristics, $g_m \approx 5$ mA/V; this gives an output resistance of ≈200 Ω. This is not a particularly accurate answer, since the method of determining g_m was crude, but this does not matter, since it is usual to operate an audio cathode follower with ≈1 kΩ resistor in series with its output to ensure High Frequency stability into capacitive loads – so this swamps the slight inaccuracy. Nevertheless, 1.2 kΩ is a low output resistance for a valve stage.

As shown, the stage does not have a particularly high input resistance because the fixed-bias network is in parallel with the input, although this configuration is useful for making Sallen & Key active filters. If we need a high input resistance, we must change our bias arrangements (see Figure 2.31).

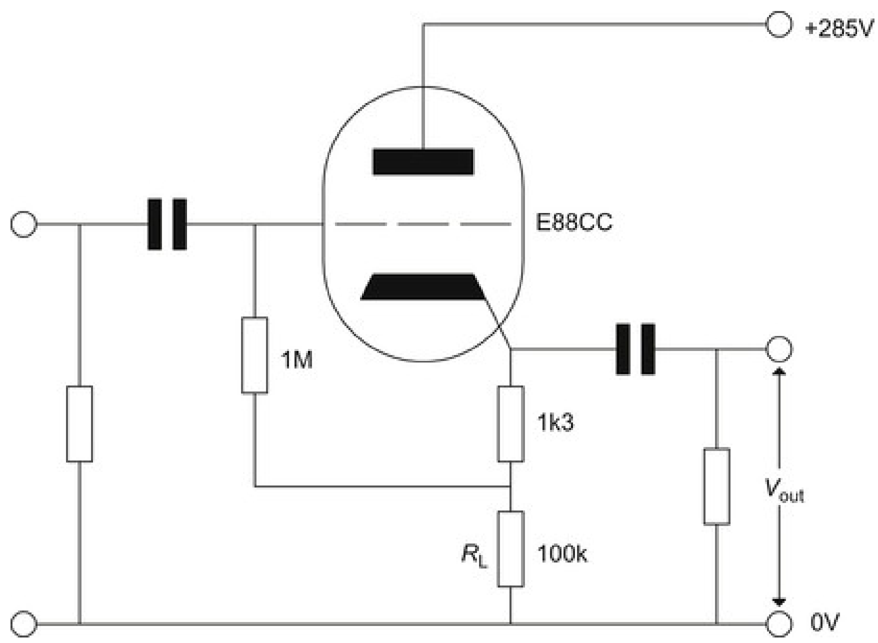


Figure 2.31 Cathode bias cathode follower.

We now have cathode or *self*-bias provided by the 1.3 kΩ resistor, whose value is calculated in the normal way. You will note that by adding this resistor, we have slightly increased the value of R_L and, indeed, this was also the case in the common cathode amplifier, but this ≈1% increase has a negligible effect on DC conditions.

At first sight, this configuration is very little better than the fixed-bias configuration, as the input resistance appears to be only 1.1 MΩ. However, the

1 MΩ grid-leak resistor has been *bootstrapped*, which is to say that the entire input signal does *not* appear across it.
It works like this: We have just calculated the gain A_v to the cathode as being 0.97. We can calculate the attenuation of the potential divider formed by the cathode bias resistor and R_L as being 0.987, so the proportion of input-signal voltage at the lower end of the grid-leak resistor is 0.96 V_{in}. Now, since the output of a cathode follower is *non-inverting*, this means that there is only 0.04 V_{in} *across* the grid-leak resistor. The *signal* current through this resistor will, therefore, be only 4% of what it would have been had the grid-leak resistor been connected directly to ground. It presents an input resistance equivalent to 1 MΩ/0.04=25 MΩ. Formalising this argument:

$$r_{\mathrm{input}} = \frac{R_g}{1 - A \cdot ((R_L/R_L) + R_k)}$$

Note that A is the gain of the cathode follower, not the original loadline gain. A similar argument can be used to determine the input capacitance of the cathode follower:

$$C_{\mathrm{input}} \approx C_{ag} + (1 - A) \cdot C_{gk}$$

Note that this is an approximate value because there will be significant strays. Using our example with the E88CC:

$$\approx 1.4\ \mathrm{pF} + (1 - 0.96) \times 3.3\ \mathrm{pF} = 1.5\ \mathrm{pF}$$

We should add a few pF for wiring strays, as we did before, which brings the likely input capacitance of the cathode follower to 4.5 pF, which is rather less than half the value of the cascode or pentode.
Although the self-bias cathode follower has the highest input impedance, its DC conditions are less stable than the fixed-bias cathode follower that has its grid voltage defined by an external network. This means that fixed-bias cathode followers tend to be used in situations where their low output resistance and DC stability are important, such as directly driving output valves.
It has been suggested that the linearity of the cathode follower is questionable. It is hard to see how this accusation can be true, particularly if the operating point has been chosen carefully, as in the previous example, since the stage operates under 100% negative feedback. This means that any non-linearity will be reduced in proportion to the feedback factor (1+ βA_0), which in our example gives a reduction of 30:1.
Nevertheless, it *is* possible to do even better. We mentioned earlier that μ was one of the more stable valve parameters, whereas r_a varies considerably with

anode current. This is significant because it is mostly the variation of r_a that causes distortion, and we can see why this is if we look at the equation for the gain of a common cathode amplifier:

$$A_v = \mu\left(\frac{R_L}{R_L + r_a}\right)$$

If we could make R_L very large, ideally infinite, r_a would become insignificant by comparison and could no longer cause distortion. Provided that we have chosen a sensible operating point where μ does not vary greatly, we will then have a very low-distortion buffer. Unfortunately, if we simply make R_L very large, we find that there is such a voltage drop across it that we need an HT of more than 2 kV! (see Figure 2.32).

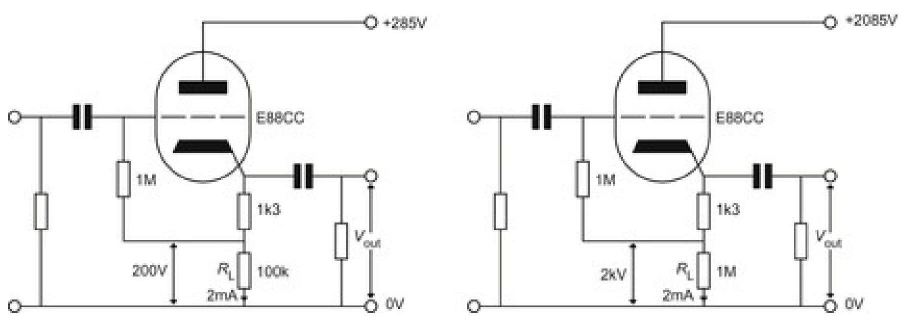


Figure 2.32 Effect of increasing R_L in a cathode follower.

We need a way around the problem of excessively large values of R_L, and to do this we need to examine some definitions.

Sources and Sinks: Definitions

A current or voltage source is a supply of energy (such as a battery) capable of supplying energy into a load whose other terminal is connected to ground, whereas a sink may *control* the characteristics of an external source of energy, but provides none of its own.

There are three important aspects that describe the performance of a constant current source or sink:

- *DC accuracy*: In some applications, the defining aspect is the accuracy with which a DC current may be set and held under changing circumstances. DC applications tend to involve chemistry, such as electroplating, where the mass deposited is assumed to be directly proportional to the total charge. Thus, provided that a current appropriate to the plating area can be accurately and reliably set, plating thickness becomes a function of time and current. (Current was once defined in terms of the mass of silver plated in a given time.)

- *Voltage compliance*: This is the range of voltage over which the source or sink is able to maintain its designed performance. The upper voltage limit is usually determined by the maximum working voltage of the device controlling the load, but it may be that power dissipation in that device limits safe operation before the explicit voltage limit is reached. Thus, a sink might employ a transistor rated at 400 V and 15 W, but if configured as a 75 mA sink, only 200 V could be allowed before the 15 W limit would be reached. The lower limit is more insidious and defines the minimum voltage required across the source or sink to ensure correct operation. The voltage across a source or sink can be increased from zero whilst monitoring current and at a certain voltage, the current will be seen to jump to its design value, and it is very tempting to take this as being the minimum voltage required. However, a more careful investigation monitoring output resistance against applied voltage would almost certainly discover that a higher minimum voltage was required than that suggested by the simple current test.

- *Output resistance*: The output resistance of a constant current source or sink is ideally infinite. However, as already noted, this cannot be maintained at all voltages or power dissipations, so it is a small-signal parameter, either described as a *slope resistance* r_{slope}, or simply as an AC resistance r.

Audio electronics often needs real-world approximations to constant current sources and sinks in order to improve the AC performance of the surrounding circuit, so the following definitions are couched in terms of their common effects on AC performance, even though practical implementations using active devices may be equally effective at DC.

A perfect constant voltage source/sink is a short circuit (zero resistance) to AC, and Ohm's law, therefore, ensures that even an infinite AC current passing through it develops zero AC voltage across it. Although active devices such as voltage regulators are frequently better, suitably sized capacitors are often inescapably used as AC approximations to constant voltage sources/sinks. Common audio examples include the reservoir capacitor in a capacitor input power supply (source), or the cathode bypass capacitor (sink).

Conversely, a perfect constant current source/sink is an open circuit (infinite resistance) to AC, and even an infinite AC voltage across it is incapable of driving any AC current through it. Active constant current sources/sinks are becoming more common, but inductors were the traditional AC approximations to constant current sources/sinks, the main audio example being the choke in a choke input power supply (source) or the primary inductance of an output transformer (sink).

Although it has been suggested that capacitors and inductors may be used as approximations to perfect sources or sinks, the implication is that topologies traditionally using these approximations may be replaced by active devices, which can be more nearly perfect. It is probably true to say that the primary difference between a valve amplifier designed in the 'Golden Age' and a modern design is that the modern design is likely to replace passive components with active devices to approximate perfect sources and sinks more closely.

The Common Cathode Amplifier as a Constant Current Sink (CCS)

We saw earlier that leaving a common cathode; amplifier with R_k unbypassed caused r_a to rise due to negative feedback. We can exploit this effect deliberately to create a CCS (see Figure 2.33).

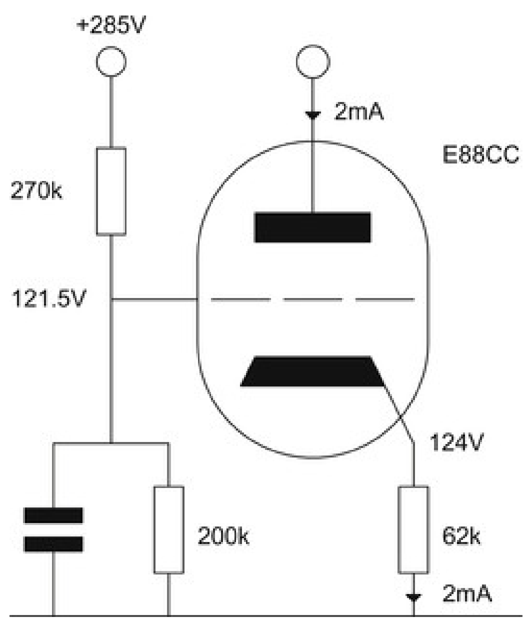


Figure 2.33 Constant current sink.

Let us suppose that we need to sink a current of 2 mA using an E88CC valve and that we have 204 V of HT available for the sink. We can treat V_a=204 V, I_a=0 mA as one end of a loadline, and plot this point on the graph (see Figure 2.34).

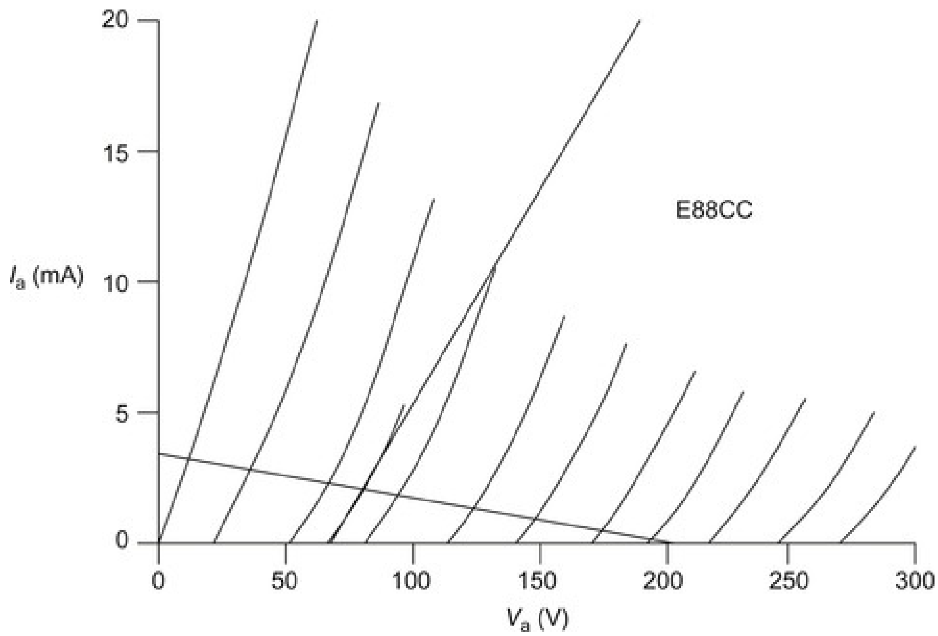


Figure 2.34 Operating conditions of constant current sink.

Plotting V_a=204 V, I_a=0 mA is easy, but we don't yet know where the other end of the loadline will be. However, we *do* know a point on the line; we know that I_a=2 mA at the operating point, although we do not know the voltage. It is up to us to *choose* a voltage, and V_a=81 V is a good choice for linearity. Linearity is still important in a CCS because the complete circuit will probably modulate the sink's anode voltage with an audio signal. If linearity is poor, this indicates non-constant r_a, which is part of the term that governs the output resistance of the sink. If the output resistance varies with applied voltage whilst it is being used as an active load for another valve, it will cause distortion in that valve. If we now draw our loadline, we can find the current through R_L when V_a=0. From this, we can calculate the value of R_L, which is then 60 kΩ. The nearest value to this is 62 kΩ, and this is what we would use.

Since I_a=2 mA, we know that the cathode of the valve will be at 124 V. V_{gk}=−2.5 V, so the grid needs to be at 121.5 V. This voltage is set in the normal way using the potential divider and capacitor combination.

The AC resistance, looking into the anode of this circuit, is:

$$r_{\text{sink}} = r_a + (\mu + 1)R_k$$

For our design, this gives a value of slightly more than 2 MΩ. Achieving this result with a pure resistance would require a 4 kV HT supply. The AC resistance is in parallel with C_{out}, and C_{ag} causes sink gain to fall, so the sink impedance falls as frequency rises. (C_{out} is the capacitance from anode to all other electrodes *except* the grid.)

Pentode Constant Current Sinks [8]

Pentodes are even better as CCSs because of their high μ, and are particularly useful when the allowable voltage drop across the sink is quite low.
If we needed a CCS of 10 mA, but were only allowed a 100 V drop across it, an E88CC could achieve an output resistance of only ≈100 kΩ, which is still a 10-fold improvement over a 10 kΩ resistor, but a pentode can do rather better.
If we leave even a 2 kΩ cathode resistor unbypassed, a pentode can increase its output resistance to >10 MΩ. This is a stunningly good CCS, but it should be remembered that pentodes tend to be noisy, so this would not be a good choice in the first stage of a sensitive pre-amplifier (see Figure 2.35).

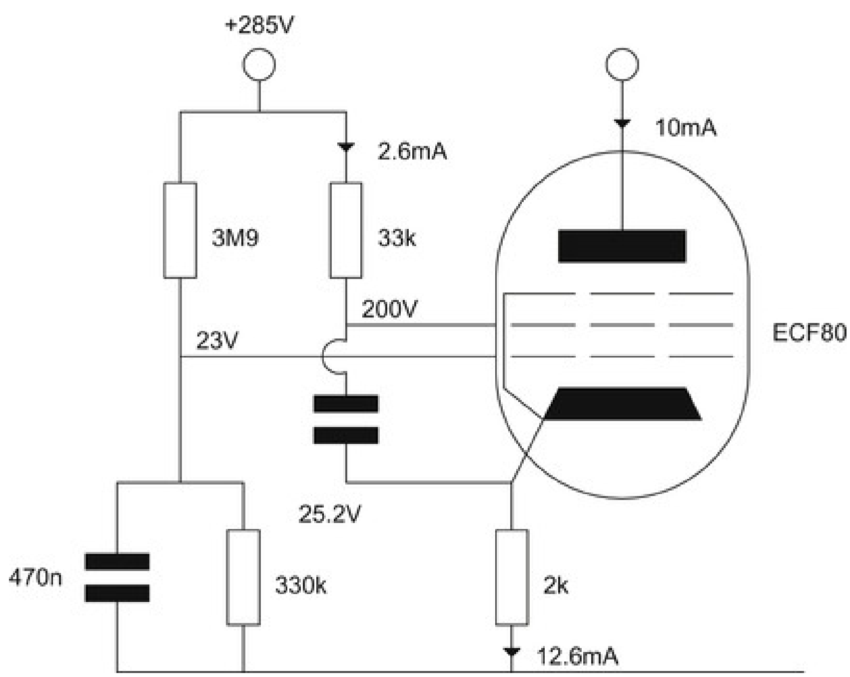


Figure 2.35 Pentode as constant current sink.

When using a pentode as a CCS, it is vital to remember that the cathode resistor passes not only the desired constant current, but *also* the g_2 current. Note also that the g_2 decoupling capacitor must be taken to the cathode, and not to ground. This is because we want cathode feedback to increase r_a, but we do not want the voltage between g_2 and the cathode to vary, as this would cause positive feedback that would reduce r_a.

Some pentodes make better CCSs than others because their anode characteristics are flatter, giving a higher output resistance, or because the flat part of their anode characteristics swings closer to 0 V. Table 2.1 shows single pentodes that are particularly suitable in CCSs.

Table 2.1 Comparison of Pentode Suitability versus CCS Current

Type	Optimum current (mA)	C_{out} (pF)	P_a (W)
EF91/6AM6	≤6	3.1	2.5
EF184/6EJ7	8–15	3	2.5
EL83/6CK6	15–30	6.6	9
EL822	20–45	6	12

Table 2.1 shows optimum currents much lower than $I_{a(max)}$, partly because at higher currents the anode curves tilt away from the horizontal, indicating reduced r_a, but mostly because the major contribution to output resistance actually comes from the unbypassed R_k, whose value is multiplied by a factor of $g_m \cdot r_a(\mu)$, which is typically 1,000 or more for a pentode. Higher currents require less bias and a reduced value of R_k, thus reducing output resistance. For maximum output resistance, it is better to use a valve with an oversized P_a, requiring a large R_k, than that with a perfectly rated P_a, requiring a smaller R_k. Unfortunately, the disadvantage of a CCS operated at very low I_a is that I_{g_2} becomes a significant proportion of I_k, making the circuit inefficient.

As an example, we might require an 8 mA CCS. However, this is the lower end of efficient operation for an EF184, and the curves show that it is likely to require $I_{g_2} \approx 3\ \mathrm{mA}$, which means that total HT current has been increased by ≈38%. If there is only one sink, then this isn't a problem, but if there are many such sinks, this can greatly increase the cost of the HT supply. If we were willing to reduce the sink current to 6 mA, this would bring it within range of the EF91, which only requires $I_{g_2} \approx 1.55\ \mathrm{mA}$ in this instance, reducing HT current from 11 mA to 7.55 mA. Although the EF91 does not have quite such attractive curves as the EF184, it is far cheaper to use, and if HT current is limited, it can be worth bending a design to allow its use.

Optimally biassed, most small-signal pentodes split I_k between I_a and I_{g_2} by ≈4:1 ratio. Thus, 8 mA of I_a typically requires 2 mA of I_{g_2}. It is very important to check that the pentode's I_a, P_a, and especially P_{g_2} are not exceeded. Successful design of pentode CCSs requires full datasheets with curves or a valve tester/rig to allow voltages to be imposed and currents to be determined experimentally (which is more reliable).

If the CCS is to be used in a stage with a low signal level, it is worth considering hum and screening. The EF184 has an integral metal screen, the EF91 has a conductive paint screen on the inside of the envelope, but the EL83 and EL822 power valves are completely unscreened.

The Cathode Follower with Active Load

You probably have realised that the requirements for the triode CCS were set by the cathode follower designed earlier, so we can now combine them to form a cathode follower with an *active load* (see Figure 2.36).

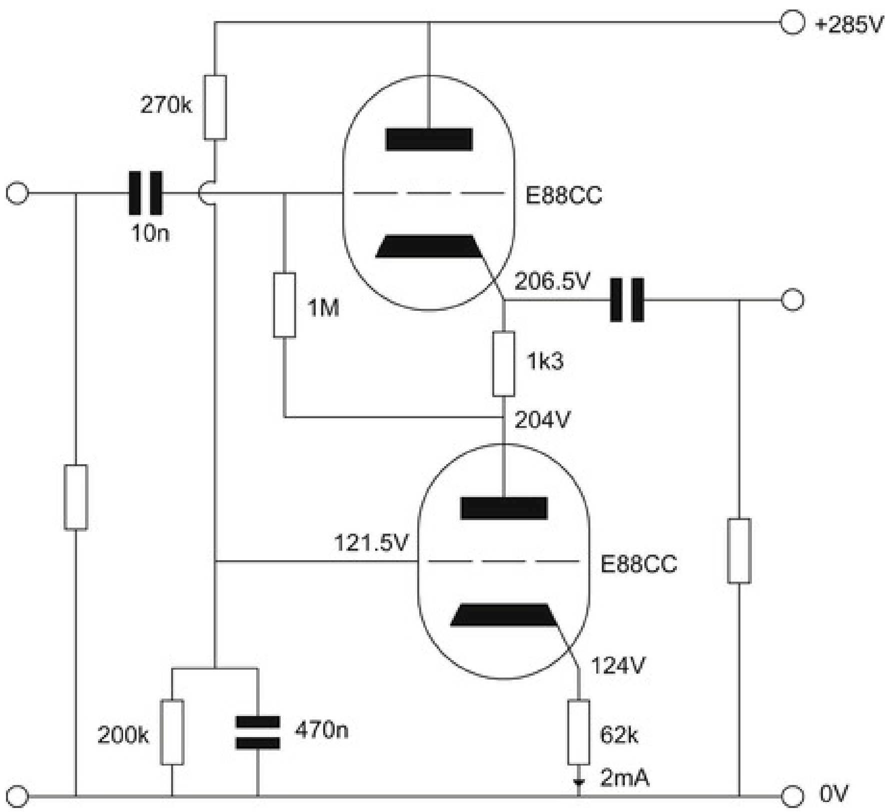


Figure 2.36 Cathode follower with active load.

Because the value of R_L for the upper stage is now so large, the gain becomes:

$$A_v = \frac{\mu}{\mu + 1}$$

The gain is, therefore, 0.97, which is only a little higher than before, but the distortion is greatly reduced. It is possible to make distortion predictions, but these are of very doubtful value, since *real* valves do not behave in the nice mathematical fashion that the equations require to generate sensible answers.

Grid current can cause a cathode follower to have very much higher distortion than expected. The author tested a self-biassed cathode follower using a 6S45 with an EF184 CCS active load, and then determined its input resistance by measuring the relative loss when driven from a 1 MΩ source resistance compared to 5 Ω. Sadly, the input resistance was not quite as high as predicted, and adjusting the value of grid-leak resistor from 150 kΩ to 1 MΩ not only changed the input resistance and slightly changed I_a (indicating grid current), but also reduced the distortion at +20dBu from 0.23% to 0.052%. Reducing the source resistance from 1 MΩ to 24 kΩ further reduced the Total Harmonic Distortion +Noise (THD+N) from 0.052% to 0.02%. Cathode followers are often used as buffers after volume controls, so this sensitivity to source resistance can be significant, particularly because, as we will see in Chapter 7, that some volume controls have higher output resistance than others.

To sum up, a carefully designed cathode follower with a resistive load produces low distortion – replacing this with an active load improves it further, and challenges test equipment, but for optimum distortion the valve should be selected/tested for low grid current.

The White Cathode Follower

Named after its inventor, the White cathode follower [9] is the basis of all output transformer-less (OTL) power amplifiers because of its low output resistance. The circuit comes in two forms: one self-contained and the other requiring an external phase splitter.

Analysis of the Self-Contained White Cathode Follower

The lower valve is fed with a signal from the upper valve, which, in turn, it feeds back into the cathode/grid circuit of the upper valve. At the input to the lower valve, the circuit may be considered to be a cascode amplifier (see Figure 2.37).

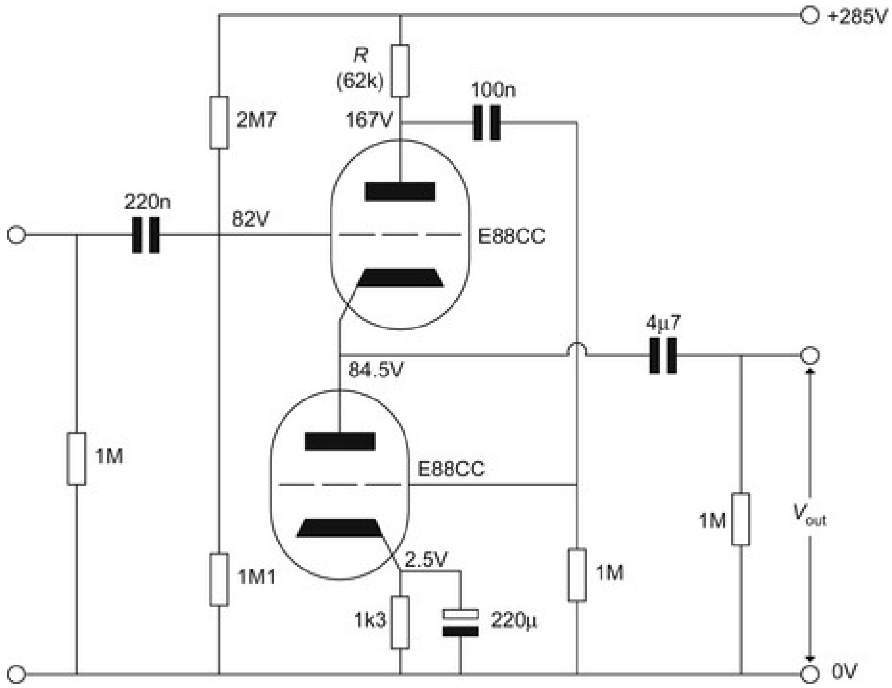


Figure 2.37 Self-contained White cathode follower.

Provided that μ is reasonably large and the cathode is bypassed:

$$A_v \approx g_m \cdot R$$

And the resistance looking into the cathode of the upper valve is:

$$r_k = \frac{R + r_a}{\mu + 1}$$

But the gain of the lower valve is devoted to reducing the output resistance at the cathode of the upper valve, so combining the two equations gives:

$$r'_k \approx \frac{R + r_a}{(\mu + 1) g_m \cdot R}$$

μ is usually rather greater than 1, even for power triodes, so if we substitute $\mu = g_m \cdot r_a$:

$$r'_k \approx \frac{R + r_a}{g_m^2 \cdot R \cdot r_a}, \text{ or } \frac{1}{g_m^2} \cdot \frac{R + r_a}{R \cdot r_a}$$

We can now recognise the $R \cdot r_a$ term as the inverted parallel combination of R and r_a. This is significant because it indicates that there is a point beyond which increasing R has no effect and that final output resistance is limited by r_a:

$$r_{out} = r'_k \approx \frac{1}{g_m^2 \cdot R'}$$

where

$$R' = \frac{R \cdot r_a}{R + r_a}$$

It should be noted that two rather dubious approximations were made to derive this result, both of which relied on high μ. The example in Figure 2.36 was optimised for low output resistance, and $R \approx 10\ r_a$, beyond which limit no practical improvement is possible.

Although superficially completely different, the self-contained White cathode follower and the shunt-regulated push–pull (SRPP) amplifier described later in this chapter are both shunt-regulated amplifiers because both valves contribute to the AC load current. Rigorous equations for gain and output resistance derived by Amos and Birkinshaw [10] are as follows:

$$A_v = \frac{\mu_1(\mu_2 R + r_{a_2})}{r_{a_2}(\mu_1 + 1) + r_{a_1} + R[\mu_2(\mu_1 + 1) + 1]}$$

$$r_{out} = \frac{r_{a_2}(R + r_{a_1})}{r_{a_2}(\mu_1 + 1) + r_{a_1} + R[\mu_2(\mu_1 + 1) + 1]}$$

where V_1 is the upper (amplifying) valve and V_2 is the lower (regulating) valve.

Using an E88CC as an example, with $g_m \approx 5$ mA/V and $\mu \approx 28$, the rigorous equation predicts $r_{out} = 6.6\ \Omega$, whereas the approximate equation is only 4% high at 6.9 Ω despite the two dubious approximations. Experimentation with a spreadsheet reveals that this version of the White cathode follower is unsuited to low- μ valves, since a 6080 ($\mu = 2$) predicts $r_{out} \approx 35\ \Omega$, which is worse than when used as a standard cathode follower ($r_{out} \approx 15\ \Omega$). However, a triode-strapped E55L ($\mu = 30$) predicts $r_{out} < 2\ \Omega$, and a triode-strapped D3a ($\mu = 80$) should easily achieve $r_{out} < 1\ \Omega$. It is easy to become excited about predicted low output resistances, but we must always remember that all such equations contain the implicit assumption that the output resistance of the HT supply is 0 Ω, which is normally only approximated by a regulated supply.

Since the White patent suggested that the circuit was particularly suitable for driving analogue video cables (which were typically 75 Ω transmission lines), it is not surprising that the stage makes an excellent output cable driver for a pre-amplifier.

Note that because the feedback that causes the low output resistance is AC coupled, output resistance rises at low frequencies not to $1/g_m$, but to:

$$r_{out(LF)} = \frac{R + r_a}{\mu + 1} \| r_a = \frac{r_a(R + r_a)}{r_a(\mu + 1) + R}$$

In this instance, r_{out} rises to 1.5 kΩ, rather than 200 Ω, which is what a normal cathode follower would achieve. The practical implication is that the stage will not short circuit low-frequency noise (such as mains hum) induced into the output cable as effectively as a stage with a true 6 Ω output resistance from DC to light.

Usually, we do not need to calculate the gain A_v precisely, and the general cathode follower approximation of $A_v = \mu/(\mu+1)$ is perfectly adequate, but if the stage was to be used as the basis of a Sallen & Key filter, the rigorous calculation of gain might be needed.

The White Cathode Follower as an Output Stage

The primary use of the White cathode follower is as an output stage for OTL amplifiers. A series resistor in either of the HT rails is a serious waste of power, so we must use the version preceded by a phase splitter (see Figure 2.38).

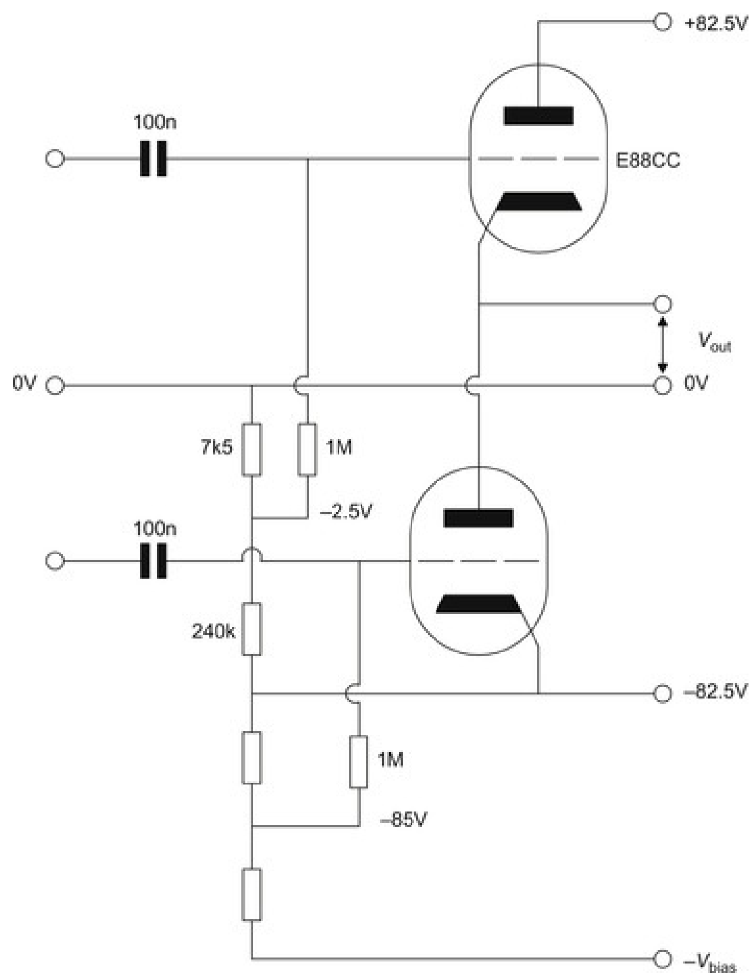


Figure 2.38 Push–pull input White cathode follower.

Assuming that neither valve switches off under any signal conditions (Class A), the gain of the lower valve is:

$$A_{v(\mathrm{lower})} = \frac{\mu \cdot R_L}{R_L + r_a}$$

The upper valve no longer has a resistor in its anode circuit, so $r_k=1/g_m$, and this is the anode load of the lower valve. Substituting:

$$A_{v(\mathrm{lower})} = \frac{\mu \cdot (1/g_m)}{(1/g_m) + r_a}$$

Multiplying by g_m and simplifying:

$$A_{v(\mathrm{lower})} = \frac{\mu}{\mu + 1}$$

A cathode follower with $R_k=\infty$ would have the same gain, and because the lower valve strives to produce exactly the same signal as a standard cathode follower if it saw $R_k=\infty$, there is no voltage difference between the two valves, so the upper valve *does* see $R_k=\infty$. However, the input to the lower valve must be inverted, requiring an external phase splitter.

The lower valve no longer reduces the output resistance of the upper valve, since with a gain of 1 it cannot apply feedback to the upper valve, which is why OTL amplifiers need considerable global feedback to bring their output resistance down to a suitable value for damping moving coil loudspeakers.

The μ-Follower

Invented by J.W. Horton [11] in 1933, but forgotten, this is a design which has attracted considerable interest since its rediscovery a few years ago [12]. (There is nothing new under the sun.) Essentially, it is a common cathode amplifier with an active load. Unlike the cathode follower, where it is arguable whether this is really necessary, the common cathode amplifier can definitely benefit from this sort of treatment (see Figure 2.39).

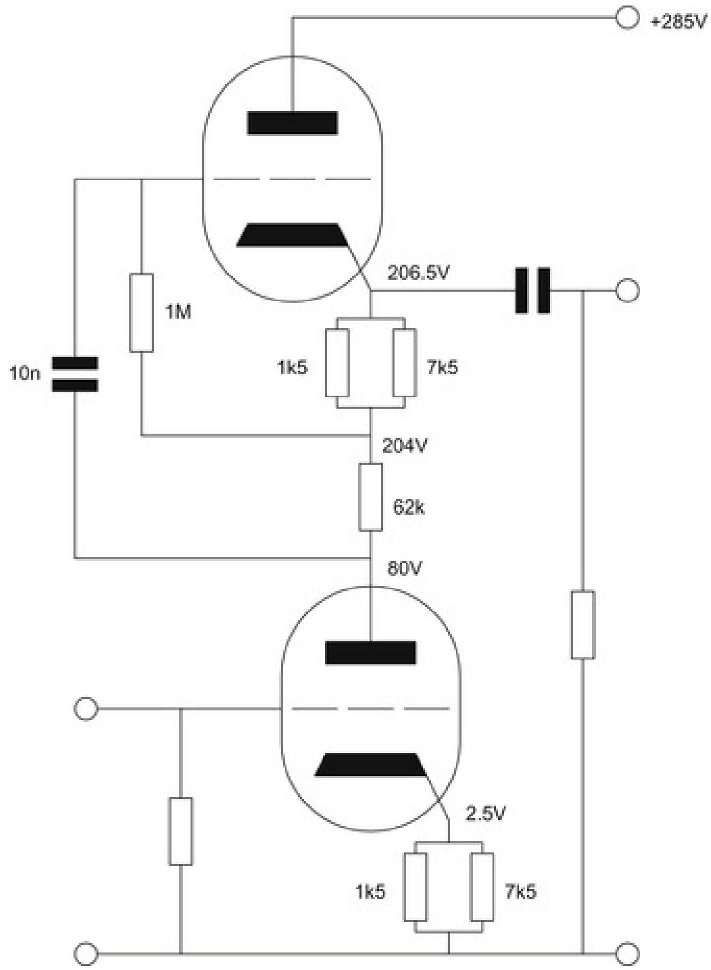


Figure 2.39 μ-Follower.

The top valve is a self-biassed cathode follower that has its input capacitively coupled from the anode of the common cathode lower stage. Since the cathode follower has $A_v \approx 1$, and is non-inverting, the signal at its cathode will be nearly equal to that at the anode of the lower valve. If this is the case, then there will be little, or no, signal voltage across the upper resistors. Little, or no, signal current flows, implying a high-resistance active load or constant current source. The lower valve achieves voltage gain $A_v \approx \mu$, and it produces low distortion (r_a is no longer a factor). As a bonus, we have two output terminals, either the direct output from the lower anode or the low resistance output from the cathode follower. It should be noted, however, that the high-resistance active load actually only operates at AC, since the coupling capacitor forms a high-pass filter in conjunction with the (admittedly high) cathode follower input resistance. If the upper valve is a constant source (even if only at AC), then we can plot the loadline for the lower valve as a horizontal line (see Figure 2.40).

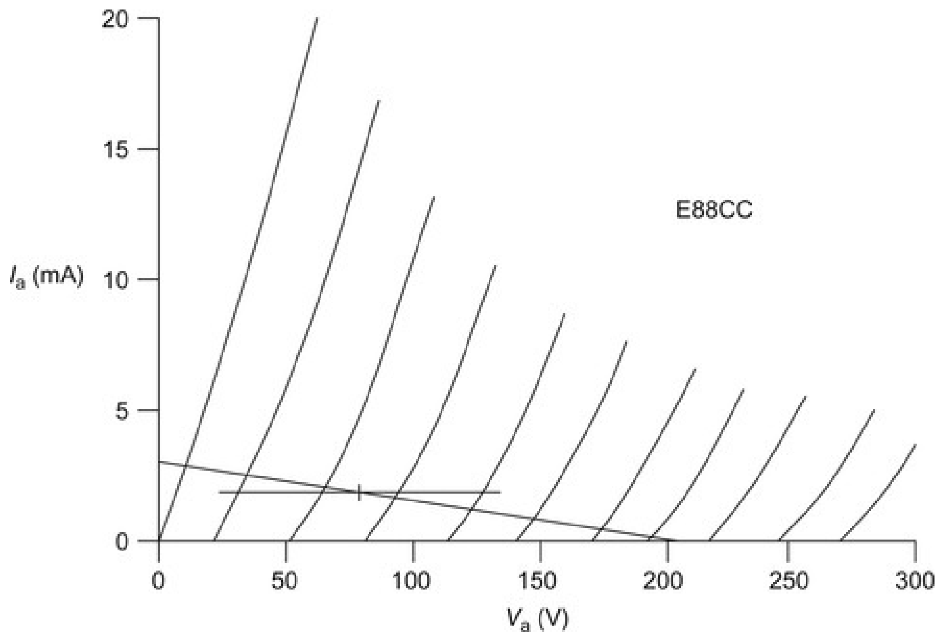


Figure 2.40 Operating conditions of lower valve in μ-follower.

This is an example of an *AC loadline*, where the slope of the loadline does not relate to DC conditions, although it must pass through the DC operating point. We can move this line to any operating point that we like. If we choose an anode current of 2 mA, and bias the anode voltage to 80 V, this gives μ=32.5, and so we would expect a gain of ≈32.

We now have to determine the operating point for the upper stage. We will supply the compound stage from an HT of 285 V, leaving 205 V for the upper stage. Since the anode currents are equal, I_a for the upper valve must also be 2 mA. If we now choose an anode voltage for the upper valve (I have chosen 80 V), we can plot the loadline. At V_a=0, we have a current of 3.25 mA, which corresponds to a 63 kΩ total cathode load for the upper valve. V_{gk} for the upper valve is 2.5 V, for I_a=2 mA; we need a 1.25 kΩ cathode bias resistance (1k5//7k5). We have now established the DC conditions of the stage.

Once we know the gain of the cathode follower, we can determine the value of the active load that it presents, and find its input resistance, which will enable us to choose an appropriate value for the coupling capacitor.

From the loadline, the gain before feedback is 29, so the gain of the cathode follower is 29/30, which is 0.97. The lower valve sees an anode load of:

$$r_{\text{load}} = \frac{R_L + R_k}{1 - A}$$

This gives a value of ≈2 MΩ, so our earlier assumptions about the gain and linearity of the lower stage were justified. We can use our earlier formula to determine the input resistance at the grid of the cathode follower:

$$r_{\text{input}} = \frac{R_g}{1 - A(R_L/R_L + R_k)}$$

This gives an input resistance of ≈19 MΩ. If we need a 1 Hz cut-off, then 10 nF is perfectly adequate. The cathode bias resistor for the lower valve is calculated in the normal way.

The lower valve's r_a and stray capacitance form a low-pass filter, and although this effect is generally negligible with medium- μ valves such as the 6J5, it becomes significant when high- μ valves such as the 7F7 are used. There are two ways in which we can inadvertently raise r_a and cause significant High Frequency loss:

- For high- μ valves in particular, the choice of operating point is a balance between maximum swing (requiring low I_a, but causing high r_a) and High Frequency $f_{-3\ \text{dB}}$ point (low r_a, but requires high I_a).
- Although the high value of load resistance for the lower valve causes β to be so small that not bothering to bypass the lower cathode resistor in a μ-follower doesn't cause noticeable loss of gain at 1 kHz, it raises r_a.

As an example of these points, a μ-follower using a 7F7 as the lower valve suffered 0.9 dB loss at 20 kHz due to the interaction between excessive r_a and stray capacitances when the cathode bypass was removed (1 kHz gain was unchanged).

An extremely useful secondary advantage of the μ-follower is its excellent immunity to noise on the HT supply, known as *Power Supply Rejection Ratio* (*PSRR*). At the output of any common cathode amplifier, PSRR can be found:

$$\text{PSRR} = \frac{R_L + r_a}{r_a}$$

This is quite simply because r_a forms a potential divider with R_L. For maximum rejection of HT noise and ripple, R_L should be as high as possible compared to r_a.

A pentode has $r_a > R_L$ and, therefore, has *no* rejection of HT noise.

Cathode feedback considerably increases r_a, but does not reduce total gain by a proportionate amount and, therefore, destroys HT rejection. In our (bypassed) example, r_a for the lower valve is equal to 6 kΩ and the active load ≈2 MΩ, resulting in 50 dB rejection of HT noise, but removing the bypass would raise r_a

for the lower valve to 47 kΩ and reduce HT rejection to 33 dB, despite leaving gain relatively unchanged.

Strictly, we should include the loss of the cathode follower in any calculation of gain to the low resistance output ($A_{total} = \mu \times A_{cathode\ follower}$), giving a gain of 31.5 in this instance.

The Importance of the AC Loadline

Up until now we have tacitly assumed that the input resistance of the following stage had little, or no, effect on the performance of the preceding stage. This would not be true if we used the anode output of the μ-follower because the value of the following grid-leak (typically ≈1 MΩ) is not merely comparable with the value of R_L, it is actually *less* than R_L and, therefore, lowers the effective value of R_L from ≈2 MΩ to 670 kΩ. This has a negligible effect on gain, but trebles the distortion, so using the anode output is not recommended.

Whilst $R_g \geq 10\ R_L$, it is legitimate to ignore its effect on the preceding stage, but once it becomes smaller than this, we should consider drawing an AC loadline to investigate whether it will cause a problem. Stages with active loads *must* take into account the input resistance of the following stage.

An accurate AC loadline is easily drawn. First we find the AC load, which is usually just the anode load and the following grid-leak in parallel. We know that the AC loadline must pass through the operating point, so all we need is a second point. The simplest way to do this is to move a convenient number of squares horizontally (change the voltage by 100 V or so), and calculate the increase, or decrease, in current through the AC load to give our second point. The line through these points is then the AC loadline, so inspection of this line gives the gain and linearity of the stage *including* the effects of the following load resistor.

Upper Valve Choice in the μ-Follower

There is no reason why the upper valve should be the same as the lower valve in a μ-follower. As a rule of thumb, the AC load resistance seen by the lower valve can be found by:

$$r_L \simeq \mu_{upper} \cdot R_L$$

Maximising r_L minimises distortion in the lower valve, but experiment shows that once $r_L \geq 50\ r_a$, there is no further benefit to be gained and it is more profitable to investigate the distortion produced by the upper valve. Because a cathode follower operates with 100% feedback, increasing μ increases feedback and reduces distortion. However, high- μ valves need a higher value of V_a to avoid grid current, reducing available V_a for the lower valve and lowering maximum voltage swing.

High g_m is also useful in the upper valve if the stage is to feed a passive equalisation network because the resultant low (but changeable) r_{out} is a smaller proportion of the network's series resistance.

The 6545P single triode has μ=52 and g_m≈20 mA/V at sensible anode currents, but its main advantage as the upper valve is that it can swing V_{gk} close to 0 V without distortion, allowing a high output voltage swing from a given HT voltage.

When triode strapped (g_2, g_3 to *a*), the D3a pentode is also a good choice as μ=80, and g_m≈20 mA/V is easily achievable even at quite low currents, but significant grid current begins at V_{gk}≈−1.1 V. The D3a has gold-plated pins and was produced in the era when gold plating genuinely meant special quality. Not only does it meet its published specification, but it is very consistent from one sample to another. By contrast, the (Soviet era) 6545P generally only just meets the lower limits of its specification and is rather variable, although its anode curves are extremely linear.

Limitations of the µ-Follower

Although the μ-follower is an excellent gain stage, it does have limitations. We have seen that it has a low output resistance and low distortion, and it is, therefore, tempting to use it as a line stage to drive long cables or low input resistance transistor amplifiers.

However, a low-impedance load steepens the AC load line of the cathode follower upper valve. Although this valve has 100% feedback, the steeper load line slightly reduces gain and the cathode follower can no longer bootstrap the lower stage's R_L as effectively, so the lower valve sees a reduced load resistance, and distortion in the lower valve rises; connection of an external load always increases distortion within a μ-follower. As an extreme example, the r_{out} of a 6J5/6J5 μ-follower was tested at 0 dBu. At +28 dBu, the stage produced 0.29% THD+N, so it was predicted to produce ≈0.01% THD at 0 dBu. However, an output resistance measurement that dropped the output from 0 dBu to −6 dBu by loading it with a 720 Ω resistance increased THD+N to 0.85%.

If a low-load resistance must be driven with minimum distortion, the µ-follower can be buffered by a cathode follower. To drive the load effectively, the cathode follower should pass ≥10 mA, and the valve should be a frame-grid type with high g_m and high μ; 6545P and triode-strapped D3a are ideal. The cathode follower is a high-impedance load, so it can be direct coupled to the lower output of the μ-follower to avoid the inevitable distortion of the upper valve in the μ-follower. For lowest possible distortion, the cathode follower should have a CCS as its load.

A well-designed μ-follower enters overload very suddenly. A 6J5/6J5 μ-follower driven from a 51 kΩ source was driven into grid current, resulting in an output of +38.1 dBu (61.6 V_{RMS}) at 0.87% distortion. A high source resistance causes hard clipping at the onset of grid current, so we should expect a decaying series of odd harmonics. However, because grid current only clips one half-cycle and asymmetry causes even harmonics, we can expect all possible harmonics (see Figure 2.41).

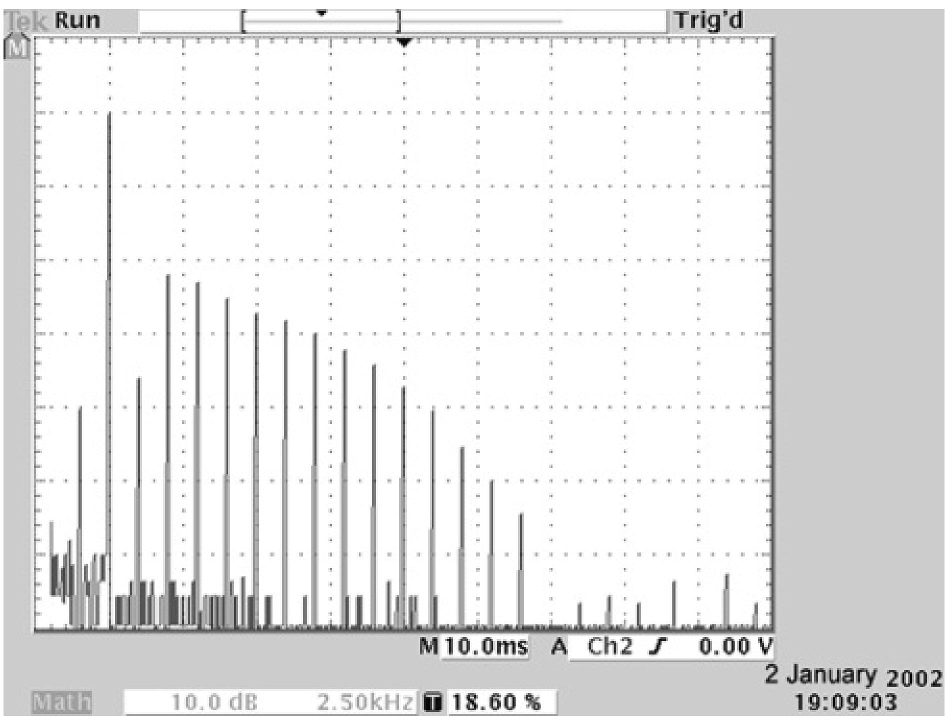


***Figure 2.41** Distortion spectrum of 6J5/6J5 μ-follower entering grid current.

Dropping the level by 1 dB to +37.1 dBu reduced the distortion to 0.54%, and the higher harmonics entirely disappeared (see Figure 2.42).

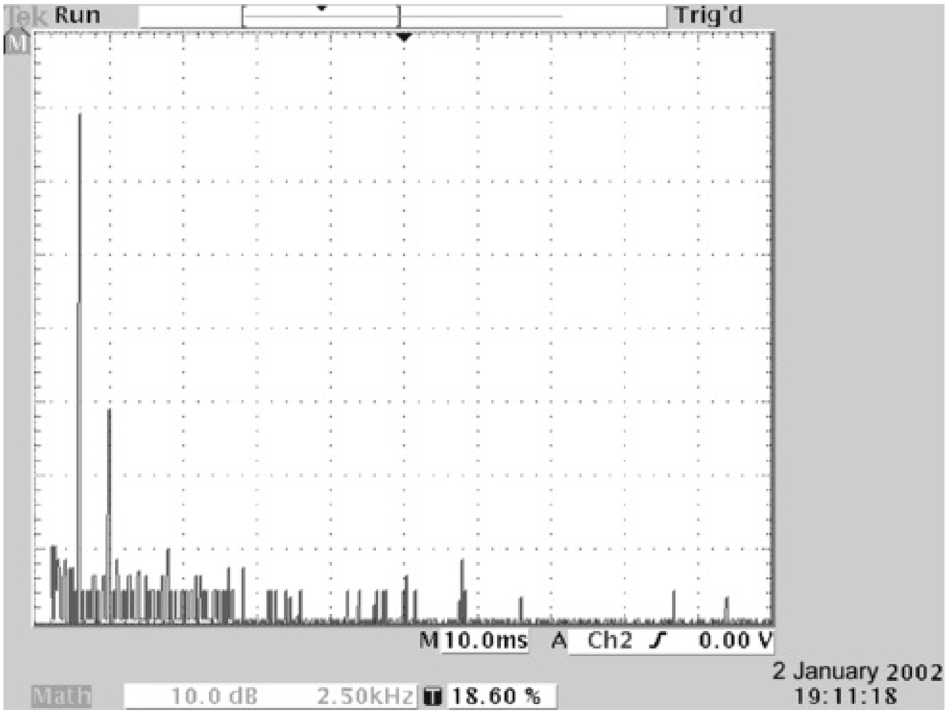


***Figure 2.42** Distortion spectrum of 6J5/6J5 μ-follower 1 dB below grid current.

The μ-follower is a gain stage and cathode follower connected in such a way that it provides a CCS load to the gain stage, making it much more linear. The downside is that the cathode follower is in series with the gain stage and consumes excess HT voltage because the (normally triode) cathode follower cannot swing to V_a=0, and also because of the voltage dropped across R_L. If we don't combine the two valves in this way, but instead use a semiconductor CCS load for the gain stage and another for the (DC-coupled) cathode follower, we can obtain the same distortion as the μ-follower but a higher maximum output swing for a given HT voltage. We will investigate semiconductor CCSs at the end of this chapter.

The Shunt-Regulated Push–Pull Amplifier (SRPP)

Patented by RCA [13], the SRPP was developed in the early 1950s to be used as a power amplifier or modulator in television transmitters, where it was typically required to drive 1,100 V_{pk-pk} into a load of 400 Ω in parallel with 500 pF with low distortion [14]. Far more distortion can be tolerated in video than in audio, and the standards of video at the time were comparatively poor, so 'low distortion' meant ≈2% and 'negligible distortion' meant <1%.

Although we are unlikely to use the SRPP in its original application, it is useful to understand the problems that the transmitter engineers faced and how they were solved. Developing 1,100V $_{pk–pk}$ across 400 Ω wasn't really a problem – it just needed a large valve, but maintaining that voltage across the 500 pF capacitance was. The highest frequency produced by the 405 line 'high definition' system was 3 MHz, and at this frequency, $X_C \approx 100\ \Omega$, requiring considerably more current than the 400 Ω resistance. The obvious solution was to increase the standing current in the stage, but this would have been wasteful of electricity because full amplitude High Frequency signals are very rare in real pictures (as opposed to test signals). What was needed was a means of sensing when the extra current was required, and then allowing a second valve to furnish that current. A resistor in series with the output of the lower valve senses the load current, so the voltage across this is used to drive the regulator (upper) valve. Because the regulator valve could typically quadruple the total signal power of the stage without requiring any additional standing current, this stratagem allowed a considerable increase in efficiency – a very important consideration in amplifiers dissipating kilowatts of heat.

The same type of valve is invariably used for both upper and lower sections (see Figure 2.43).

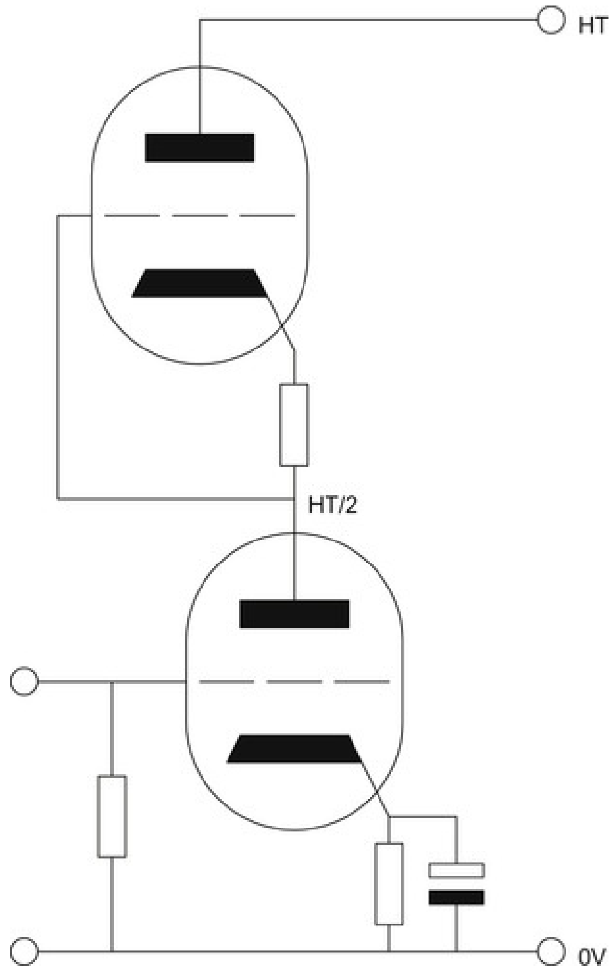


Figure 2.43 Shunt-Regulated Push–Pull (SRPP) amplifier.

The same current passes through both valves, so the associated cathode bias resistor R_k is the same. From a DC point of view, the section above the lower anode is identical to that below it, so each section sees half the HT voltage. If we draw a vertical line on the anode characteristics at 285 V/2=142.5 V and choose our anode current, this determines the required bias. The −4 V curve crosses 142.5 V at 4.5 mA, so 4 V/4.5 mA=889 Ω, and a 910 Ω resistor is fine (see Figure 2.44).

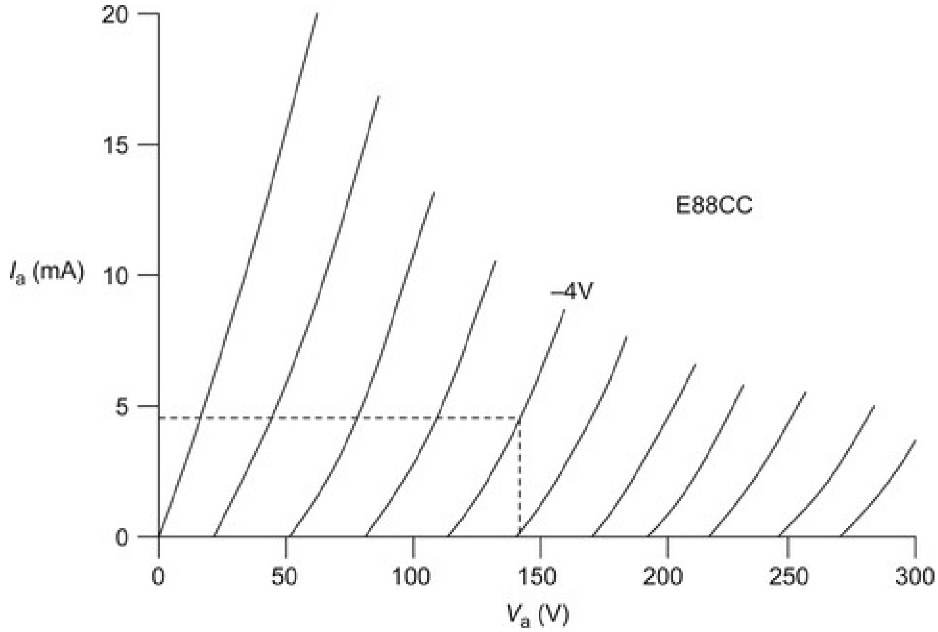


Figure 2.44 Choosing SRPP operating current.

Conceivably, differing valves or differing DC conditions could be used for upper (V_2) and lower (V_1) valves, in which case the full equations derived by Amos and Birkinshaw [15] give the gain of the stage as:

$$A_v = \frac{\mu_1(\mu_2 R_k + r_{a_2})}{r_{a_1} + r_{a_2} + R_k(\mu_2 + 1)}$$

And the output resistance may be found from:

$$r_{out} = \frac{r_{a_2}(R_k + r_{a_1})}{r_{a_1} + r_{a_2} + R_k(\mu_2 + 1)}$$

The SRPP is intermediate between the common cathode amplifier with a resistor as R_L, and the μ-follower with an active R_L, but the low value of upper cathode resistance R_k means that the value of R_L seen by the lower valve is inevitably quite low, implying that the SRPP must have $A_v < \mu$ and significantly increased distortion compared to a μ-follower.

A pair of typical 6J5GTs whose characteristics had previously been measured was set up as an SRPP:

μ=21

g_m=2.95 mA/V

r_a=7.11 kΩ.

The equations predicted A_v=14.3 and r_{out}=2.3 kΩ. Measurement found A_v=13.5

and r_{out}=2.3 kΩ. The SRPP was compared with the μ-follower by swapping the same valves between stages, and testing with identical DC conditions for all valves. Unsurprisingly, given its heritage, not only did the SRPP deliver a significantly higher output voltage swing than the μ-follower, but also the μ-follower required a higher HT voltage because it wastes HT across its additional 10 kΩ R_k (see Figure 2.45).

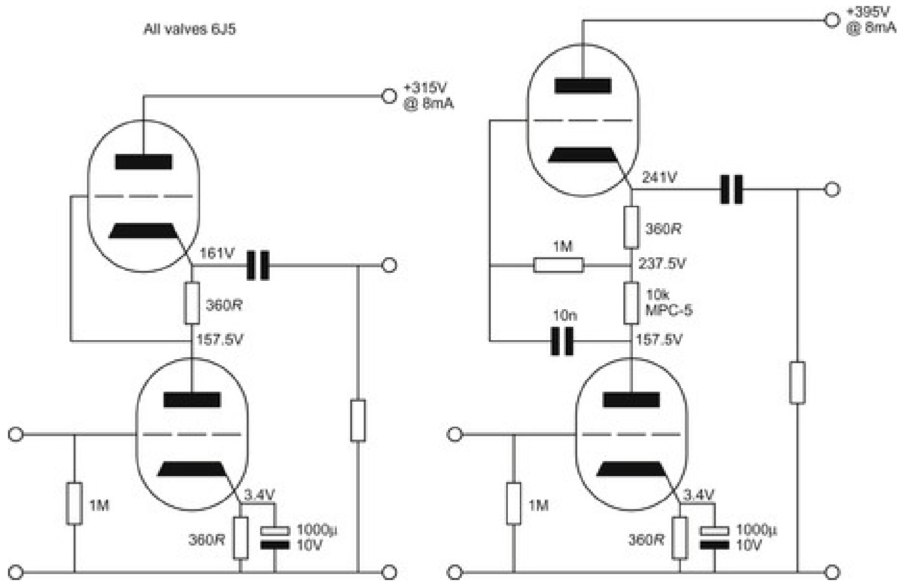


Figure 2.45 Circuits used for comparison of SRPP and μ-follower.

At an output of +28 dBu (19.5 V_{RMS}), the μ-follower produced 0.24% THD+N, but the SRPP produced 1.32%, an increase of 15 dB. As predicted, the SRPP produces significant distortion, and although this falls with level, it is still rather high for use in a pre-amplifier gain stage. The effect of HT voltage on distortion at +28 dBu was also investigated (see Figure 2.46).

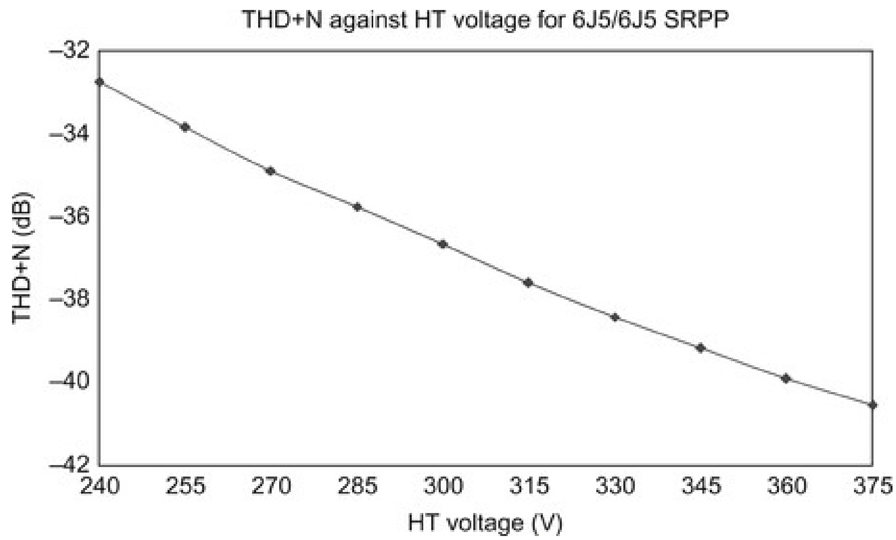


Figure 2.46 THD versus HT voltage for 6J5/6J5 SRPP at +28 dBu.

Although the SRPP seems a poor choice compared to the μ-follower, it does have the advantage that it is DC coupled internally (the μ-follower needs a coupling capacitor to the upper valve), and it is, therefore, immune to blocking (see Chapter 3).

The β-Follower

The β-follower [16] seeks to exploit the advantages of the μ-follower with the efficiency and DC coupling of the SRPP stage (see Figure 2.47).

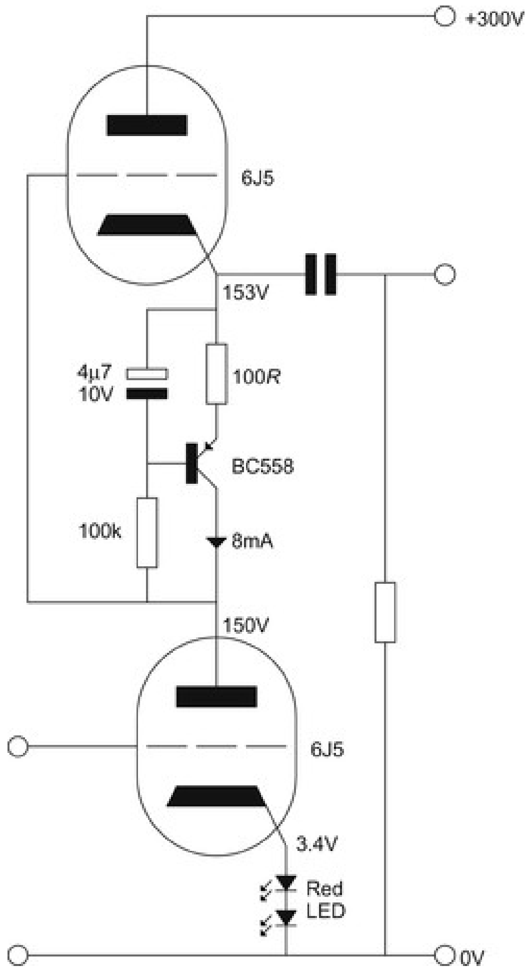


Figure 2.47 *β*-Follower.

Replacing the cathode bias resistor with a bipolar transistor allows the large (perhaps 10 kΩ) R_L to be discarded, reducing wastage of HT, and allowing the two valves to be DC coupled.

Bipolar transistors are usually treated as if their output characteristics are constant current, which implies horizontal output curves, but real transistors have curves that slope slightly (see Figure 2.48).

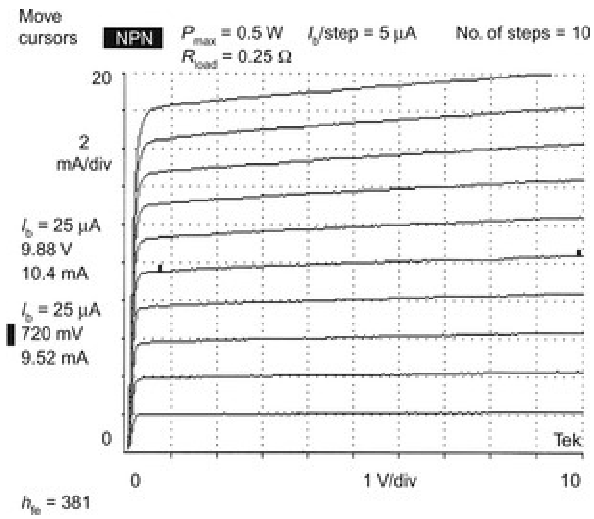


Figure 2.48 I_C versus V_{CE} for BC549 NPN transistor.

Looking into the anode, a valve multiplies R_k by μ; similarly, a bipolar transistor multiplies any resistance in the emitter circuit by β, or h_{fe}, so the curves can be flattened by adding an emitter resistor. Since h_{fe} for a small-signal transistor is likely to be ≈400, a 100 Ω resistor in the emitter makes the output resistance $1/h_{oe}\approx 40$ kΩ. The cathode follower then multiplies this resistance by its μ, perhaps 20, to give $R_L\approx 8$ MΩ, which is even better than a μ-follower can achieve.

The β-follower easily achieves $r_L \geq 50\ r_a$, even with a low- μ upper valve, so the upper valve must be chosen for minimum distortion, otherwise it will compromise the excellent performance of the lower valve.

The β-follower is an excellent test bed for determining irreducible distortion. If the lower valve is fed from $r_s\approx 0$, and loaded by $r_L\approx\infty$, then the remaining distortion is due to errors in valve geometry, such as uneven grid winding. The 6J5/6J5 β-follower shown in Figure 2.47 gave distortion performance that challenged the author's test equipment, with only the second harmonic being reliably measurable at −55 dB below the fundamental at an output level of +28 dBu – all other harmonics were better than −100 dB!

In theory, we could replace the bipolar transistor and its associated components by a depletion-mode junction field-effect transistor (JFET). If the gate is connected directly to the source, a typical 2SK147 becomes a ≈9 mA constant current source. However, r_d (the output resistance looking into the drain) is typically <10 kΩ, so it is not as effective at reducing distortion as the β-follower (see Figure 2.49).

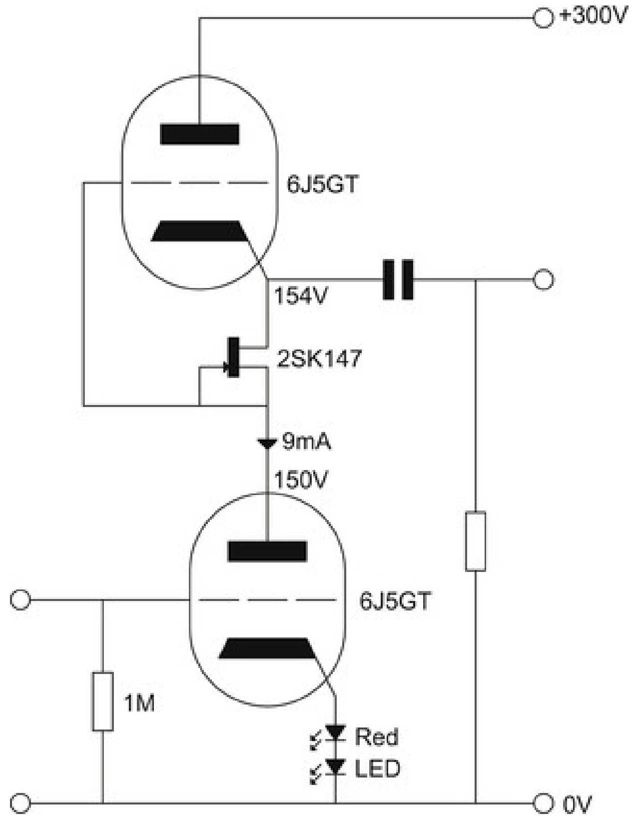


Figure 2.49 μ-Follower with 2SK147 JFET constant current source.

The Cathode-Coupled Amplifier

The cathode-coupled amplifier was a popular oscilloscope amplifier because directly coupling a cathode follower to a grounded grid stage produced a non-inverting amplifier having wide bandwidth; the circuit has recently become popular in OTL headphone amplifiers (see Figure 2.50).

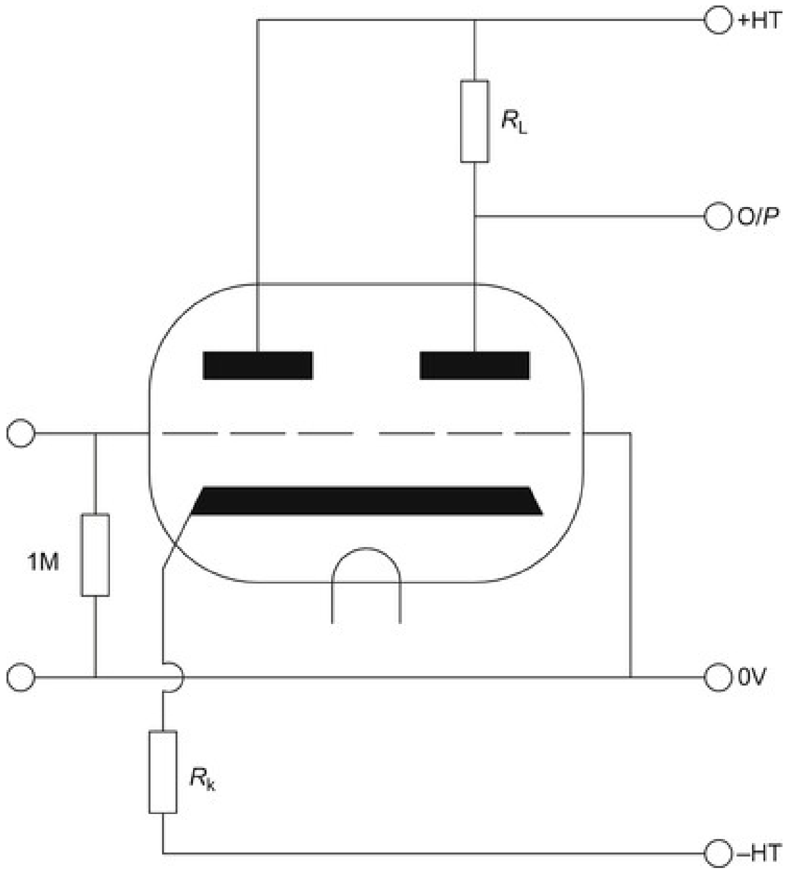


Figure 2.50 The cathode-coupled amplifier.

The amplifier can be considered to be a cathode follower (non-inverting) coupled to grounded grid amplifier (also non-inverting).
The cathode follower has 100% negative feedback, and its gain before feedback is:

$$A_0 = \frac{\mu R_k}{R_k + r_a}$$

Feedback reduces gain:

$$A = \frac{A_0}{1 + \beta A_0}$$

Combining the two to give the gain after feedback:

$$A = \frac{\mu R_k}{R_k(\mu + 1) + r_a}$$

Looking into the cathode of the grounded grid stage, we see:

$$r_k = \frac{R_L + r_a}{\mu + 1}$$

But the cathode follower sees this r_k in parallel with its load resistance R_k, so its gain becomes:

$$A = \frac{\mu R_k(R_L + r_a)}{R_k(\mu + 1)(R_L + 2r_a) + r_a(R_L + r_a)}$$

The total gain of the amplifier is the product of grounded grid gain and cathode follower gain:

$$A = \frac{\mu^2 R_L}{(\mu + 1)(R_L + 2r_a) + r_a(R_L + r_a/R_k)}$$

As Tektronix [17] pointed out, once R_L approaches r_a (as would be likely in an oscilloscope) and μ is large, cathode follower gain simplifies to ½, implying a total gain roughly half that of a common cathode stage. Otherwise, provided that μ is reasonably large, the R_k term becomes negligible, and the gain simplifies to:

$$A \geq \frac{\mu R_L}{R_L + 2r_a}$$

which is slightly lower than the gain we would expect from a conventional common cathode stage, but non-inverting.

The grounded grid section intrinsically has wide bandwidth because the control grid screens the cathode from the anode, and the cathode follower section has wide bandwidth because it does not suffer Miller capacitance from a changing anode voltage. Thus, the combined amplifier has wide bandwidth – which is why it was so popular as an oscilloscope amplifier.

However, there is a price to be paid for this bandwidth. Unless R_L is very large, the load resistance seen by the cathode follower is only a few kΩ, which greatly increases its second harmonic distortion (H2). It is fortunate that the signal level at this point is likely to be low, because this load resistance causes very little gain before feedback and, therefore, very little feedback is available to reduce distortion. The increased H2 would not have been a problem in an oscilloscope because oscilloscope display tubes necessitated push–pull operation, which cancels H2.

The Differential Pair

All of the circuits that we have so far studied have been *single-ended*, which is to say that they have only one output. (The μ-follower-type circuits were single-ended because although they had two outputs, they were of the same polarity.)

By contrast, the *differential pair* has two inputs and amplifies the *difference* between them to provide two outputs, one inverted with respect to the other; this

makes the differential or *long-tailed* pair [18] a very useful stage.
A differential pair can be made using the basic common cathode triode amplifier or with cascodes. (The μ-follower is not suitable because differential pairs attempt to exploit the normally large ratio between R_L and R_k.) For simplicity, we will analyse the differential pair using the basic common cathode triode amplifier (see Figure 2.51).

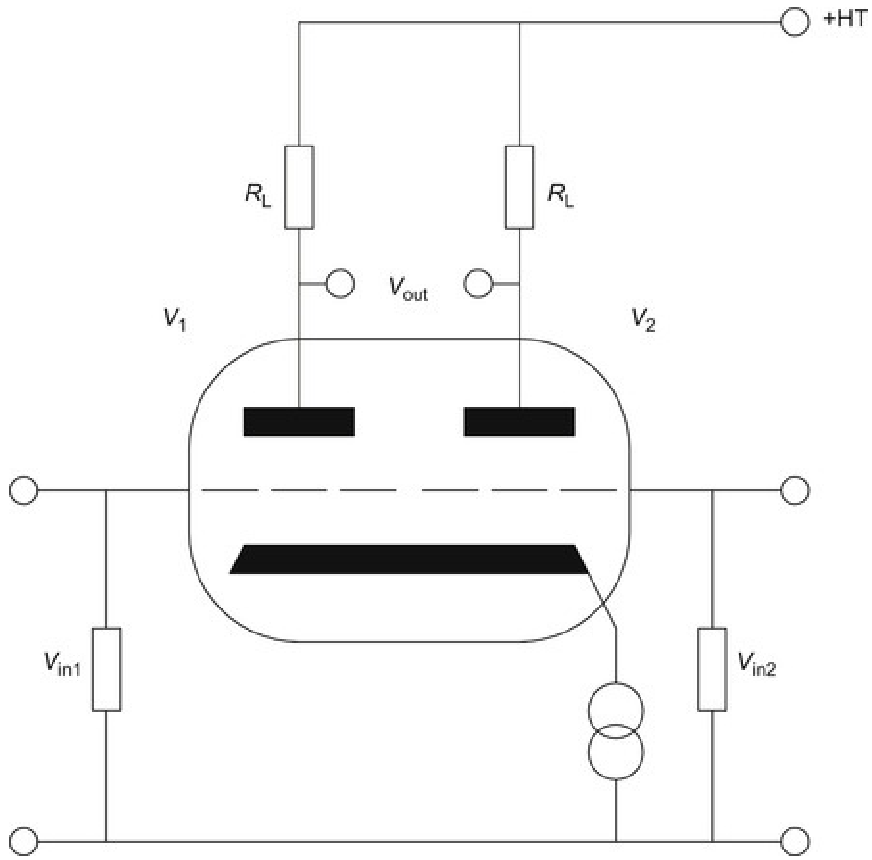


Figure 2.51 The differential or long-tailed pair.

The circuit consists of two identical triodes, often in the same envelope, with their cathodes tied together, passing anode current to ground via a CCS and each driving equal value anode load resistors.
Suppose that we apply an input signal such that the voltage on the anode of V_1 rises by 1 V. The current through V_1 must, therefore, have fallen, but since both valves are sitting on a CCS, this can only occur if the current through V_2 has *risen* by an *equal* amount. Since the anode load resistors are equal, it follows that the voltage on the anode of V_2 must have fallen by 1 V.
The outputs of the two anodes are equal in voltage, but one is inverted with

respect to the other.
Returning to the inputs: if we short circuit g_{V_2} to ground, and apply a sine wave to g_{V_1}, then the cathode will ‘follow’ that signal because, ignoring the anode loads, the circuit is a cathode follower. This means that V_2 is driven by its *cathode*, an amplified signal appears on its anode, and, therefore, an equal and opposite signal appears on the anode of V_1. The argument works in the same manner for a signal applied to g_{V_2}.

Gain of the Differential Pair

When driven by a signal connected between the two grids, the gain of the differential pair is identical to that of a standard common cathode stage, but the output voltage is found *between* the anodes of the stage. Therefore, if we look between one anode and ground, we only see half of the output voltage, and the gain appears to be halved.
If we use the differential pair as a phase splitter, and apply the same input voltage as before between one grid and ground, instead of each grid seeing half the input voltage, one sees the entire input voltage and the other none. Because the voltage difference between the two grids is the same, the gain remains the same.

Output Resistance of the Differential Pair

Provided that the output of the differential pair is not unbalanced *in any way*, r_{out} at each terminal is identical to that of a simple common cathode amplifier ($r_a \| R_L$).
However, if only one output is loaded, the output resistance rises considerably. Working backwards from the path to ground (HT supply) via the first R_L, we see:

$$r_k = \frac{R_L + r_a}{\mu + 1}$$

But we now also see a path R_k to ground (0 V), which is in parallel with r_k:

$$r'_k = \frac{R_k \cdot ((R_L + r_a)/(\mu + 1))}{R_k + ((R_L + r_a)/(\mu + 1))}$$

Multiplying through by (μ+1):

$$r'_k = \frac{R_k(R_L + r_a)}{R_k(\mu + 1) + R_L + r_a}$$

Looking down through the second anode, we see r_a in series with this multiplied by a factor of (μ+1):

$$r'_a = r_a + \frac{R_k(\mu+1)(R_L+r_a)}{R_k(\mu+1)+(R_L+r_a)}$$

If we divide by R_k(μ+1), we obtain:

$$r'_a = r_a + \frac{R_L+r_a}{1+((R_L+r_a)/R_k(\mu+1))}$$

As R_k tends to ∞, the right-hand term on the bottom line reduces to zero, resulting in a maximum value of r'_a:

$$\boxed{r'_a \approx R_L + 2r_a}$$

This high value of r_a will become significant when we investigate the PSRR of the differential pair.
If R_L>> r_a, then the output resistance (only one terminal loaded) is:

$$r_{out} \approx \frac{R_L(R_L+2r_a)}{2(R_L+r_a)} \approx \frac{R_L}{2}$$

AC Balance of the Differential Pair and Signal at the Cathode Junction

---

Provided that the anode load resistors are perfectly matched, the only way that there can be an AC imbalance at the anodes is if some signal current is lost to ground by a finite tail resistance. Thus, an infinite tail resistance would allow two entirely different valves to achieve perfect balance provided their anode loads were matched.
If the two valves are perfectly matched, then they must have identical gain to their identical anode loads and there can be no signal at the cathode junction (grids driven by signals of equal and opposite polarity). If they are not perfectly matched, their gains must differ, so there must be a signal at the cathode junction.
If the two valves are not driven by signals of equal and opposite polarity, there must be a signal at the cathode junction. This statement is easily understood by considering the extreme case of a phase splitter (signal at only one grid); the only way the valve with no signal on its grid can have its V_{gk} modulated is by

moving its cathode, requiring a signal at the cathode junction.
The two previous statements show that any imperfection of drive or valve matching results in a signal at the cathode junction – there should ideally be no signal. Worse, if there *is* a signal, it implies that the valves are being forced to amplify that signal in order to produce their output signal, and any interference signal at the cathode junction will also be amplified.

Common-Mode Rejection Ratio (CMRR)

Rather than assuming signals of equal and opposite polarity on each grid, we will now consider the response to identical signals on each grid. If we apply +1 V to *both* grids, the cathode voltage simply rises by 1 V, the cathode current remains constant, and the anode voltages do not change, because we have not modulated V_{gk}. The amplifier only responds to differences between the inputs or *differential* signals. Applying the same signal to both grids is known as a *common-mode* signal.
This property of rejecting common-mode signals is significant, since it implies that the circuit can reject hum on power supplies, or common-mode hum on the input signal, so we will investigate it further.
The signal at each output can be expressed in terms of currents using Ohm's law:

$$V_{\mathrm{out}(1)} = i_1 \cdot R_{\mathrm{L}(1)}$$
$$V_{\mathrm{out}(2)} = i_2 \cdot R_{\mathrm{L}(2)}$$

Each output will be an exact inverted replica of the other *if* $i_1 = i_2$, provided that the two load resistors are equal. There are two main ways in which this nirvana may be eroded:

- If signal current is lost through an additional path to ground. A signal current i_1 flowing down V_1 out of its cathode must split, with some current being lost down R_k and the remainder flowing up into the cathode of V_2, to become i_2. However, if $R_k = \infty$, no current can flow down R_k, and $i_1 = i_2$. If $\mu_1 = \mu_2$, and $R_L(1) = R_L(2)$, then:

$$\boxed{\mathrm{CMRR} \approx \frac{\mu R_{\mathrm{k}}}{R_{\mathrm{L}} + r_{\mathrm{a}}}}$$

This indicates that we should use high- μ valves, and maximise the *ratio* of R_k to R_L. As an example, an EF184 constant current source ($r_{sink} = R_k \approx 1$ MΩ) and E88CC differential pair (μ=32) and R_L=47 kΩ would have CMRR≈57 dB.

- Results from the above equation will be degraded if $\mu_1 \neq \mu_2$, or if $R_{L(1)} \neq R_{L(2)}$, or a combination of the two. With the easy availability of low-cost, accurate digital multimeters, mismatching of load resistors is avoidable, but matching the valves is harder. If $\mu_1 \neq \mu_2$, then:

$$\text{CMRR} \propto \frac{\mu_1 \cdot \mu_2}{\mu_1 - \mu_2}$$

which indicates that high- μ valves are still desirable, but that matching is important.

Because the simple equation for CMRR ignores mismatched valves, mismatched load resistors and stray capacitances, any predictions of CMRR>60 dB should be treated with caution. Nevertheless, it is handy for checking that your tail resistance R_k is sufficiently high to ensure that predicted CMRR>40 dB, since 40 dB is an easily achievable CMRR in practice.

Power Supply Rejection Ratio (PSRR)

Since hum and noise on the power supply line are a common-mode signal, they must also be attenuated by the CMRR. We might also expect the potential divider formed by r_a and R_L to give significant additional attenuation. However, at either terminal, the only path to ground is via the other anode and R_L up to the HT supply, which is exactly the scenario we investigated when determining r_{out} with only one output loaded, therefore:

$$r'_a \approx R_L + 2r_a$$

And the attenuation of power supply noise (solely due to potential divider action) is thus:

$$\text{Attenuation} = \frac{R_L + 2r_a}{2(R_L + r_a)} \leq 6 \text{ dB}$$

If $R_L \gg r_a$, we achieve the maximum attenuation of 6 dB! Our previous example (R_L=47 kΩ and r_a=4.95 kΩ) attenuates by 5.2 dB, and together with the 57 dB due to CMRR, PSRR=62 dB.

It is now well worth comparing the PSRR of the common cathode stage, μ-

follower and differential pair (same DC conditions for the amplifying valve) (Table 2.2).

Table 2.2 Comparison of PSRR for Different Topologies

Stage	PSRR (dB)
Common cathode (R_L=47 kΩ)	20
μ-Follower (r_L≈742 kΩ)	44
Differential pair (r_{sink}≈1 MΩ)	62

The differential pair is the best, and will remain the best, since an improved constant current source for the μ-follower could be adapted to become an improved CCS for the differential pair.

Knowing the PSRR enables us to design power supplies correctly because it gives an indication of the allowable hum on the HT supply.

For example, the second (differential pair) stage of a balanced pre-amplifier might need the 100 Hz power supply hum to be 100 dB quieter than the maximum expected audio signal. At this point, the signal has not received RIAA 3,180 µs/318 µs correction, so the level at 100 Hz is 13 dB lower than at 1 kHz. However, peak levels from LP are +12 dB compared to the 5 cm/s line-up level, so the maximum audio signal at 100 Hz is 1 dB lower than the 1 kHz calculated signal level at the anode (2.2 V_{RMS})=2 V. We want 100 dB signal/hum, but because 62 dB of this will be provided by the differential pair's PSRR, we only need the hum on the power supply to be 38 dB quieter than 2 V, so we could tolerate 25 mV of hum on the power supply – which is easily achievable.

Semiconductor Constant Current Sinks

The differential pair demonstrated the need for CCSs, but the pentode CCS is profligate with HT voltage (although it is a good sink), and a differential pair with grids at ground potential would require a subsidiary negative supply for the sink of −100 V. This is often undesirable, so a solution is needed.

Unlike the original valve designers, we are in the fortunate position of being able to use transistors, and even op-amps, if we consider them to be necessary. This is a perfect example of where a transistor or two can be very helpful.

The simplest form of a transistor CCS is very similar to our triode version. The red LED sets a constant potential of ≈1.7 V on the base of the transistor. V_{be} is ≈0.7 V, so the emitter resistor has 1 V held across it – irrespective of its resistance. We can, therefore, use Ohm's law to determine the resistance required to programme a particular current:

$$R_{\text{programming}} + \frac{1\ \text{V}}{I_{\text{required}}}$$

Thus, if we need to sink 5 mA, we would need a 200 Ω programming resistor. The AC resistance looking into the collector is:

$$r_{\text{out}} = R_{\text{E}} \cdot h_{\text{fe}} + \frac{1}{h_{\text{oe}}}$$

In this instance, a BC549C (h_{FE} guaranteed>420 and $1/h_{\text{oe}} \approx 12$ kΩ) gives $r_{\text{out}} \approx 96$ kΩ. Note that an expensive 2 W resistor is required to bias the LED (see Figure 2.52a).

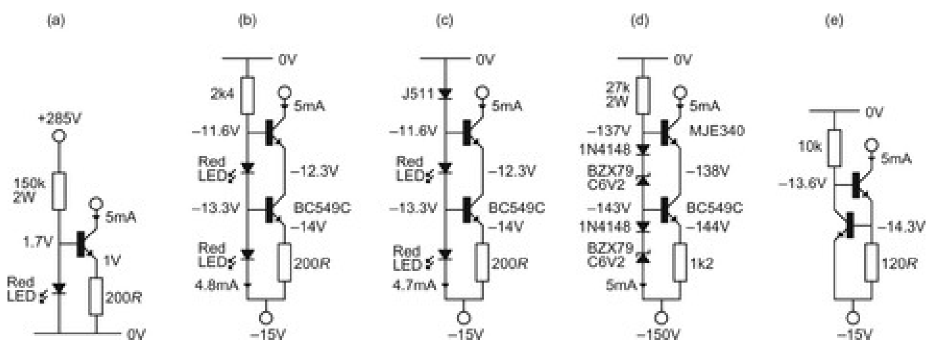


Figure 2.52 Semiconductor constant current sources.

The simple circuit can easily be improved upon, and since silicon is cheap, it seems worthwhile to do so. There are two problems to be addressed. Firstly, the transistor needs $V_{\text{CE}}>0.5$ V for it to operate as a CCS, which is uncomfortably close to typical bias voltages for high- μ valves such as the ECC83. Secondly, 92 kΩ output resistance is not especially high, and we can do much better.

A transistor cascode is broadly similar to a pentode, but a practical circuit requires a negative supply, although this may not be a problem in a power amplifier, because there is often a negative bias supply for the output valves that we can use. (Even though the bias winding normally supplies <1 mA, wire rated for 1 mA is very fragile, so transformer manufacturers typically use thicker wire, allowing us to draw 10 mA from this winding, and the increase in transformer VA loading is usually negligible.)

A cascode CCS has much higher output resistance than a single transistor CCS:

$$r_{\text{out}} = R_{\text{E}} \cdot h_{\text{fe(upper)}} \cdot h_{\text{fe(lower)}} + \frac{1}{h_{\text{oe(upper)}}}$$

The AC output resistance of the initial design has been multiplied by the h_{fe} of the second transistor (assumed to be 400), which improves it from ≈96 kΩ to

≈40 MΩ, and the value of $1/h_{oe}$ (which we probably don't know accurately) is now negligible.

The qualities we want from a CCS are high output resistance and low output capacitance – because capacitance to ground reduces output impedance at high frequencies. All transistors have internal capacitance between their collector and base C_{cb}, and the higher the power and/or voltage rating of a transistor, the higher C_{cb} tends to be. Maximising performance at high frequencies means choosing the most fragile transistor we dare in order to minimise C_{cb}.

However, it is possible to accidentally make C_{cb} higher than expected (see Figure 2.53).

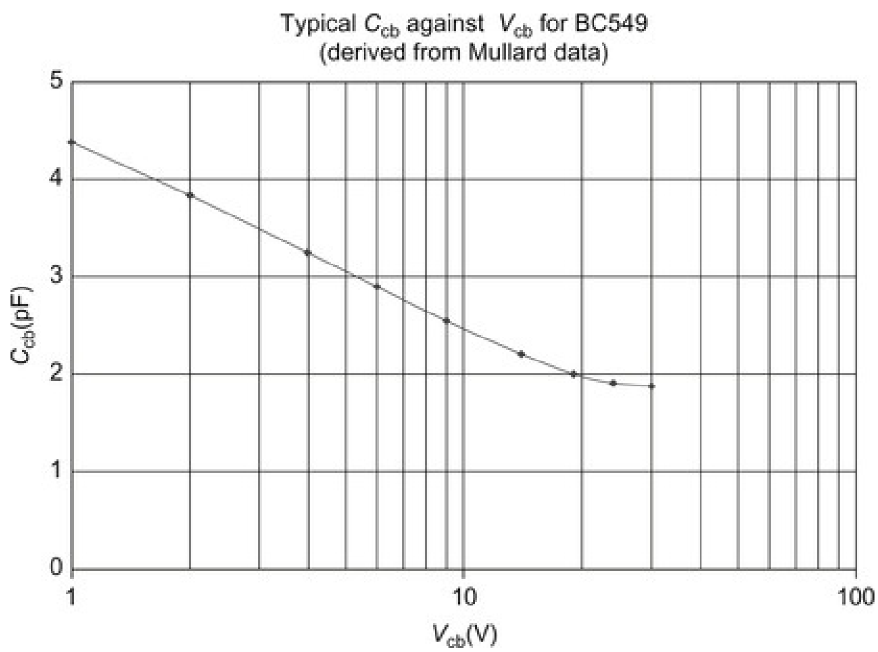


Figure 2.53 Collector-to-base capacitance (C_{cb}) for BC549.

C_{cb} is the capacitance across the depletion region between base and collector, and its thickness is proportional to $\sqrt{V_{cb}}$, so if there's a choice between having 1 V and having 10 V across the transistor, choose 10 V – this is another justification for a negative supply. However, a more obvious advantage is that the negative supply allows the output port to be taken down to 0 V without linearity problems. High Frequency stability of the cascode CCS is excellent (see Figure 2.52b).

As shown, the cascode current sink is relatively sensitive to hum and noise on the negative supply because of current changes through the voltage reference. The cheapest and best solution is to regulate the negative supply, perhaps using a

337 three-terminal regulator, but if there isn't enough negative voltage available for the 4 V drop needed across the regulator, a second-best solution is to insert a constant current diode into the chain that feeds the voltage reference (see Figure 2.52c).

The cascode CCS is easily adapted to withstand a larger voltage by replacing the outer transistor (the one connected to the load) with a higher-voltage type, but this generates two new problems:

- High-voltage transistors inevitably have a low h_{fe}, and because r_{out} is the product of the two current gains and the programming resistor, reduced h_{fe} translates directly into reduced r_{out}. Don't over-specify the device – if you only need to withstand 100 V and dissipate 100 mW, then a 2N5551 ($h_{fe(min)}$=80) doubles the output resistance compared to an MJE340 ($h_{fe(min)}$=30).
- If the outer transistor has a significant voltage across it, it probably also dissipates significant power. As an example, 120 V×10 mA=1.2 W – which is too much for a small transistor, forcing an MJE340 and a heatsink. Elsewhere, the author has vilified small heatsinks on transistors, preferring to use the chassis as a heatsink, but this is the one instance where it is *essential* to use an individual heatsink. The collector of almost all power transistors is connected to the case, so we place an insulating washer between it and the heatsink. For an MJE340, the transistor, insulating washer and heatsink form a capacitance of ≈8 pF, and one end of this capacitor is connected to the output terminal of our CCS. If we connect the heatsink to ground, we connect 8 pF in parallel with the output of our CCS and reduce output impedance at high frequencies. *Solution*: Use an individual heatsink (oriented to lose heat efficiently), do not connect it to ground, and minimise its capacitance to ground.

Because we have voltage to spare, some of the lost output resistance can be recovered by setting a higher reference voltage, allowing a higher value of R_E. The 1N4148 diode compensates for variation of the lower transistor's V_{be} due to temperature, but should not be added to LED references because the temperature coefficients of V_{be} and V_{LED} are quite similar, so they are already temperature compensated [19] (see Figure 2.52d).

It is essential to realise that it is *only* the outer transistor (the one connected to the load) that has to withstand any external high voltage. Not only is it unnecessary to use a high-voltage device for the inner transistor, but it would significantly degrade performance. High-voltage transistors always have low h_{fe} − 40 is typical for the MJE340 – so a cascode CCS required to withstand +100 V using a pair of MJE340s would have an output resistance one-tenth that of exactly the same design using a BC549C as the inner transistor, simply because the BC549C has 10 times the h_{fe} of an MJE340. *Remember*: choose the outer transistor to survive external conditions and the inner transistor to optimise performance.

The Williams [20]‘ring-of-two’ circuit works by holding 0.7 V across the 120 Ω sense resistor. If that voltage rises, due to increased current through the resistor, T_1 turns on harder, which causes the base voltage of T_2 to fall. T_2 begins to turn off, so the current through the 120 Ω resistor and, therefore, the sink current, is held constant (see Figure 2.52e).

Beware that because the ring-of-two relies on feedback applied over two transistors, there is a possibility of oscillation at high frequencies due to stray capacitances. By comparison, the cascode CCS is far less likely to become unstable.

Using Transistors as Active Loads for Valves

All the previous sink circuits can be mirrored about 0 V and PNP transistors substituted for NPN. If the circuit is then connected to the HT supply, all the previous sink circuits become constant current sources, allowing a triode to achieve $A_v = \mu$. More significantly, they permit a valve to achieve low distortion from a low HT voltage.

As an example, in common with all high- μ valves, the ECC83 needs considerable V_a before it can be biassed out of grid current; 150 V is typical. As a general rule of thumb, $R_L > 2\ r_a$, and as $r_a \approx 75$ kΩ for the ECC83, we might use $R_L = 150$ kΩ. If $I_a = 0.7$ mA, we would drop 105 V across R_L, so we would need 255 V of HT. But we might only require the stage to produce an output swing of 5 $V_{pk–pk}$, so most of the HT is wasted. If we replace the 150 kΩ resistor with a constant current source, the valve sees a much higher value of R_L, and we can set the HT voltage independently to accommodate the maximum required output swing (see Figure 2.54).

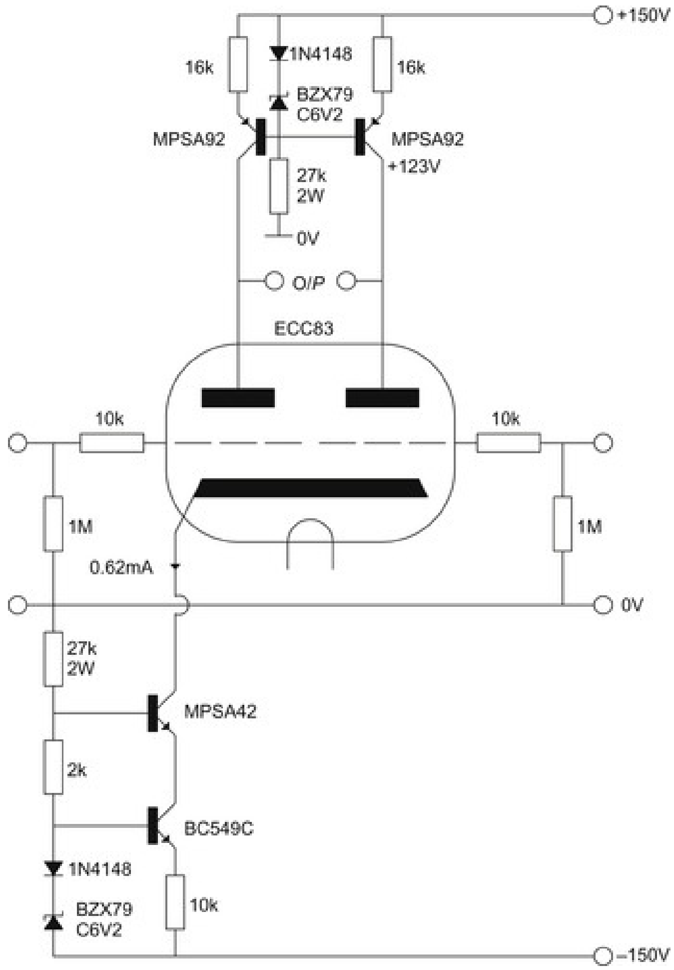


Figure 2.54 Semiconductor anode loads.

In Figure 2.54, the concept of operating a high- μ valve from a low HT was taken to the extreme because the author needed a high gain differential pair stage (ECC83: μ=100), but only had 150 V of positive HT available. Note that high-voltage transistors are required to withstand either anode swinging towards 0 V.

At fist sight, the circuit has two constant current sources, both trying to define the current in the same wire. The trick is that the cathode CCS is deliberately superior to the anode constant current sources, enabling it to enforce its behaviour upon them.

The anode constant current sources have an output resistance of ≈600 kΩ (h_{fe}=40 and R_E=15 k), but the cathode CCS has an output resistance of ≈130 MΩ (h_{fe1}=420, h_{fe2}=40 and R_E=7k5). Thus, fine adjustment of the cathode CCS simply changes the voltage drop across the 600 kΩ output resistance of the anode constant current sources. The output resistance of the anode constant current sources doesn't change appreciably with temperature, so the stability of the final circuit is down to the behaviour of the cathode CCS. The cathode CCS enforces 0.82 mA, ideally split equally between the anode CCSs so that each passes 0.41 mA. By Ohm's law, a 1 μA change in current through the 600 kΩ output resistance of an anode CCS would cause a change in voltage drop of 0.6 V. The LED reference voltage and ordinary resistor in the cathode CCS circuit should be stable to <1% drift in current, resulting in <2.5 V drift in each anode voltage, which is perfectly acceptable.

Although Zener diodes are normally bypassed to reduce noise, the noise generated by both diodes is a common mode and is, therefore, rejected by the next (differential) stage. On test, the circuit achieved the required differential swing of 7 $V_{pk–pk}$ at 1 kHz with only 0.04% distortion.

A cascode greatly increases r_{out}, flattening the loadline and reducing distortion in the valve. If we wanted to maximise output swing and minimise distortion, we might operate a 7N7 (loctal equivalent to 6SN7) at I_a=8 mA because μ becomes more nearly constant when I_a>6 mA. We assume that our cascode will provide a horizontal loadline, so we plot this at 8 mA (see Figure 2.55).

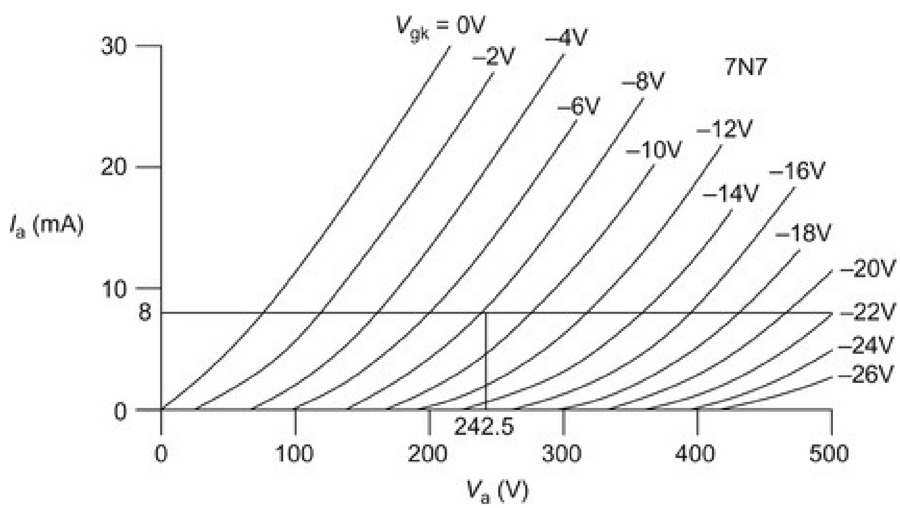


Figure 2.55 7N7 with constant current 8 mA loadline.

Normally, as V_a rises, we have to consider the cramping of curves as I_a falls and cut-off is approached, but I_a is now a constant, and the only limit to positive swing is that the cascode requires sufficient voltage to operate correctly. 15 V is quite adequate for the cascode, so a 400 V HT would allow V_a to swing to 385 V. Looking in the opposite direction along the 8 mA loadline, grid current is likely to begin at ≈100 V. The maximum possible swing is, therefore, 385−100 V=285 V $_{pk–pk}$≈100 V $_{RMS}$.

Although the cascode forces I_a=8 mA, we must adjust valve bias to set V_a. At just over maximum swing, we should clip positive and negative half-cycles equally, so the operating point should be halfway between the maximum and minimum permissible anode voltages – which is their average:

$$V_a = \frac{V_{max} + V_{min}}{2} = \frac{385 + 100}{2} = 242.5\ \text{V}$$

Looking at the curves, we see that $V_{gk} \approx 8$ V is required to set V_a correctly, and this can be provided by an 8.2 V Zener diode (see Figure 2.56).

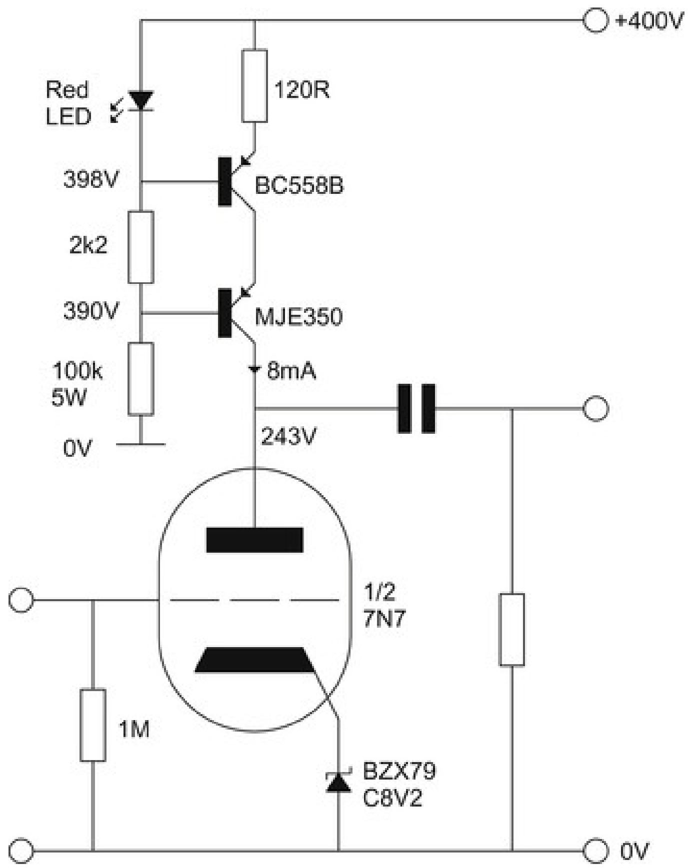


Figure 2.56 Cascoded semiconductor anode load.

Since the stage is intended to swing large voltages, noise is not a problem, so it is not essential to bypass the Zener diode with a capacitor.

If there is 242.5 V across the valve, then there is 147.5 V across the lower transistor, so it must dissipate 1.18 W when quiescent. When V_a swings to 100 V, the transistor must withstand 285 V at 8 mA, so it momentarily dissipates 2.28 W, and it might seem that this is the required rating of the transistor. However, the flat loadline has reduced distortion almost to zero, so the positive and negative swings are equal, and the average power dissipated in the transistor over one cycle of audio is equal to the quiescent power.

As before, it is only the transistor nearest the valve that must be able to withstand high voltages and dissipate significant power, so the additional transistor can be as fast and fragile as we like.

Optimising r_{out} by Choice of Transistor Type

Table 2.3 compares transistors that are useful in support circuitry for valves.

Table 2.3 Abbreviated Data for Bipolar Transistors Useful in Valve Support Circuitry

		$V_{CE(max)}$ (V)	$I_{C(max)}$ (mA)	P_{max}	f_T	$h_{FE(min)}$	$1/h_{oe(typ)}$ (kΩ)
BFR90	NPN	15	25	300 mW	5 GHz	40	5
BC549C	NPN	30	100	500 mW	300 MHz	420	12
BC558B	PNP				200 MHz	220	6
2N3904	NPN	40	200	500 mW	250 MHz	100	15
2N3906	PNP			625 mW			5
2N5551	NPN	160	600	625 mW	100 MHz	80	35
MPSA42	NPN	300	500	625 mW	50 MHz	40	50
MPSA92	PNP						35
MJE340	NPN	300	500	20 W	10 MHz	30	150
MJE350	PNP				4 MHz		50

$V_{CE(max)}$: The maximum allowable voltage between collector and emitter. (There are various ways of specifying this limit, so unless you know your precise circuit conditions, it is wise not to exceed 2/3 V_{CE}.)

$I_{C(max)}$: The maximum allowable collector current.

P_{max}: The maximum allowable power dissipation in the device ($P = I_C \times V_{CE}$).

$h_{FE(min)}$: Minimum DC current gain from base to collector. (The author's measurements suggest that h_{FE} is generally double the manufacturer's specified minimum value at the typical currents required by valves.)

f_T: AC current gain h_{fe} falls with frequency. At f_T, h_{fe}=1; this is known as the transition frequency.

$1/h_{oe(typ)}$: This is the typical AC resistance (equivalent to r_a) seen looking into the collector. It is very rarely specified by manufacturers, so these figures were measured on a curve tracer at V_{CE}=10 V and I_C=10 mA to allow comparison between the types. PNP transistors tend to have a lower Early voltage than their NPN counterparts, so $1/h_{oe}$ is lower *and* falls faster at higher currents.

Output resistance at low frequencies is partly determined by $1/h_{oe}$, but is dominated by h_{fe}, since any resistance in the emitter circuit is multiplied by h_{fe}. Impedance at high frequencies is shunted by the capacitance seen at the collector of the transistor, which will partly be determined by strays, but also by the transistor itself. In general, high-voltage/high-power transistors have a larger silicon die area, and greater capacitance, which is reflected in their lower f_T. Additionally, f_T varies significantly with I_C, and operating a transistor below its optimum I_C could reduce f_T by a factor of five. If in doubt, download the transistor's datasheet from the Internet – all the semiconductor manufacturers have excellent websites. As a consequence of these considerations, a small cascode CCS would ideally use two BC549Cs, or if low output capacitance (≈0.5 pF excluding strays) was essential and sufficient voltage was available, a triple cascode composed of a BFR90 with two BC549Cs beneath it. It is only critical for the outer transistor to have low output capacitance because the capacitances of the other transistors are at lower impedance points.

Any bipolar transistor needs a minimum V_{CE} for it to operate linearly. For a low-voltage transistor at currents ≤30 mA, ≈1 V is sufficient, but higher currents may require 2 V (see Figure 2.57).

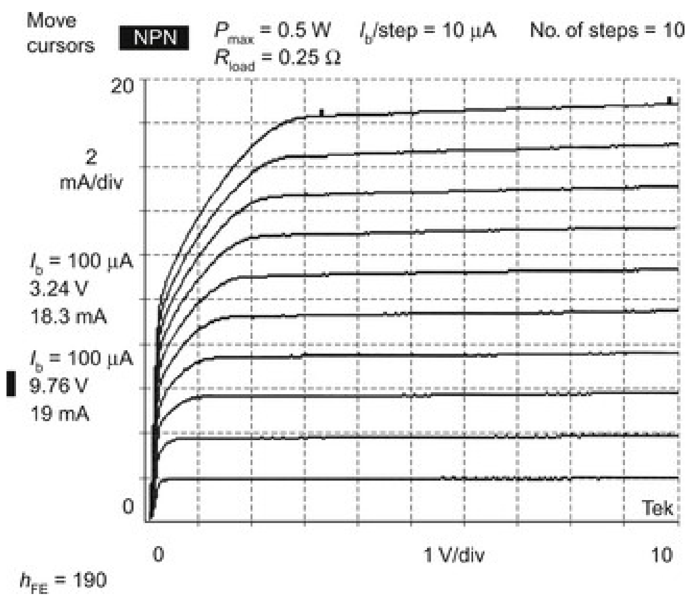


Figure 2.57 I_C versus V_{CE} for 2N3904 transistor showing minimum V_{CE} required.

High-voltage transistors such as MPSA42 or MJE340 may require V_{CE}>2 V. A CCS in a differential pair operating as a phase splitter has half the input signal across it, so this point can become significant. In a cascode sink, the lower transistor has hardly any AC signal across it, so it can operate with only 2–3 V_{DC}, leaving the remaining DC for the upper transistor, which supports the bulk of the AC signal.

Field-Effect Transistors (FETs) as Constant Current Sinks

FETs are described as being depletion or enhancement mode. An enhancement mode device needs bias similar to a bipolar transistor to turn it *on*, whereas a depletion mode device needs bias similar to a valve to turn it *off*. Both can be used for making CCSs, but a depletion mode device can be self-biassed, whereas the enhancement mode device requires an external voltage, making it more cumbersome to use. ‘Constant current’ diodes are actually depletion mode FETs with their gate and source tied together. They tend to have a low maximum voltage, and disappointingly low slope resistance for their price.

As stated earlier, the output capacitance of any constant current circuit must be minimised if it is to maintain its performance at high frequencies, and that means choosing transistors with low output capacitance. Sadly, many high-voltage FETs simply aren't suitable for valve electronics because of excessive output capacitance. As an example, the 700 mW power rating of the ZVN0545A suggests that it ought to be useful but it actually has higher output capacitance than the 15 W (TO220) DN2540N5. The reason for this is that semiconductor manufacturers make far fewer different devices than you think – they just package the dies differently. Thus, the DN2540N3 is the same silicon die, but in a TO92 package, reducing its power rating to 1 W.

Happily, the 400 V 15 W Supertex DN2540N5 depletion mode JFET is very nearly ideal for supporting valve audio. Unfortunately, the phrase 'device variation' was invented for FETs and never more so than for the voltage $V_{GS(ID=0)}$ required to turn the device on (or off); Supertex specify a spread of $V_{GS(ID=0)}$ from −1.5 V to −3.5 V. Fortunately, it's easy to measure V_{GS} for an individual device under any operating conditions, so devices can be matched or other circuit values can be determined for a specific device (see Figure 2.58).

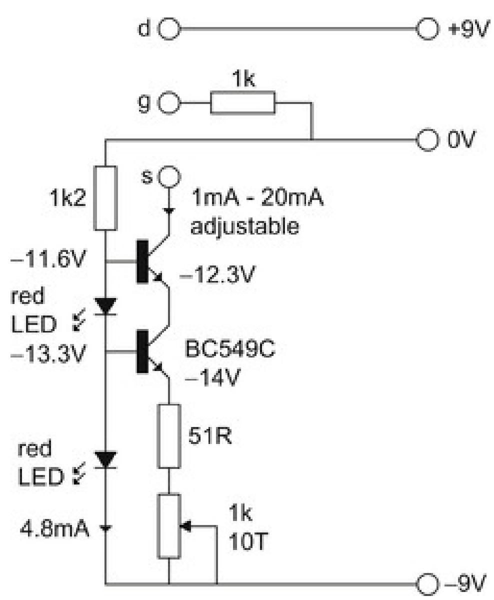


Figure 2.58 This jig allows easy determination of V_{GS} for any FET conditions.

The drain of the device is connected to a positive supply having the same voltage expected in the final circuit. The gate is connected to ground via a 1 kΩ gate-stopper resistor to prevent RF oscillation. The source is connected to a convenient negative supply via a programmable CCS adjusted to the required current. A Digital Volt Meter (DVM) connected between ground and source measures V_{GS}. If we had more than one device, we would quickly run them all through the test jig to find each individual V_{GS}. Devices with similar V_{GS} would be deemed to be pairs.

Designing Constant Current Sinks Using the DN2540N5

The very simplest CCS using the DN2540N5 is little more than the contents of the constant current diode vilified earlier (see Figure 2.59).

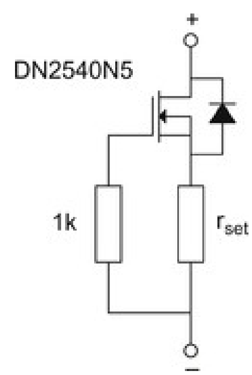


Figure 2.59 The simplest JFET CCS.

The carbon gate-stopper resistor prevents Very High Frequency (VHF) oscillation, and the source resistor programmes the current. Unfortunately, the value of the source resistor has to be found empirically for each DN2540N5, using the test jig of Figure 2.58. Alternatively, a variable resistor could be used and Adjusted On Test (AOT), but we still need to know the approximate value required in order to buy the correct value of variable resistor and if accidentally set to 0 Ω at switch-on, the saturation current might cause damage. It really is easier (and cheaper) to use the jig and fit a fixed resistor.

As an example, we might want to use a DN2540N5 as a CCS shared by the cathodes of a pair of EL84s. We know that when the EL84s are working, their cathodes will be at ≈11 V and will pass a total of 80 mA. The programmable CCS probably needs a minimum of 6 V across it to operate correctly, so we connect it to a negative supply of perhaps −9 V. We then apply +11 V to the drain of our Device Under Test (DUT), and using an ammeter adjust the programmable CCS to draw 80 mA through the DN2540N5. Having set the required current, we measure the voltage between 0 V and the FET's source. (We don't attempt to genuinely measure V_{GS} on the FET because connecting a DVM to it will almost certainly cause oscillation, so we assume zero gate current and measure from 0 V.) Perhaps, we find that for our particular sample, V_{GS}=−1.873 V at I_{ds}=80 mA. All we need to do to turn the JFET into a 80 mA CCS for our EL84 is to add a source resistor that will drop 1.873 V when 80 mA passes through it (just like a triode), 1.873 V/0.08 A=23.4 Ω. The nearest standard value of 24 Ω will almost certainly do.

Like the pentode CCSs we investigated earlier, the FET multiplies the value of its programming resistor by μ. Unsurprisingly, μ is not given by the manufacturer, but can be found from the relationship:

$$\mu = g_m \cdot r_{ds}$$

The drain slope resistance r_{ds} can be found from the device curves by measuring output currents on one V_{GS} curve at two widely spaced voltages (see Figure 2.60):

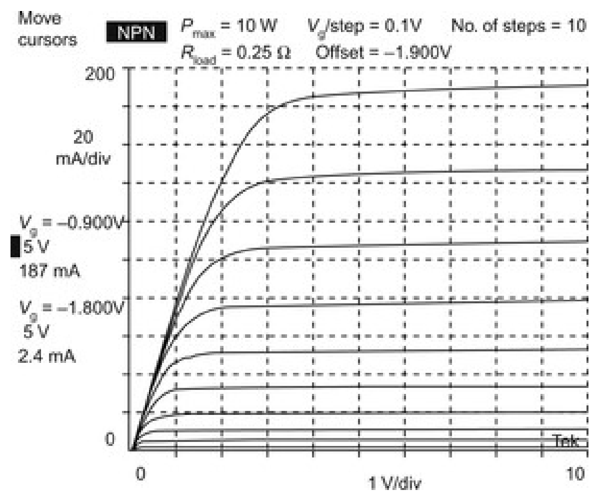


Figure 2.60 Determining FET r_{ds} is just like determining pentode r_a.

$$r_{ds} = \frac{V_1 - V_2}{I_1 - I_2} = \frac{100 - 20}{32.6 - 30} \geq 31 \text{ k}\Omega$$

The mutual conductance g_m at V_{ds}=100 V can also be found from the device curves:

$$g_m = \frac{I_1 - I_2}{V_{GS_1} - V_{GS_2}} = \frac{33.2 - 18.4}{-1.375 \text{ V} - (-1.475 \text{ V})} = 148 \text{ mA/V}$$

Note that although this value of g_m is much lower than given on the datasheet, it is far higher than a valve could achieve at the same current. Combining the two values, we obtain μ≈4,500, which is a little higher than we might expect from a pentode. More significantly, a 39 Ω programming resistor in this FET would produce a ≈40 mA CCS having an output resistance of 175 kΩ. To put this into context, a genuine 175 kΩ resistor would drop 7 kV rather than 100 V and dissipate 282 W rather than 4 W.

The previous worked example was given not to suggest that such a process is needed, but to show just how good a CCS can be made with the DN2540N5.

Nevertheless, considerable improvement is possible, and like the simple bipolar transistor, the cascode is the way to go. Remembering how the single BJT CCS evolved into a cascode, we could add a voltage reference and second FET (see Figure 2.61a and b).

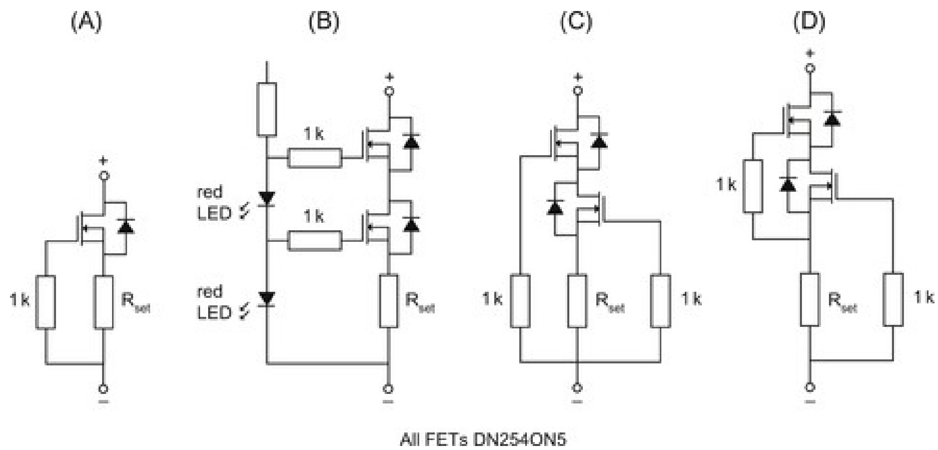


Figure 2.61 Evolution of the JFET cascode CCS.

Unfortunately, by adding the voltage reference, we lose the very desirable property of the circuit being a two-terminal device. Because the DN2540N5 is a depletion mode device, the voltage reference is not strictly necessary, and we could use one DN2540N5 as the reference resistance for the other (see Figure 2.61c).

However, there are disadvantages because the lower device in the cascode is forced to operate with a very low V_{ds}, but they can be ameliorated by returning the upper device's gate stopper not to 0 V but to the source of the lower device (see Figure 2.61d).

This simple circuit is deservedly popular because (being two-terminal) it can be used equally well as an anode load or as the cathode tail resistance in a differential pair. The programming resistance is multiplied by both amplification factors, and even though the lower device's performance suffers due to the low V_{ds}, our previous example is likely to improve from ≈175 kΩ to ≈30 MΩ.

However, things are not quite as rosy as they might seem. The high resistance applies at DC, but not at AC because of output capacitance C_{OSS}[21] (see Figure 2.62).

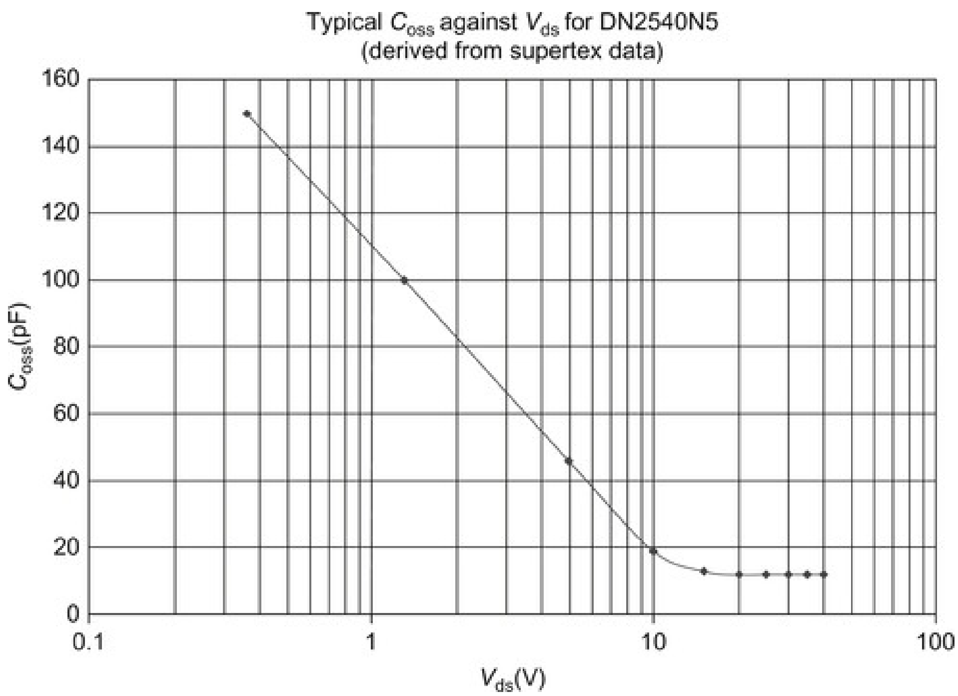


Figure 2.62 C_{OSS} output capacitance of the DN2540.

As can be seen from the graph, C_{OSS} is alarmingly high at low voltages and it is not until V_{ds}>15 V that it falls to its asymptotic value of ≈12 pF. Fortunately, we would always operate at V_{ds}>15 V (preferably 40 V, or more) because a BJT cascode is a better choice at lower voltages. An EF184 pentode CCS could achieve an output capacitance of 3 pF, but once we add typical strays of 3 pF to both circuits, the valve advantage degrades to 6 pF against 15 pF. This somewhat poorer output capacitance is the price we pay for the undoubted convenience of a two-terminal CCS.

The device has a maximum continuous current rating of 500 mA, so it should come as no surprise to learn that performance degrades at low currents. If you need a current <10 mA, then you owe it to yourself to see if there's an alternative solution because the device is much better >10 mA, and at 25 mA it really comes alive (r_{out} of a cascode CCS at 25 mA was four times that at 10 mA, all other factors kept constant).

References

[1] Mullard. Book 2 (part 1): receiving valves and television picture tubes. General operational recommendations. Mulland, Londen 1969:17.

[2] Gewartowski, JW; Watson, HA, *Principles of electron tubes*. (1965)Van Nostrand; pp. 112–21.

[3] Improved vacuum electric discharge apparatus having an incandescent cathode. Siemens & Halske Aktiengesellschaft. British Patent 145,421; 1921.

[4] Improvements in or relating to circuit arrangements and discharge tubes for amplifying electric oscillations N V Philips. Gloeilampenfabrieken. British Patent 287,958; 1928.

[5] Hunt, FV; Hickman, RW, On electronic voltage stabilizers, *Rev Sci Instrum* **10** (January) (1939) 6–21.

[6] Johnson WZ. Cascode amplifier. US Patent 4,647,872; 1987.

[7] Blumlein AD. Improvements in and relating to thermionic valve circuits. British Patent 448,421; 1934.

[8] Puckle, OS, *Time bases*. 1st ed. (1943)Chapman & Hall; p. 92; 2nd ed. Wiley, New York, 1951. p. 129.

[9] White E. Improvements in or relating to thermionic valve amplifier circuit arrangements. UK Patent 564,250; 1940.

[10] Amos, SW; Birkinshaw, DC, *Television engineering: principles and practice*. General circuit techniques. **Vol. 4** (1958)Iliffe; p. 264.

[11] Horton JW, J Franklin Inst. Vol. 216. December 1933.

[12] Kimmel, A, The mu stage, *Glass Audio* **5** (2) (1993) 12.

[13] Balanced Direct and Alternating Current Amplifier. US Patent 2,310,342. 9 February 1943 [filed November 1940].

[14] Cooper, VJ, Shunt-regulated amplifiers, *Wireless Engineer* **May** (1951) 132–145.

[15] Amos, SW; Birkinshaw, DC, *Television engineering: principles and practice*. General circuit techniques. **Vol. 4** (1958)Iliffe; p. 250.

[16] Ignacio, Vilá, The beta follower amplifier, *Glass Audio* **10** (4) (1998) 1.

[17] Typical oscilloscope circuitry. Rev ed. Tektronix. 1966. p. 7-1.

[18] Blumlein AD. Improvements in or relating to thermionic valve amplifying circuit arrangements. British Patent 482,740; 1936.

[19] Lefferts, PA, LED used a voltage reference provides self-compensating temp coefficient, *Electronic Design* **February** (1975) 92.

[20] Williams, P, Ring-of-two reference, *Wireless World* **July** (1967).

[21] DN2535, DN2540 data sheet. Supertex Inc; 2001.

Recommended Further Reading

- Yarwood, J, *An introduction to electronics*. (1950)Chapman & Hall; Quite a small book covering most electronic principles but majoring on the physical background behind them. Essential reading.
- Valley, GE; Wallman, H, *Vacuum tube amplifiers*. (1948)McGraw-Hill; Now reprinted by Audio Amateur Press (2000). A very general book ranging from audio through nuclear pulse amplifiers to radio and modulation. Has a good section on noise.
- Terman, FE, *Electronic and radio engineering*. 4th ed. (1955)McGraw-Hill; As implied by the title, majors on radio but contains plenty of general electronic principles.
- Reich, HJ, *Theory and applications of electron tubes*. (1939)McGraw-Hill; As implied by the title, fundamental electronics with audio and radio applications. An ideal counterpart to Terman.
- Langford-Smith, F, *Radio designers handbook*. 4th ed. (1953)Iliffe; Generally considered to be the bible of valve audio, especially as it contains considerable detail on output transformer design. However, many of the audio circuits are crude by modern standards, and the understanding of loudspeakers is flawed.
- Reich HJ. Principles of electron tubes; 1941. Now reprinted by Audio Amateur Press US. (1995). Readily available, but not outstanding.
- Ryder, JD, *Electronic fundamentals and applications*. 3rd ed (1964)Prentice-Hall; As should be expected from a book written at the transition from vacuum state to solid state electronics, considerable space is given to early transistors. However, the strength of this book is that a principle explained in one technology is then shown to be general.
- Amos, SW; Birkinshaw, DC, *Television engineering: principles and practice. Video-frequency amplification, Vol. 2*. (1956)Iliffe; Ignore the word ‘video’ in the title. Lucidly written as part of the BBC’s internal engineering training, this book analyses a number of circuits common to video that have been stolen by modern audio.
- Alexander, RC, *The inventor of stereo: the life and works of Alan Dower Blumlein*. (1999)Focal; Not an engineering tome, but a very readable biography of a genius killed at the age of 39 having generated 128 fundamental patents. Blumlein invented most of audio and video and (as if that was not enough), the differential pair and stereo.
- Burns, R, *The life and times of AD Blumlein*. (2000)The Institution of Electrical Engineers; The other biography of Blumlein, this time with more

of an engineering bias. Read the two as a pair.

Chapter 3. Dynamic Range

Distortion and Noise

In engineering terms, dynamic range is the ratio between the largest and the smallest signals, and it is primarily wide dynamic range (not bandwidth) that distinguishes quality audio from rubbish. Provided that the largest signal is not clipped, maximising dynamic range becomes a problem of minimising the three unwanted signals that otherwise mask the smallest wanted signal:

- Distortion
- Random noise
- Artificial interference, frequently referred to as ElectroMagnetic Compatibility (EMC).

EMC is largely a construction issue and will be considered in *Building Valve Amplifiers*, so this chapter will investigate how to minimise distortion and noise.

Distortion

To investigate and minimise distortion, we must begin by looking at the fundamentals of distortion measurement. Even designers of test equipment would probably concede that this is not the sexiest of topics, but unless we understand how errors can creep into our measurements, we will not have the confidence to compare the results of one measurement with those of the other, leaving us unable to test and improve our designs.

Defining Distortion

Although we may glibly use the word ‘distortion’ while talking about amplifiers, there are actually two distinct types of distortion.
Linear distortions do not change with amplitude. If we consider the transfer characteristic of a device producing linear distortion, it is a straight line – hence the term linear distortion (see Figure 3.1a).

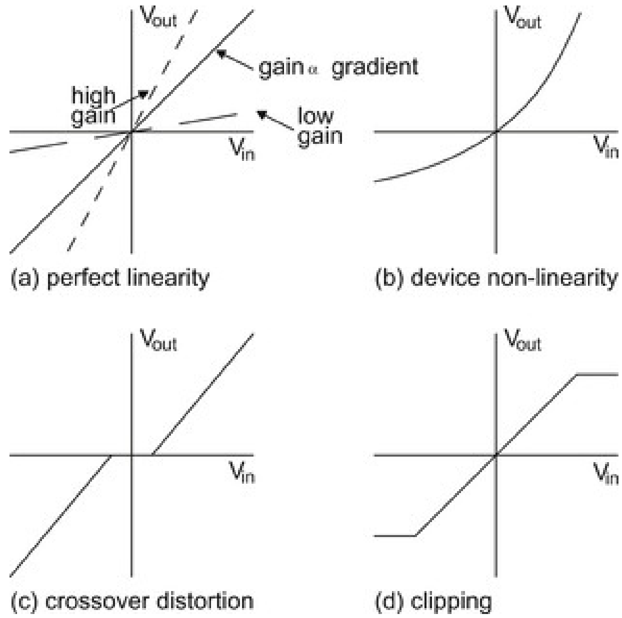


Figure 3.1 Transfer characteristics and the distortion they produce.

Although a device causing linear distortion changes the shape of the waveform, there are no *additional* frequencies at the output of the device. Linear distortion typically causes errors in the amplitude against frequency response – and this is the way that it is usually assessed. However, it is perfectly possible to change the shape of a waveform without changing the amplitude against frequency response by distorting the *time* at which frequencies arrive – loudspeaker crossover systems without delay compensation inevitably generate this distortion. The shape of a square wave's leading edge is particularly sensitive to timing errors, so an oscilloscope quickly reveals problems. Alternatively, timing errors between sine waves of differing frequencies can be determined by measuring the gradient of a plot of phase against frequency (linear scale for frequency). Deviations from the expected straight line imply phase errors – hence the term *linear phase* for an ideal device.

Unsurprisingly, the transfer characteristic for *non-linear* distortion is not a straight line, and a device causing non-linear distortion has frequencies at the output that were not present at the input (see Figure 3.1b–d).

Measuring Non-Linear Distortion

We can assess the linearity of a device in two fundamental ways:

- We plot the transfer characteristic directly. Since the definition of non-linear distortion was that the transfer characteristic should deviate from a straight line, we could measure the deviations. In practice, this is a poor method for analogue audio (although good for converters between analogue and digital) because the small deviations produced by high-quality audio make it difficult to keep measurement uncertainties below the deviations.

• We look for frequencies at the output of the device that were not present at the input. This is a very sensitive and easily applied test, so there are two common variations on the theme.

The simplest expression of the second test is to apply a single sine wave to the device. At the output of the device, we expect to see a single sine wave. However, if the device produces non-linear distortion, there will also be harmonics of the original sine wave. The test is popular because removing the original sine wave at the output is easy, leaving only the harmonics – which can then be measured, either individually or collectively, as Total Harmonic Distortion (THD).

A more complex method is to apply *two* sine waves to the device. Again, we should only see these two frequencies at the output, but a device producing non-linear distortion causes the two frequencies to amplitude modulate one another, producing sum and difference frequencies known as intermodulation products. Intermodulation distortion measurement is popular with Radio Frequency (RF) engineers because it is easy to tune to each intermodulation product and measure its amplitude, but this is not quite so easily done at audio frequencies.

It is most important to realise that measuring harmonic distortion is no more ‘correct’ than measuring intermodulation distortion, or vice versa. Both forms of measurement simply reflect the same non-linearity in the device’s transfer characteristic. What *is* important is how the measurement is made and how the results are interpreted.

Distortion Measurement and Interpretation

In an ideal world, everybody would make their distortion measurement in the same way, with the same equipment, and interpret their results identically. All results would then be comparable, allowing us to state that device ‘A’ was better than device ‘B’.

In practice, there are many different measurement techniques. For example, the intermodulation distortion measurement requires two (or more) frequencies. Which frequencies should be chosen, and what should their relative amplitudes be? There are at least three standards for this measurement. Similarly, which frequency should we use for measuring harmonic distortion? Should we make the measurement at a single frequency, or should we sweep frequency through the entire audible range? Which results should we include, and which ones should we exclude? Standards attempt to answer these questions and allow results to be compared. Engineers love standards – that's why we have so many of them.

If we have designed a piece of equipment, we already know where its failings are likely to be, so we plan our test to expose those failings. This allows us to quantify the failings, make a change to the design and measure to see if we have made an improvement.

The previous paragraph strikes to the heart of the measurement problem, and raises various points:

- We need to be aware of the limitations of the test equipment. There is little point in attempting to measure the distortion of an amplifier suspected to produce <0.01% Total Harmonic Distortion+Noise (THD+N) if the test oscillator itself produces 0.01% THD+N.
- We must know the relevance of our measurements. Measuring wow and flutter on an analogue turntable is useful because this measurement exposes known mechanical failings. Conversely, early CD player specifications quoted pointless wow and flutter measurements (essentially zero for a digital source), but failed to measure jitter (an insidious problem in the conversion between analogue and digital domains).
- A designer, seeking to improve their design, makes the test critical. Conversely, the marketing department requests engineering tests that the device is known to pass comfortably – such as harmonic distortion at 10 dB below full output for a digital source – because these tests give such good figures.
- Hopefully, nobody understands a given design as well as the designer – who is best placed to decide which tests should be made.
- Measurements are of most use to the designer.

For these reasons, measurements quoted by manufacturers or reviewers are not necessarily particularly useful – and this is part of the reason for subjective reviews. (Another reason is that good test equipment is expensive.)

However, if we intend to design and build valve amplifiers, then carefully chosen measurements employing cheap test gear and careful interpretation can be very useful indeed.

Choosing the Measurement

Transistor amplifiers typically have plenty of global negative feedback to reduce distortion. Because applying feedback can easily turn an amplifier into an oscillator, the amplifier is deliberately made to have an amplitude response that falls with frequency before the feedback is applied. Since negative feedback reduces both linear and non-linear distortions, when it is applied the frequency response reverts to flatness and non-linear distortion is also reduced. However, because the amplifier's response was falling with frequency before the feedback was applied, less negative feedback is available at high frequencies to correct non-linear distortion. This means that high-feedback amplifiers *must* have THD that rises with frequency, so a single-frequency measurement is inappropriate, and a swept measurement is better.

If we test a circuit that does not have global negative feedback, then a single-frequency measurement may be appropriate – if we know what causes the distortion.

A valve distorts because of the curvature in its I_a versus V_{gk} transfer characteristic, and does so at all audio frequencies without fear or favour. Harmonic distortion is the easiest measurement, and we are at liberty to choose any test frequency that we feel is convenient. We might choose 50 Hz or 60 Hz because we have a DVM specified to be accurate to 0.1 dB at that frequency. If we do, we will find that we cannot even measure the fundamental amplitude accurately because stray hum picked up from nearby power wiring beats with our wanted signal to give a gently fluctuating measurement. We need to change our test frequency so that it is clear of mains hum and its harmonics.

Perhaps we could use 10 kHz. This is nicely clear of mains hum, but has problems of its own. Some non-linearities produce mainly higher harmonics, but if the amplifier's amplitude against frequency response was already falling, this would attenuate the very harmonics we were trying to measure, and give a falsely good result. We need a lower frequency.

In terms of octaves, 1 kHz is in the middle of the audio band, so it is least affected by errors caused by reduced bandwidth and is sufficiently far away from AC mains frequency for hum not to upset the result. Marketing people love 1 kHz because it measures so very well.

Refining Harmonic Distortion Measurement

Classical harmonic distortion measurements were made at 1 kHz by removing the 1 kHz fundamental and measuring the amplitude of the remaining signal. Although they were appropriate for the valve amplifiers of the time, these tests were rightly criticised when applied to transistor amplifiers because they took no account of the distribution of harmonics and their subjective annoyance.

Weighting of Harmonics

Various proposals have been made for weighting the levels of individual harmonics to allow harmonic powers to be summed to give a single-figure measure of subjective distortion.

Shorter [1] suggested in 1950 that levels should be weighted by a factor of $n^2/4$ (where n is the number of the harmonic):

$$\mathrm{dB} = 20\log\frac{V_1}{V_2} = 20\log\frac{n^2}{4} = 40\log\frac{n}{4}$$

From n to 2 n is one octave, so the gradient in dB/octave is:

$$\begin{aligned} &= 40\left(\log\frac{2n}{4} - \log\frac{n}{4}\right) \\ &= 40\log\left(\frac{2n/4}{n/4}\right) = 40\log 2 \\ &= 12\ \mathrm{dB/octave} \end{aligned}$$

Thus, rather than measuring amplitudes of individual harmonics and calculating THD, a rising response of 12 dB/octave could be applied to a conventional distortion meter. In order that the measurement should be comparable with a conventional meter measuring pure second harmonic distortion from a 1 kHz source, the weighting filter would need a −12 dB gain offset to ensure 0 dB gain at 2 kHz. Note that the combination of the weighting filter and its gain offset means that weighted distortion measurements are only valid for the single specified fundamental frequency.

However, there are problems with the $n^2/4$ weighting technique. Using the 1 kHz measurement example, 20 kHz is 10 times higher than 2 kHz, so the filter would add a minimum of 40 dB of gain to harmonics that are inaudible. Since the whole point of the exercise was that the measured result should agree with the subjective nuisance, a 20 kHz low-pass filter is also required.

Although the Shorter recommendation successfully ranked measured distortion against the subjective nuisance, the test suffered from the limitations of its time. The levels of deliberate distortion were quite high (0.41–3.7% RMS unweighted), and the loudspeaker was a ‘wide-range coaxial horn’ of unspecified distortion.

Peter Skirrow of Lindos Electronics argued that distortion should be measured at

1 kHz using a weighting filter conforming to CCIR468-2 because the response of this filter was determined by the subjective nuisance of different frequencies. Broadly speaking, CCIR468-2 rises with frequency at 6 dB/octave, has 0 dB gain at 1 kHz and peaks by 12 dB at 6.3 kHz, after which it falls swiftly (see Figure 3.2).

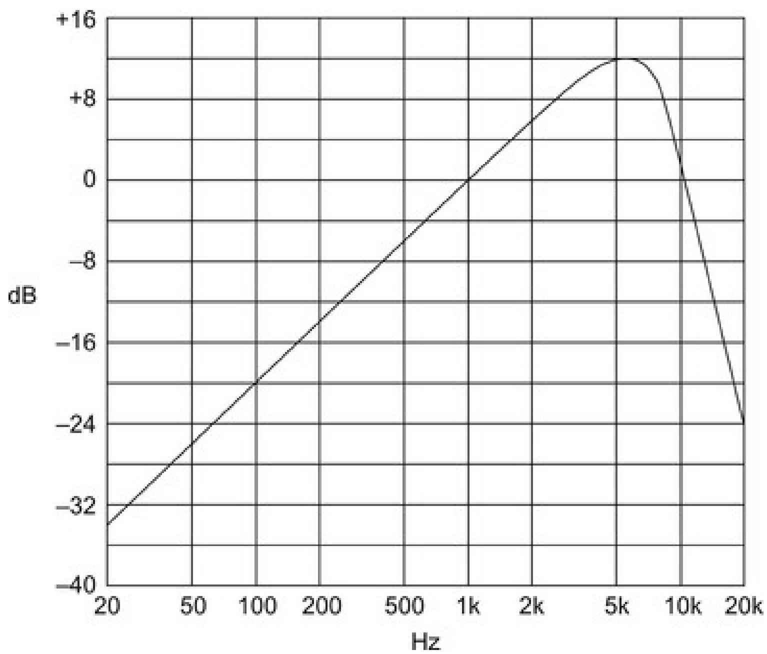


Figure 3.2 Frequency response of CCIR468-2 weighting filter.

Summation and Rectifiers

The waveform that remains after the fundamental has been removed is known as the *distortion residual* and is composed of a number of harmonically related frequencies. How should we measure the amplitude of this waveform? This is not nearly as easy a question as it first appears. Perhaps we could measure V_{pk-pk} (see Figure 3.3).

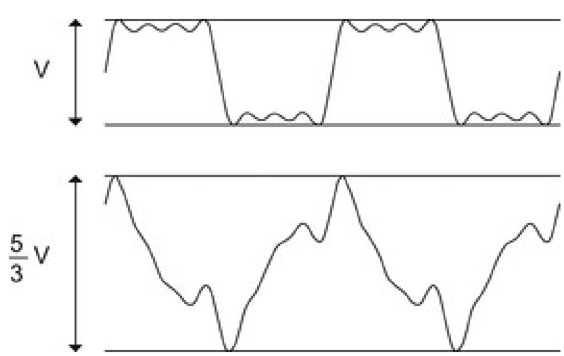


Figure 3.3 The effect of phase on waveform shape.

Both waveforms are square waves with correct amplitude harmonics up to the seventh harmonic, and none thereafter, but one has had the phase of the fundamental shifted by 90°, which significantly changes V_{pk-pk}. Mathematically, the correct way to sum individual harmonics is to turn the voltages into powers by squaring ($P = V^2 / R$), take the mean of the powers, and then convert this back into a voltage by taking the square root, otherwise known as an RMS measurement. Thus, classical distortion measurements measure the amplitude of the distortion residual using a meter incorporating a true RMS rectifier, and their measurements may reflect this fact by quoting THD in $\%_{RMS}$.

Alternative Rectifiers

CCIR468-2 specifies that the rectifier should be peak detecting, because noise is impulsive, and we want to capture the amplitude of these noise spikes. Crossover distortion produces narrow spikes that would contribute very little to an RMS summation, but are subjectively extremely annoying, so the peak-detecting rectifier of CCIR468-2 would seem ideal for detecting these spikes.

CCIR468-2 is not quite ideal because it needs the previously mentioned gain offset, so the CCIR/ARM recommendation offsets the gain of the CCIR468-2 weighting filter by 6 dB to give 0 dB gain at 2 kHz, allowing it to be used for 1 kHz distortion weighting. Unfortunately, it also changes the rectifier from peak detecting to average reading (ARM, Average Reading Meter), marking it less likely to detect the spikes produced by crossover distortion, so you might prefer to use CCIR468-2 and add the gain offset manually.

Noise and THD+N

Although using a CCIR468-2 filter to weight distortion is cheap and effective, its rising response with frequency may create another problem. Properly designed circuitry creates very little distortion. To put it another way, the distortion could be of comparable amplitude to the noise that all electronics generates. When we make our THD measurement, using our meter, how do we know that we are not actually measuring the amplitude of the noise?

The best solution is to view the distortion residual on a 20 MHz oscilloscope (or one with the 20 MHz filter engaged). If the waveform appears clean, we are measuring mostly distortion; if a repetitive waveform is difficult to discern, we are probably measuring noise. Thus, all practical measurements made by a meter are actually THD+N, and we have to be certain that the noise is sufficiently

small to be ignored. Digital oscilloscopes are excellent for making this decision objectively. The oscilloscope can be set to measure the RMS amplitude of the distortion residual, and this measurement can be compared with the same measurement but with averaging engaged. Averaging multiple waveforms cancels the noise (which is random), but maintains the repetitive distortion residual. If there is negligible (<10%) difference between the two measurements, we are measuring distortion. But if the averaged value drops to 71% of the unaveraged value, we are measuring equal amounts of noise and distortion residual, and it is time to stop recording distortion figures.
By definition, white noise has constant amplitude with frequency, whereas distortion harmonics occur at very specific frequencies. Our meter is a broadband device, which means that it is equally sensitive to all frequencies across the audio bandwidth. Thus, although the noise power in a particular frequency band could be quite low, and possibly significantly less than the amplitude of an adjacent distortion harmonic, when summed, the noise powers could easily swamp the distortion powers. This wouldn't be a problem if it were not for the fact that the ear/brain combination can pick distortion harmonics out of the broadband noise because it works like a *spectrum analyser*.

Spectrum Analysers

A spectrum analyser plots amplitude against frequency, allowing us to distinguish easily between noise and distortion harmonics. If we measure individual amplitudes of distortion harmonics, and then mathematically apply a subjective weighting such as Shorter's recommendation or CCIR468-2 to those numbers, we avoid noise problems.
Analogue audio spectrum analysers were traditionally expensive, but the digital alternative simply relies on raw computing power, and now that this is cheap, many digital oscilloscopes have facilities that convert them into spectrum analysers. Alternatively, a PC with a recording quality (24-bit 192 kS/s) soundcard can perform the entire function of distortion measurement and analysis at audio frequencies. All that is required is some analogue interfacing hardware (such as Pete Millett's 'Soundcard Interface/AC RMS Voltmeter' design) and some audio analysis software (such as AudioTester). Such a solution does not have quite the flexibility and ease of use of a dedicated audio test set, but it's a fraction of the price and the performance is excellent. Many modern audio test sets are simply outstandingly good soundcards having optimum analogue scaling and dedicated audio analysis software.
However, the process of analogue to digital conversion and its subsequent

analysis is *not* transparent, so we do need to understand its limitations.

Digital Concepts

An analogue signal is continuously variable both in voltage (or current, distance, etc.) and in time. Conversely, a digital signal can only change its parameter in discrete steps (*quanta*) and at fixed intervals. Taking measurements to plot a graph is a crude form of analogue to digital conversion, because we freeze the variation, make a numerical measurement, and then move on to make another measurement. The power of the technique is that reducing the measurements to a series of numbers allows us to analyse those numbers using a supremely powerful tool – *mathematics* – to find patterns.

Analogue to digital conversion is a two-part process. We freeze the parameter at fixed intervals, and we take numerical measurements of the parameter. These processes can be done in either order. We could take continuous measurements, but record only those measurements that occur at particular intervals. Alternatively, we can first freeze the parameter at fixed intervals, and then make the numerical measurement. It does not matter which way round these two quite distinct processes are applied.

Sampling

The process of freezing the parameter at regular intervals in time is known as *sampling*. If we take 192,000 samples in a second, then the *sample rate* is 192 kS/s; alternatively we can quote the *sampling frequency* as 192 kHz. The sampling frequency is significant because the *Nyquist* criterion states that *alias* (fictitious) frequencies will appear if we attempt to sample a signal containing frequencies at, or above, half the sampling frequency.

Mild abuse of the Nyquist criterion produces low-frequency aliases that were not in the original waveform. You can demonstrate aliasing to yourself by laying two pieces of fine netting or perforated metal one on top of the other and then sliding and rotating one against the other. Large circles appear and disappear, which are known as *Moiré* patterns (after a type of lace). The reason that Moiré occurs is that one piece of netting is sampling the other, but the sampling frequency is the same as the sampled frequency. As the netting slides, the relative phase changes, which changes the frequency of the aliases.

To avoid aliasing, the analogue to digital convertor must be preceded by a low-pass filter known, predictably, as an anti-aliasing filter. As an example, a non-recording quality computer soundcard operating at a sample rate of 44.1 kHz should be preceded by an anti-aliasing filter having a cut-off frequency of ≈20 kHz. Thus, if we used such a soundcard for distortion measurements, it would be blind to frequencies above 20 kHz and probably attenuate frequencies just below 20 kHz. Conversely, digitising oscilloscopes cannot be preceded by anti-aliasing filters (because their sample rate changes over a wide range). We must either choose a sufficiently high sample rate that we are confident that aliasing will not occur (such as a 192 kHz recording quality soundcard), or add an external anti-aliasing filter.

Scaling

When we plot numbers on graph paper, we must choose a scale that conveniently fits our numbers to the lines on the paper. As an example, if the graph paper has 10 large squares, each composed of 10 small squares, and we had a current measurement ranging from 0 mA to 8 mA, then we would set a scaling of one large square=1 mA. This may seem obvious, but what if we chose a scale of one large square=0.1 mA, or even 10 mA? In the first instance, our data would overload the graph paper, and in the second, it would hardly be seen. The purpose of scaling is to match the range of the parameter to the range of our measurement system.

Similarly, when we convert an analogue parameter to a number, we first scale the parameter to be measured, and then we can measure it. Incidentally, this is why 4¾ digit DVMs specify their *basic accuracy* on the 0–5 V range. Their measurement system actually measures from 0 V to 5 V, and the range switch selects attenuators or amplifiers to scale external voltages or currents to fit this system. Practical problems mean that the scaling cannot be perfect, hence increased errors on all ranges bar 0–5 V. (3½ digit DVMs typically measure from 0 mV to 200 mV, so they specify their basic accuracy on the 0–200 mV range.)

Quantisation

If we have correctly scaled the parameter to be measured, the precision by which we make the numerical measurement is determined by the number of *quanta* or steps available. The process of comparing the continuously variable parameter against the series of fixed steps and finding the step that is closest is known as *quantisation*. The result of quantisation is a number, although it is commonly known as a digital *word* that is a code for the input voltage, so this is sometimes known as Pulse Code Modulation or *PCM*.

We now have a succession of digital words appearing at regular intervals that we write into digital memory known as a *waveform record*.

Number Systems

Computers count in the *binary* (0, 1) system, rather than the *denary* (0–9) system used by humans. This seems rather limited because it means that we can count to nine, but no higher, and the computer can only count to one. The solution in both cases is to scale the counting system. Each time we reach 9, and want to add 1, we record the new number as a scaled 1, but this is an inconvenient term, so we call it 'ten'. There is no reason why we should not scale tens, so you will probably remember 'hundreds, tens and units' from when you were first taught addition. The scaling is shown more formally in Table 3.1.

Table 3.1 Powers in the Denary Number System

Thousands	Hundreds	Tens	Units	Tenths	Hundredths	Thousandths
1,000	100	10	1	1/10	1/100	1/1,000
10^{3}	10^{2}	10^{1}	10^{0}	10^{-1}	10^{-2}	10^{-3}

The terms 'hundreds, tenths', *etc.* are simply powers of the *base*, in this case 10. The binary system works in exactly the same way, but because it uses 2 as its base, rather than 10, its table is slightly different.

Thus, even though the binary system only counts from 0 to 1, if we use a *word* with enough *bits*, we can have any number we like (see Table 3.2).

Table 3.2 Powers in the Binary Number System

32	16	8	4	2	1	1/2	1/4	1/8	1/16	1/32
2^{5}	2^{4}	2^{3}	2^{2}	2^{1}	2^{0}	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}

Precision

Computers use binary numbers, so if we make our numerical measurement more precise by using smaller quantising levels, there must be more of them, and a binary word comprising more bits is required. CD used a 16-bit word, and because there are two possible states for each bit, the total number of different levels that can be described by a 16 bit word is 2^{16}=65,536. Similarly, a 24 bit system can describe 2^{24}=16,777,216 different levels, but requires one-and-a-half times as much memory to store each word (24/16=1½).

As a rule of thumb (ignoring dither), the Dynamic Range (DR) of a digital system is:

$$\mathrm{DR}_{\mathrm{dB}} = 6n$$

where n=number of bits.

Thus, a 16 bit system has a theoretical dynamic range of 6×16=96 dB and a 24 bit system 144 dB (never achieved because Analogue to Digital Convertors (ADCs) simply aren't that good).

We could decide to be more precise by making *more* numerical measurements. Sampling twice as often doubles the memory required.

To sum up, a more precise description generates more data, requiring more memory, and this will become significant later.

The Fast Fourier Transform (FFT)

The reason for converting our analogue signal to a digital signal was to allow mathematical techniques to be applied to the resulting numbers and allow patterns to be seen. (Humans are good at recognising patterns, so any technique that reveals patterns helps understanding.) An oscilloscope allows us to spot patterns that repeat in time, such as a spike that occurs each time a sine wave changes polarity, but a spectrum analyser allows us to spot patterns in frequency, perhaps a small spike that indicates fifth harmonic distortion.

The FFT is a mathematical tool that converts data initially presented as a graph of a parameter plotted against time into that parameter plotted against frequency. The FFT is immensely powerful, but it has its limitations.

The Periodicity Assumption

In converting from time to the frequency domain, the mathematics of the FFT makes the assumption that the waveform to be analysed repeats itself periodically. This assumption may seem trivial, but it has *major* repercussions.

If we capture a *single* cycle of the waveform, and draw it around a circular drum (like an old-fashioned seismograph) such that the end of the cycle just meets the beginning, then by rotating the drum we can replay the waveform ad finitum and reproduce our original signal. Unfortunately, any uncertainty as to the precise timing of the end of the cycle causes a step in level when we attempt to loop the recorded cycle back to itself on replay. However, if we capture more cycles on our drum, the glitch occurs proportionately less frequently and causes less of an error. Thus, capturing 1,000 cycles reduces the error by a factor of 1,000, but multiplies the length of the waveform record by the same amount.

Windowing

Another way of reducing the step is to force periodicity by applying a *window* to the waveform record. In this context, a window is a variable weighting factor that multiplies the values of the samples at the ends of the waveform record by zero, but applies a greater weighting (≤1) to samples towards the middle. Since any number multiplied by zero is zero, this forces the end samples to zero and allows the waveform record to repeat without glitches (see Figure 3.4).

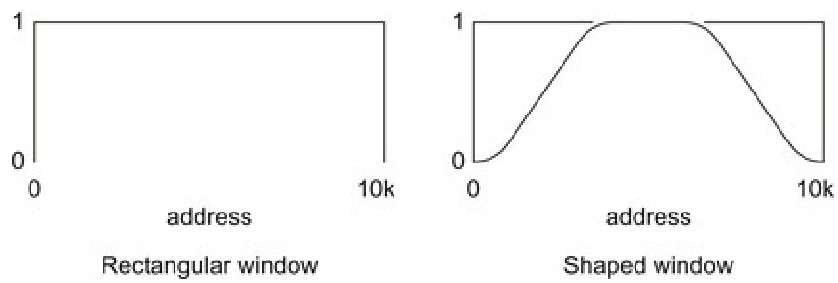


Figure 3.4 Windowing forces periodicity.

If windowing distorts the waveform record, it must distort the results of the FFT. Windowing either spills energy from high-amplitude bins into adjacent bins, which produces visible skirts around frequencies having high amplitude, or changes bin amplitudes. (Because the process of sampling broke time into discrete slices, the results of an FFT must produce frequencies in discrete slices, and these are known as *bins*.) All windows are, therefore, a compromise between frequency and amplitude resolution.

A window that does not modify sample values is known as a rectangular window (because it multiplies by a constant value of 1 over the entire waveform record). Because the rectangular window does not modify sample values, it does not cause spreading between bins, and offers the best frequency resolution, but any periodicity violation causes amplitude errors. FFT software generally offers user-selectable windowing, such as Blackman–Harris, which shapes the ends of the waveform record to avoid periodicity violation and minimise consequent amplitude errors at the expense of spreading between frequency bins. Alternatively, Hamming or Hanning windows improve frequency resolution (reduced bin spreading) at the expense of increased amplitude errors.

Best results are obtained by synchronising the oscillator to the FFT system so that only complete cycles without periodicity errors are captured in the record, allowing a rectangular window to be used. If true *synchronous* FFT is not possible, then a practical compromise is to trigger the analyser from the fundamental frequency and finely adjust oscillator frequency for minimal skirts on the highest amplitude bin.

If multiple waveform records are captured, they can be averaged together to reduce errors. This is a very powerful technique, although it slows measurement speed.

How the Author's Distortion Measurements Were Made

An MJS401D analogue audio test set and a Tektronix TDS3032 oscilloscope with FFT option were used in combination.

Distortion measurements were made at 1 kHz, and a 400 Hz 36 dB/octave high-pass filter was engaged to reject hum. The meter used an RMS rectifier to sum harmonic amplitudes correctly, and its bandwidth was restricted to the audible range by a 22 kHz 36 dB/octave low-pass filter. The distortion residual was then passed to the spectrum analyser.

The 9 bit oscilloscope/spectrum analyser used a sample rate of 50 kS/s to maximise the number of cycles captured, and the 22 kHz filter in the MJS401D formed the anti-aliasing filter. The oscilloscope was triggered from the 1 kHz fundamental, and rectangular windowing was used, with the MJS401D oscillator frequency fine-tuned for minimum skirts to give quasi-synchronous FFT. To reduce the contribution of random noise, the FFTs were averaged over 16 records (12 dB noise reduction), each 10 kbits long, resulting in >50 dB of reliable spectrum analyser dynamic range.

Because the dynamic range of the audio test set is added to that of the spectrum analyser, the main limitation becomes the distortion residual of the test set, and figures below −90 dB should be viewed with caution.

Now that the distortion measurement method is understood, we can use it as necessary to test and compare low-distortion valve circuits.

Designing for Low Distortion

There are many ways of reducing distortion, or to put it less charitably, it's easy to generate distortion inadvertently. To simplify investigation, we will consider the distortion generated by a single stage before progressing to multiple stages. We will consider:

- Signal amplitude
- Grid current
- Distortion reduction by parameter restriction
- Distortion reduction by cancellation
- DC-bias problems
- Individual valve choice
- Coupling from one stage to the next.

We will investigate each aspect in turn and test our hypotheses with practical measurements.

Signal Amplitude

In theory, the distortion generated by triodes is predominantly second harmonic

(H2). A common-cathode amplifier using 417A/5842 was set up to test the theory (see Figure 3.5).

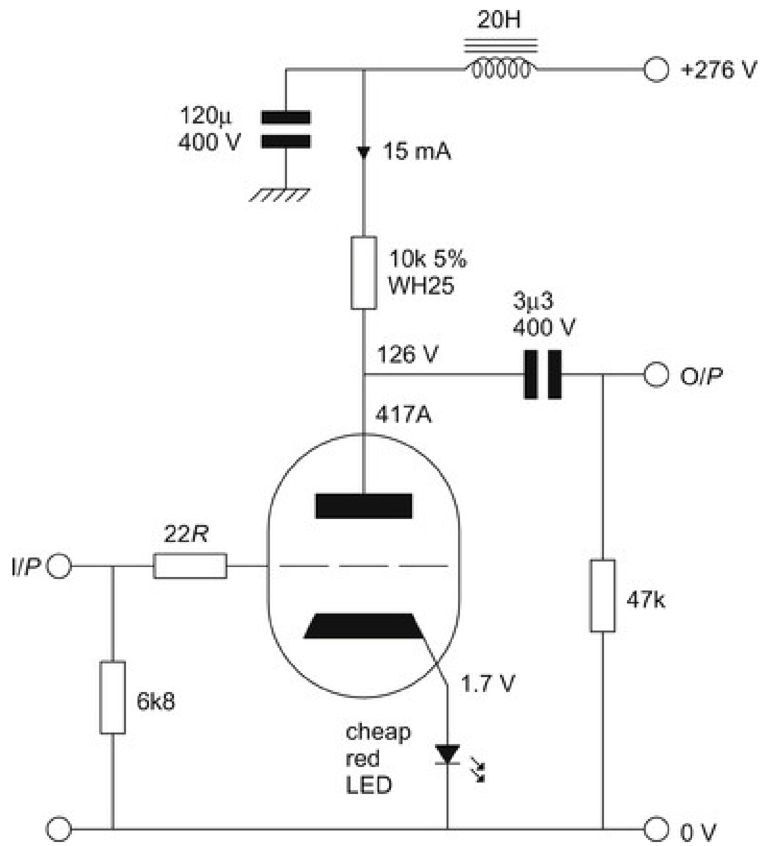


Figure 3.5 Common cathode test amplifier circuit.

Twenty-two 417A/5842s were tested at an output level of +18 dBu (6.16 V $_{RMS}$); their results were averaged and are presented in Table 3.3.

Table 3.3 Average Level of Distortion Harmonics Produced by 417-A Common-Cathode Amplifier at +18 dBu

Harmonic	Level (dB)
H1 (fundamental)	0
H2	−41
H3	−100
H4	−95

The distortion generated by the 417A/5842 clearly *is* dominated by H2. The 417A/5842 type turned out to be a particularly good example, but even for the worst valves H2 is more than 20 dB higher than any other harmonic. This is useful because it means that we can use the following formula to estimate distortion when drawing and comparing loadlines on reliable graphs:

$$\%D_{\text{2nd harmonic}} \approx \frac{V_{\text{quiescent}} - ((V_{\max} + V_{\min})/2)}{V_{\max} - V_{\min}} \times 100\%$$

The shape of a triode's transfer characteristic is a simple curve ($I_a \propto V_{gk}^{3/2}$), so traversing a smaller distance of the curve becomes a closer approximation to a straight line, and there should be less distortion. This hypothesis was tested by a 7N7/D3a μ-follower circuit (see Figure 3.6).

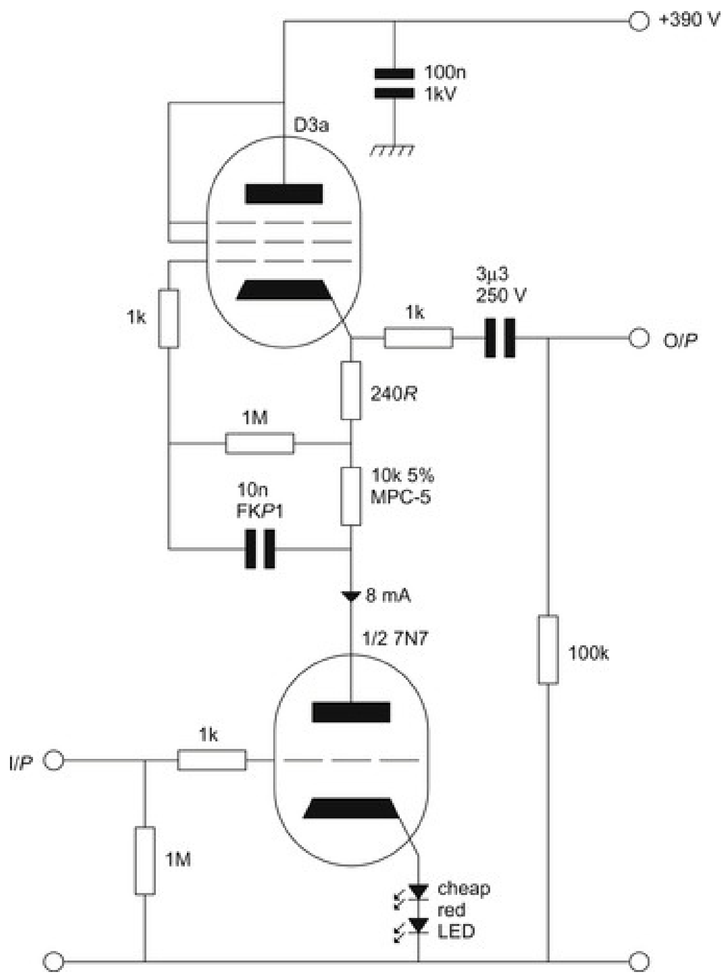


Figure 3.6 μ-Follower linearity test circuit.

In order that the test circuit should not be falsely good when approaching grid current, it was driven from a source resistance of 64 kΩ, thus replicating typical conditions of use. The upper limit of measurement was set by the onset of grid current at an output of +34 dBu (THD+N=−43 dB). The lower limit of reliable measurement was set by the ability of the analogue analyser to lock cleanly to the distortion waveform, which began to degrade at an output of +14 dBu (THD+N=−63.5 dB). Between these limits, the output level was changed in 1-dB steps, and a graph of THD+N was plotted against output level (see Figure 3.7).

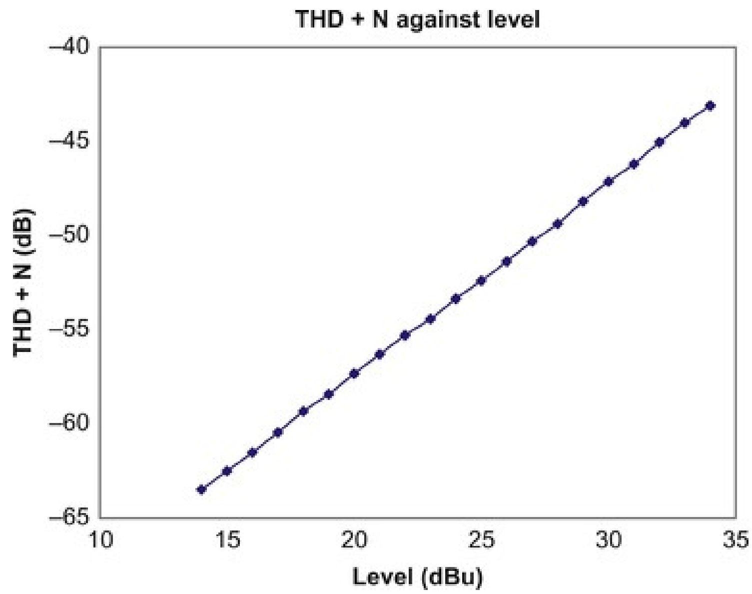


Figure 3.7 Graph of distortion versus level for μ-follower test circuit.

The graph clearly shows that THD+N is directly proportional to output level. Thus, a distortion measurement of 1% at 15 V_{RMS} implies distortion of 0.1% at 1.5 V_{RMS}. This observation is extremely useful if it is necessary to estimate the distortion of a triode handling small-signal voltages – such as would normally be encountered early in an RIAA stage.

The supposition that triode distortion is predominantly H2 and is proportional to level is true for all triodes when used with practical resistive anode loads. The effect of an active load ($R_L \Rightarrow \infty$) is to minimise H2, but barely changes higher harmonics. Once H2 has been minimised, the effects of the higher harmonics become more significant, with the result that some triodes do *not* then have distortion that is proportional to level. If you use an active load, you may need to check whether distortion remains proportional to level for that particular type of valve.

Cascodes and Distortion

The cascode is an ideal small-signal RF circuit. It works by placing two devices in series across the high tension (HT) so that the upper device provides most of the gain and loads the lower device with the upper device's load divided by its voltage gain. This means that the lower device sees a very low resistance load (near vertical loadline), lowering its gain (and in a semiconductor circuit) forcing its voltage gain to unity. From an RF point of view, the reduction of gain in the lower device is extremely advantageous because it reduces Miller capacitance from the output of the device to its input. From an audio point of view, the low impedance load greatly increases H2. Worse, the series connection means that each device has a greatly reduced HT voltage, which tends to increase distortion. For a phono stage, low noise is the primary consideration, so the lower device requires high g_m because it reduces noise, but a careful balance of DC current between the upper and lower devices is required to optimise the dynamic range between noise and distortion. However, because distortion is proportional to amplitude, minimising signal amplitude minimises distortion, and the dominant H2 can be removed by cancellation if a pair of cascodes is configured as a differential pair.

Grid Current

Distortion changes with V_a and I_a will be investigated later because they are caused by changes in the small-signal parameters μ, r_a and g_m that are normally assumed to be constant. Thus, unless we also need to maximise voltage swing, our choice of operating point simply needs to be wary of grid current and cut-off. The problems of cut-off are obvious, but grid current causes far more problems.

Distortion due to Grid Current at Contact Potential

As V_{gk} approaches 0 V, grid current begins to flow, and the input resistance of the valve falls dramatically. If the source driving the valve had r_{out}=0, this would not be a problem, but it is highly likely that it has significant output resistance, and the potential divider that is momentarily formed at the positive peaks of the waveform where grid current flows clips the input signal. Symmetrical clipping produces a square wave composed of odd harmonics, but grid current clips asymmetrically, so even harmonics can also be expected.

The distortion caused by grid current is obnoxious because it is composed of high-order harmonics. The following traces were obtained by driving the lower valve of a μ-follower into grid current from a source resistance of 47 kΩ. The input signal was increased until distortion of the output waveform was just visible on a carefully focussed analogue oscilloscope. The THD+N was measured to be 2%, and the distortion residual had a very distinctive waveform (see Figure 3.8).

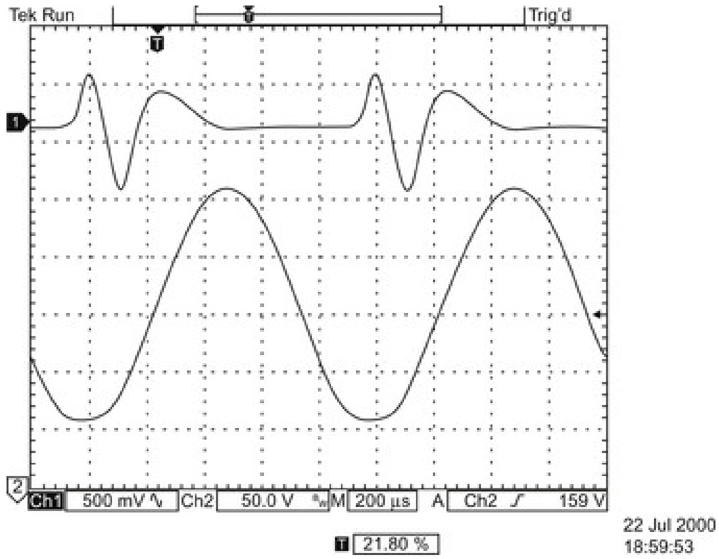


Figure 3.8 Upper: The distinctive distortion waveform caused by grid current. Lower: Soft clipping caused by grid current.

FFT analysis of the distortion residual revealed a spray of high-level, high-order, odd and even harmonics (see Figure 3.9).

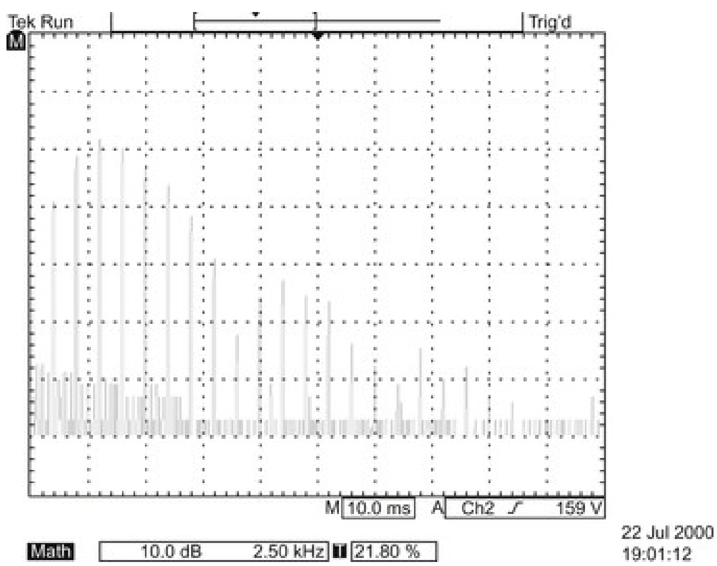


Figure 3.9 Distortion spectrum produced by grid current acting on 1-kHz sine wave. Vertical scale: 10 dB/div. Horizontal scale: 2.5 kHz/div. (DC−25 kHz.)

Although grid current occurs at $V_{gk}=0$ V in an ideal valve, practical valves enter grid current a little earlier due to the thermocouple effect of heated junctions between dissimilar metals within the valve, and the average energy of electrons in the electron cloud above the cathode surface. Typically, grid current begins at $V_{gk} \approx -1$ V, and this is known as the *contact potential*. Measuring distortion whilst driving from a high impedance source is an excellent way of determining contact potential for a given valve.

Distortion due to Grid Current and Volume Controls

It is perfectly possible to change the design of the volume control preceding an amplifier stage and *measure* a change in distortion. The most common type of volume control is a resistor from which a variable tapping is taken, either a wiper moving along the resistive track, or a switch wiper selecting a tapping from a chain of fixed resistors (see Figure 3.10a).

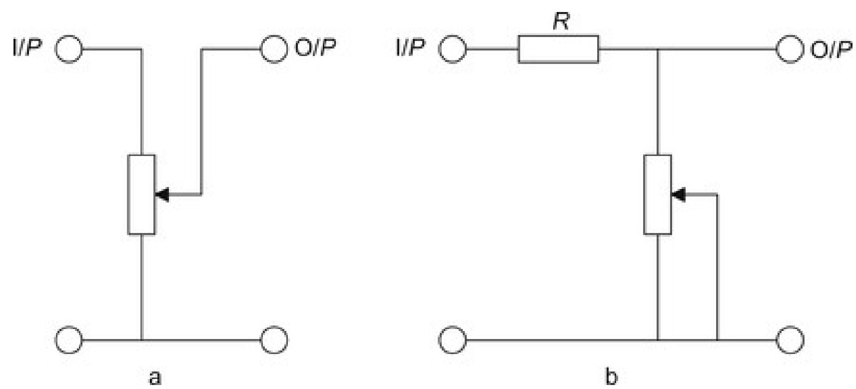


Figure 3.10 Fundamental basis of most volume controls.

Alternatively, we can use a fixed series resistor followed by a variable shunt resistor (see Figure 3.10b).

Unfortunately, the circuit of Figure 3.10b has much higher output resistance than that of Figure 3.10a. Measuring distortion whilst driving from a high source resistance is a very sensitive test of gas current because the (non-linear) gas current develops a voltage across the source resistance, which is in series with the signal. As the source resistance rises, so does the distortion.

A 6545P cathode follower biassed by an EF184 constant current sink was tested at +20 dBu (7.75 V_{RMS}). When driven from a 5 Ω source, the distortion was 0.02%. A 100 kΩ potential divider volume control has a maximum output resistance of 25 kΩ, so distortion was also measured with a 25 kΩ source resistance, and found to be unchanged. However, when the source resistance was increased to 1 MΩ, the distortion rose to 0.2%. Admittedly, it is unlikely that the source resistance would be as high as 1 MΩ, but 100 kΩ would be quite possible from the volume control in Figure 3.10b.

Operating with Grid Current (Class A2)

Most Class A amplifiers operate without grid current because this allows a high grid resistance that is easily driven. Once V_{gk} becomes positive, rather than repelling electrons from the cathode, the control grid becomes a weak anode and attracts electrons, most of which are then captured by the true anode that is at a much higher voltage, but some flow out of the grid as grid current. Grid current has important consequences:

- The electron stream from the cathode is divided between grid and anode current, implying partition noise. However, as the most likely use of Class A2 is in the output stage, where signal voltages are high, this noise is unlikely to be a problem.
- There is a potential difference between the grid and the cathode (V_{gk}) and current flowing through the grid (I_g), so it must be dissipating power in exactly the same way as an anode. If the grid was not designed to dissipate power, it will quickly heat, distorting its shape and possibly destroying the valve.
- Because the input resistance of the grid when operating as an anode is very low, imposing a signal voltage on the grid requires considerable power ($P = V^2/R$), which must be provided by the driver stage.
- However, because the grid is driven positive, it is possible to drive the anode of a triode far closer to 0 V than if V_{gk} was negative. The efficiency of the output stage is thus significantly increased.

Driver stages for Class A1 are voltage amplifiers that only need to supply sufficient current to charge and discharge the Miller capacitance of the output stage, but a driver stage for Class A2 must provide power. There are two ways in which this power can be delivered.

The driver stage can be designed to be a small power stage. One possible choice is the common-cathode dual triode 6N7 that can be operated either in push–pull or single-ended with the two triodes paralleled to double P_a. A transformer reflects its load impedance by n^2, so a step-down transformer with a voltage ratio of 2:1 increases the load impedance seen by the driver valve by a factor of four. Because a transformer in the anode circuit of a valve theoretically allows V_a to swing to 2 V_{HT}, requiring double the anode swing is not a problem. Additionally, the low R_{DC} of the secondary reduces the chances of thermal runaway in the output stage. Sadly, good driver transformers are even more difficult to design than output transformers because they operate at higher impedances.

Alternatively, the Class A2 stage can be driven DC-coupled from a cathode follower. A power valve is still required, but it no longer needs to be able to swing many volts. Power frame-grid valves that have high mutual conductance, but low $V_{a(max)}$, such as the 6545P and E55L, are ideal as power cathode followers. Unfortunately, frame-grid valves tend to have modern, efficient heaters (thin heater/cathode insulation), which means that their $V_{hk(max)}$ is quite low, possibly causing a problem if the Class A2 stage requires significant grid voltage swing. To bias the Class A2 stage correctly, the cathode of the cathode follower can only be slightly positive, but we need a reasonably large value of R_L to ensure linearity of the cathode follower, so a negative supply is required (see Figure 3.11).

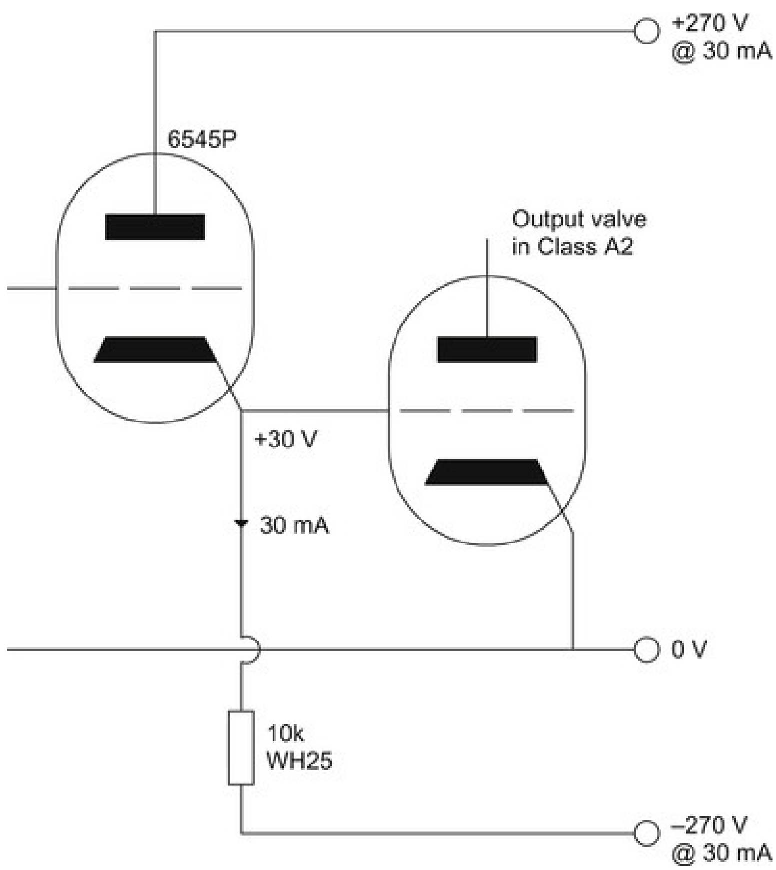


Figure 3.11 Using a DC coupled power cathode follower to drive Class A2.

A low output resistance is offered by both of the previous solutions, but it is not zero. Because $r_{out} \neq 0$, it forms a potential divider with the input resistance of the Class A2 stage, causing attenuation. If V_{gk} swings negative, the input impedance of the Class A2 stage becomes infinite, and there is no longer any attenuation, causing distortion. No distortion advantage can be gained using a constant current sink load for the cathode follower because it faces the low-resistance load of the Class A2 grid, but it allows a lower-voltage negative supply, reducing cost. Whereas a Class A1 stage should never be driven into grid current for fear of distortion, the Class A2 stage must never be allowed to stray *out* of grid current, or distortion will result. Thus, Class A2 power stages typically use high- μ (μ>100) transmitter valves originally intended for Class B use because these valves are designed to pass low anode currents at V_g=0 V even at their intended anode operating voltage.

Distortion Reduction by Parameter Restriction

Triodes produce primarily H2 distortion because as r_a changes with I_a, the attenuation of the potential divider formed by r_a and R_L changes, with more attenuation on one-half cycle of the waveform than on the other. However, there are ways of reducing this distortion:

- Use a large value of R_L. If $R_L \gg r_a$, then the changing attenuation of the potential divider is insignificant because the attenuation itself becomes negligible.
- Hold I_a constant so that r_a cannot vary. This implies an active load such as a constant current source is the basis of the μ-follower.

These two methods are actually very similar because both seek to make $R_L \gg r_a$. (For an ideal constant current source, $r_{slope}=\infty$.) In general, for a given HT voltage and I_a, replacing the load resistor R_L with a constant current source can be expected to reduce H2 by a factor of ≈7.

Once the previous methods of distortion reduction have been used, the amplifying valve sees an almost horizontal AC loadline, and when R_L>50 r_a, the far lesser effect of variation of μ with V_a becomes observable. The variation of μ with V_a can be reduced by avoiding operation at low I_a (where the anode curves begin bunching) and by choosing a valve whose curves bunch less as I_a tends to 0 (see Figure 3.12).

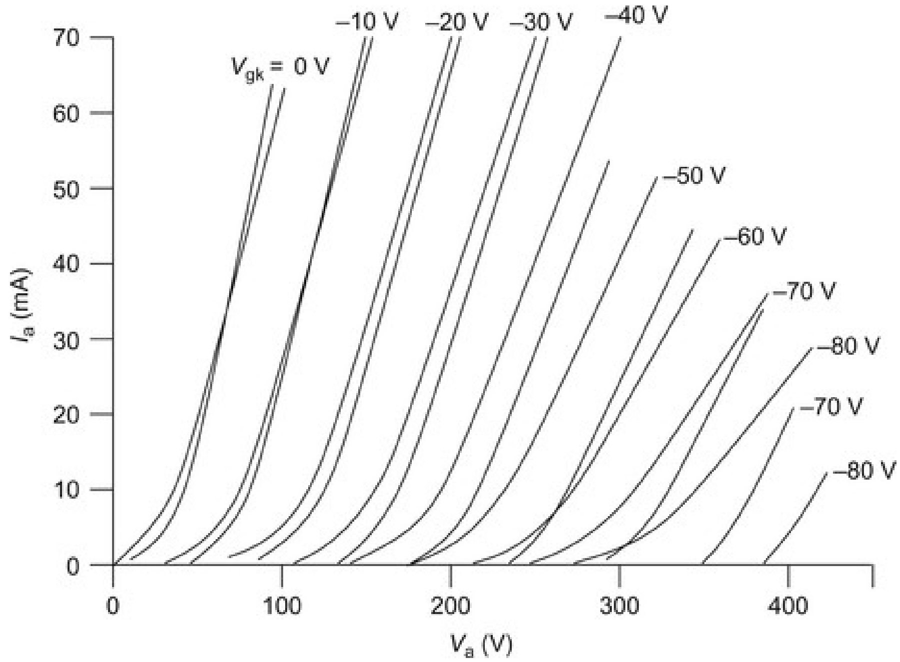


Figure 3.12 Grid construction: More fine turns (dashed) versus fewer coarse turns (solid).
(After Henderson (GEC) [2]).

Bunching of anode curves is caused by the inevitable non-uniformity of the electric field between the grid wires at the grid/cathode region, so the graph compares two GEC [2] directly heated triodes having similar μ, but the solid curves are due to a grid wound with a few turns of coarse wire, and the dashed curves are due to a grid wound with more turns of fine wire. Unfortunately, as the grid wire becomes finer, it is less able to support itself, but a frame-grid allows arbitrary thickness of wire, which is why the E88CC, and particularly the 6545P (both frame-grid), exhibit very little bunching.

Alternatively, it may be possible to hold V_a constant. Clearly, this cannot be done if the stage has gain, but a cathode follower can be arranged to have constant I_a and V_a simultaneously [3] (see Figure 3.13).

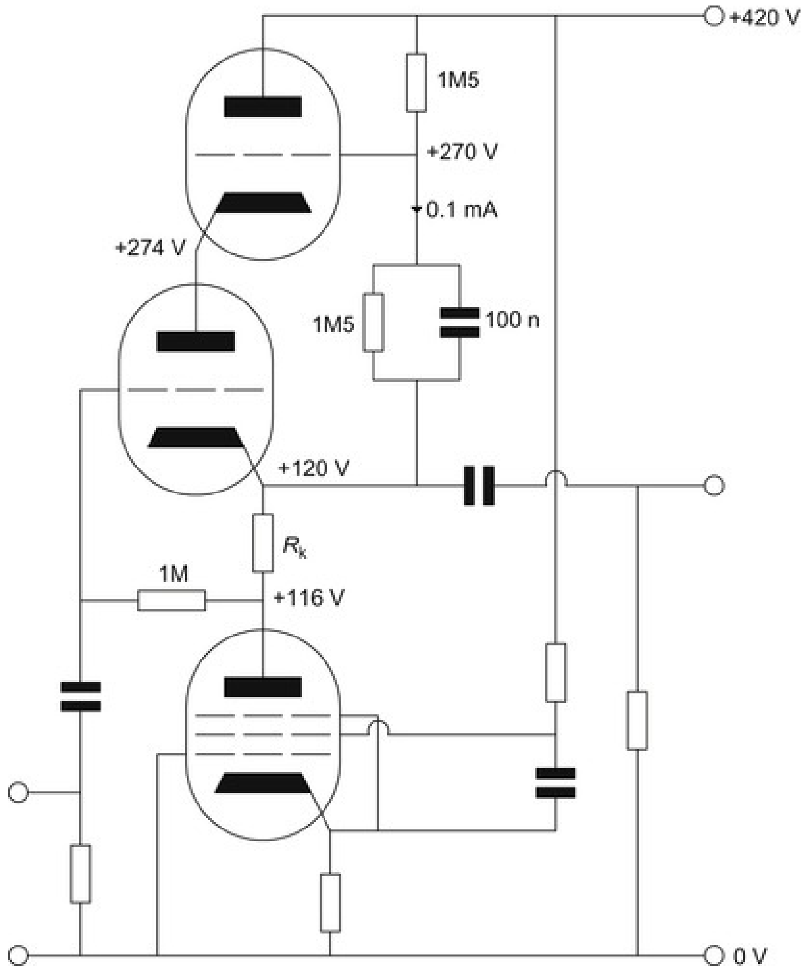


Figure 3.13 The cathode follower allows constant I_a and V_a to be forced.

The middle valve is the cathode follower. The lower valve is the traditional pentode constant current sink that forces I_a in the cathode follower to be constant. The upper valve is also a cathode follower and should have high μ and g_m, so the 6545P (μ=52) is ideal. The upper valve sees a high impedance load, so its gain is:

$$A_v = \frac{\mu}{(\mu + 1)} = \frac{52}{(52 + 1)} = 0.98$$

The upper cathode follower's grid is AC coupled to the output of the middle cathode follower, and because its gain is almost unity, its cathode is at the same AC voltage as its grid. Thus, even when the middle cathode follower swings its cathode, the upper cathode follower forces its anode to swing by an almost identical amount, and constant V_a has been enforced simultaneously with constant I_a.

Unfortunately, the improvement is accompanied by significant costs:

- The required HT voltage has been raised by V_a of the upper cathode follower.
- We need a third elevated heater supply (for the upper cathode follower).
- Cathode followers are already prone to instability, and bootstrapping the anode of one with the output of another invites further problems.

You might have a different opinion, but the author feels that a carefully designed cathode follower sitting on a constant current sink already challenges his test equipment.

Distortion Reduction by Cancellation

In theory, if two common-cathode triode amplifiers are operated in a cascode, because each stage inverts, the distortion of the second triode is inverted with respect to that produced by the first triode, and cancellation should occur. However, a moment's thought shows that this is unlikely to occur to any significant degree. Distortion is proportional to level, and because the second triode has gain, it produces a significantly higher level, and therefore proportionately higher distortion, than the first triode. A small amount of cancellation may occur, but the improvement is $\propto 1/A_2$, so if the second triode was a type 76 (μ=13) and A_v=10, we might reduce distortion from 1% to 0.9%, which is less than the sample-to-sample variation of distortion in either valve.

Perhaps we could choose the second valve to be much more linear than the first so that both produce equal amounts of distortion. Low- μ valves are the most linear, so an 845 (μ=5.3) should achieve A_v=4, and therefore we need a valve that produces four times the distortion of the 845. This can probably be done, and adjusting the bias of the first valve would allow complete cancellation to be achieved. However, this cancellation would be critically dependent on the gain of the 845, which is determined by R_L, yet R_L is a loudspeaker whose impedance changes with frequency. In practice, 6 dB reduction in H2 is feasible.

Surprisingly, it is possible to achieve distortion cancellation between a common cathode stage followed by a cathode follower stage, provided that the common cathode stage has its distortion minimised by a constant current load and the cathode follower has its distortion deliberately increased by an AC load (see Figure 3.14).

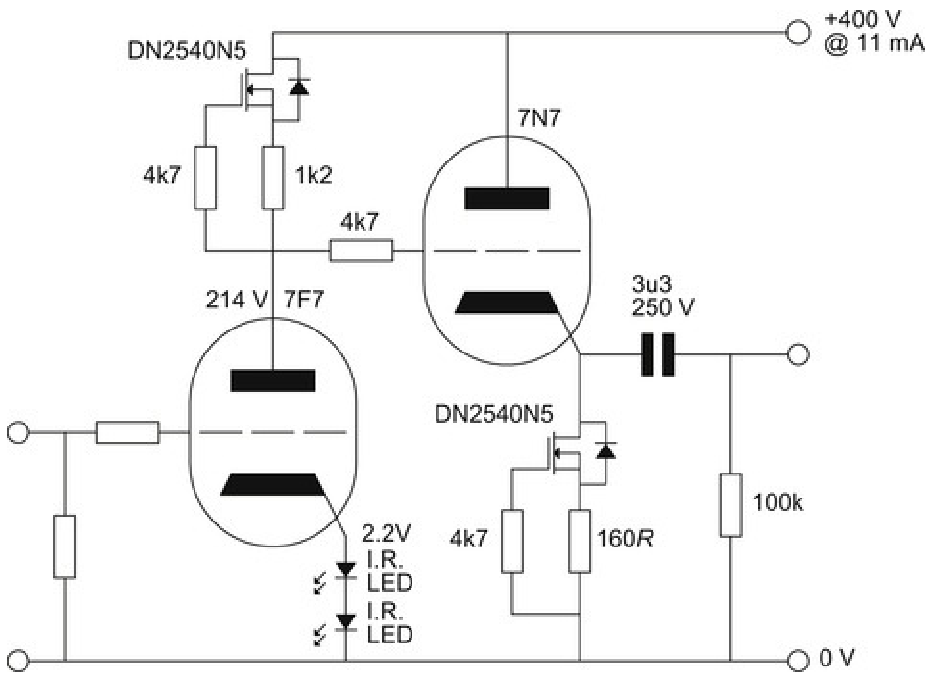


Figure 3.14 Distortion cancellation between a common cathode amplifier and a cathode follower.

The common-cathode stage produces perhaps 0.1% THD+N, and the dominant harmonic is H2, with all the others better than 20 dB down. In theory, if we could cancel H2, we would be left with only the higher harmonics, and because they are 20 dB further down, our THD+N would have dropped by 20 dB from 0.1% to 0.01%. In practice, the situation is a little more complicated (see Figure 3.15).

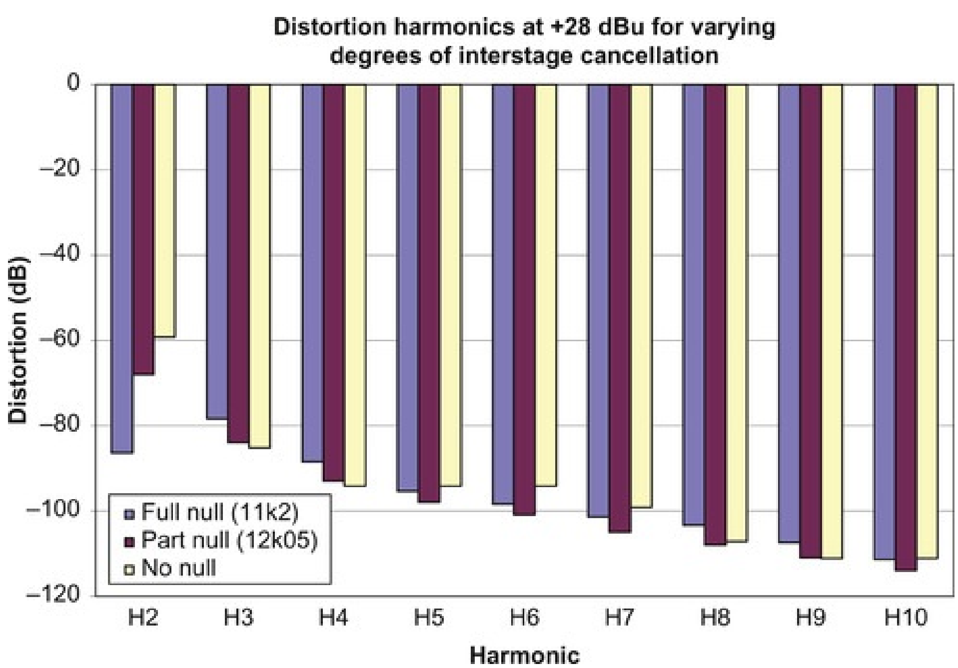


Figure 3.15 Distortion spectra resulting from no nulling, perfect and imperfect nulling.

We can see that full nulling (11.2 kΩ AC load) reduces H2 by 27 dB from −59 dB to −86 dB (0.005%), which is certainly impressive, but at the expense of skewing the distortion spectrum so that H3 is 8 dB higher than H2. Note that a 7.5% change in AC load resistance from 11.2 kΩ to 12.05 kΩ radically changes the distortion spectrum. Distortion cancellation can only be achieved reliably if the two valves are identical, carry exactly the same signal and have the same load conditions.

Differential Pair Distortion Cancellation

The differential pair with constant current sink tail provides optimum conditions for distortion cancellation because the signal *current* is forced to swing between the two valves with no loss. Provided that the load impedances are matched, the voltage swings at each anode must be equal and opposite, theoretically allowing perfect H2 cancellation. The anode load resistors can easily be matched to 0.2% by a DVM, and if each anode drives a cathode follower, then the shunt capacitance is so small that any imbalance is insignificant at audio frequencies. (Even at 20 kHz, X_c=1.6 MΩ for the 5 pF input capacitance of a typical cathode follower, so this is significantly larger than the typical 47 kΩ anode load resistors.)

A Mullard 6SN7GT with well-matched sections was compared in different configurations with I_a=7.5 mA and V_a=230 V. Each test circuit was measured at an output level of +14 dBu, but for the differential pair the signal *between* the anodes was measured (+20 dBu), corresponding to +14 dBu at each anode (see Figure 3.16 and Table 3.4).

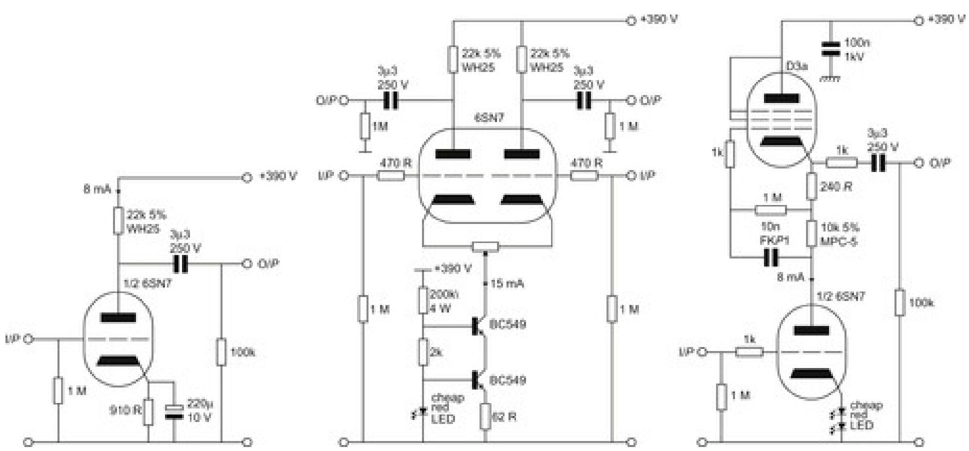


Figure 3.16 Test circuits for distortion comparison.

Table 3.4 Comparison of Distortion Harmonics for Different Topologies

Harmonic	Common cathode (dB)	Differential pair (dB)	μ-Follower (dB)
H2	−51	−77	−68
H3	−93	−89	–
H4	(−106)	–	–

As can be seen from Table 3.4, the differential pair cancels even harmonics, but sums odd ones. Although 0.0035% H3 is unlikely to be a problem, it indicates that differential pairs are ideally built with valves that produce small amounts of odd harmonic distortion. Conversely, the μ-follower was not as effective at reducing H2, but all other harmonics were below the limits of reliable measurement.

Push–Pull Distortion Cancellation

A push–pull transformer-coupled Class A output stage meets most of the requirements for distortion cancellation. In practice, unless the two valves are an accurately gain-matched pair, or provision has been made for DC and AC balance, cancellation cannot be perfect. Nevertheless, 14 dB H2 cancellation is routinely achieved because the tight coupling between the two halves of the transformer assists AC balance.

The Western Electric Harmonic Equaliser

The harmonic equaliser is a recent rediscovery championed by John Atwood and Lynn Olsen. The surprisingly clearly written patent [4] states that distortion reduction action is due to cancellation between H3 produced directly by the valve and that produced by intermodulation between H2 and the fundamental, H1. Remember that intermodulation distortion produces sum and difference frequencies, so H1+H2=H3 as stated by the patent, and H1−H2=−H1, which means that the difference frequency is at the same frequency as the fundamental, but has inverted polarity, thereby slightly attenuating it.

In push–pull form, the equaliser consists of a resistor connected from the two cathodes to ground with its value chosen purely from AC considerations – biassing is a separate issue (see Figure 3.17).

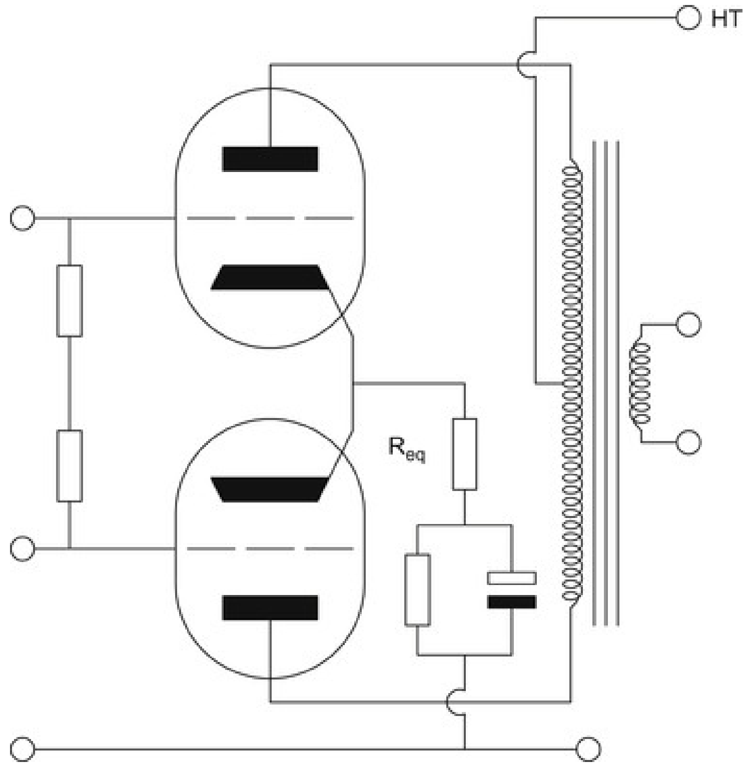


Figure 3.17 The WE harmonic equaliser.

The easiest way to determine the required value of the resistor is to temporarily substitute a constant current sink set to the required DC current and bypass it with a capacitor in series with a variable resistor (the equaliser) (see Figure 3.18).

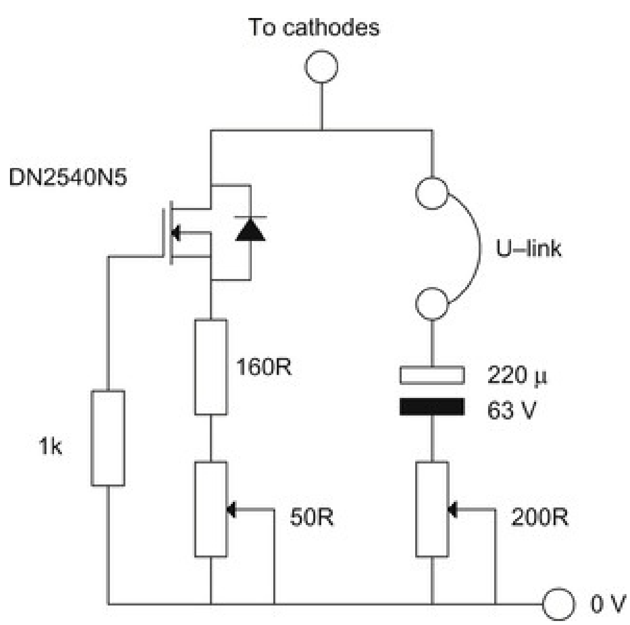


Figure 3.18 A crude JFET CCS plus capacitor-coupled variable resistor allows easy determination of the harmonic equaliser resistor

independently of DC conditions.

Sadly, the author's experiments indicate that although the equaliser affects all valves, it isn't always a positive effect. As a beneficial example, adding a 91 Ω equaliser resistor to a pair of push–pull 6S4As operating with a 9k5 $_{a–a}$ load from 320 V reduced H3 by 28 dB so that the spectrum became dominated by H2, with H3 and all others >20 dB below that (see Figure 3.19).

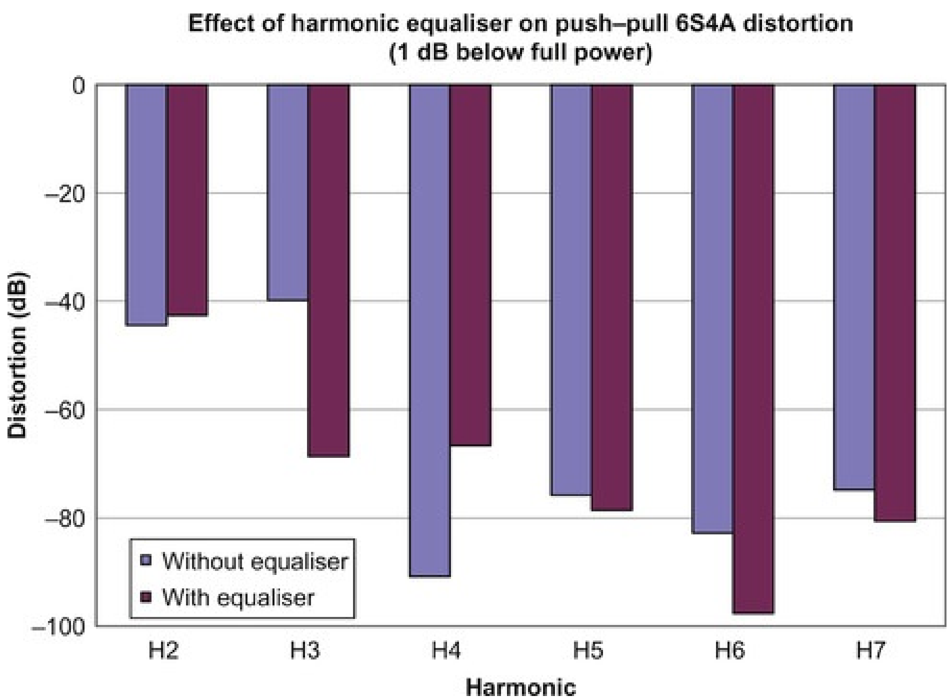


Figure 3.19 The effect of the harmonic equaliser. Note the substantial reduction of odd harmonic amplitudes.

Side-Effects of the Harmonic Equaliser

Although in a harmonic equaliser resistor simple addition can significantly improve an output stage's distortion spectrum, it does so by cancellation, and that always means that it will be sensitive to load resistance. Once the value of the harmonic equaliser resistor has been set, load resistance must not change (see Figure 3.20).

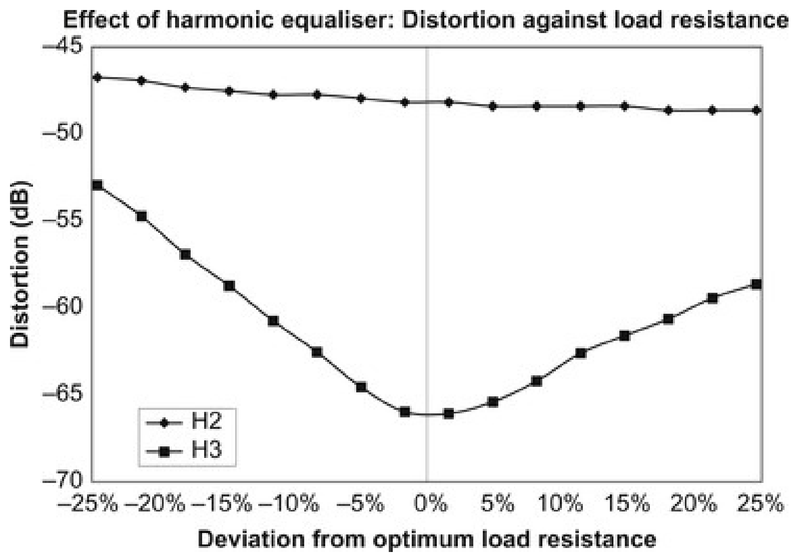


Figure 3.20 The effect of load resistance on third harmonic cancellation.

As can be seen, the load resistance ideally needs to be within ±5% of nominal to achieve the full benefit of the harmonic equaliser.

Unfortunately, the impedance of a moving coil loudspeaker rises with frequency due to voice coil inductance and has a low frequency electrical resonance that looks like an inductor and a capacitor in parallel due to the mechanical resonance formed by the mass of the cone and compliance of the suspension. In order for an output stage's harmonic equaliser to work correctly, the amplifier must be connected directly to a single resistive loudspeaker and the harmonic equaliser resistor must be set to match that specific loudspeaker. Fortunately, most moving-coil loudspeakers can be rendered adequately resistive above resonance by the addition of an appropriate Zobel network across their terminals (see Figure 3.21).

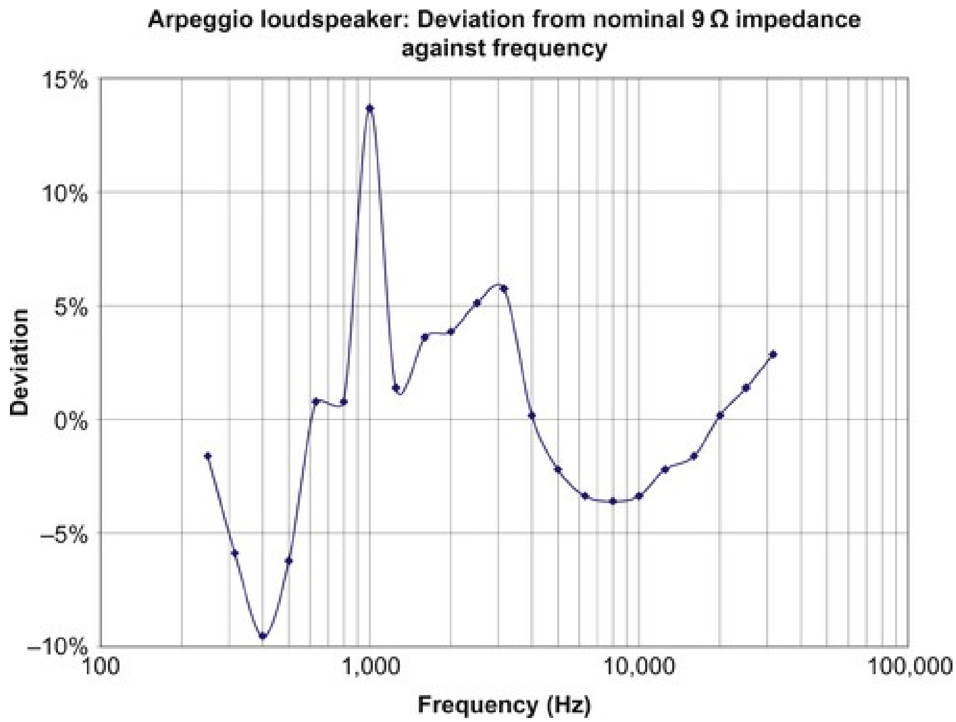


Figure 3.21 The author's 'Arpeggio' loudspeaker is an almost resistive load >250 Hz.

Unfortunately, applying the harmonic equaliser tends to increase the output resistance of the output stage. The significance of this increased output resistance is that the loudspeaker system must be designed to be driven by this specific non-zero output resistance.

In short, an amplifier having an output stage using the harmonic equaliser technique cannot be used as a universal amplifier. Whether it is part of a complex loudspeaker system employing an active crossover or a loudspeaker with a single full-range driver, it is a complementary system where the amplifier must be designed for a specific loudspeaker, and the loudspeaker's acoustic loading must be designed with the specific amplifier in mind. It is probably the combination of these requirements that has prevented widespread adoption of the harmonic equaliser, yet the requirements can be met, as we will see in Chapter 6.

Although an explicit harmonic equaliser might not have been intended, cathode feedback sometimes results in distortion reduction exceeding the feedback factor, implying cancellation and consequent load sensitivity. For both amplifiers, cathode feedback in the Quad II power amplifier and in the author's 'Scrapbox Challenge' produced better results than expected from the feedback factor alone, implying cancellation.

DC Bias Problems

Having chosen the topology of a stage with great care, we choose an operating point that cunningly maximises output swing, minimises distortion, uses standard component values, and all within the current capability of the power supply. We now need to bias the stage, which can be done in a number of ways:

- Cathode resistor bias
- Grid bias
- Cathode bias with a rechargeable battery
- Cathode bias with a diode
- Cathode bias with a constant current sink.

Cathode Resistor Bias

Bias can be achieved by inserting a resistor in the cathode path (see Figure 3.22).

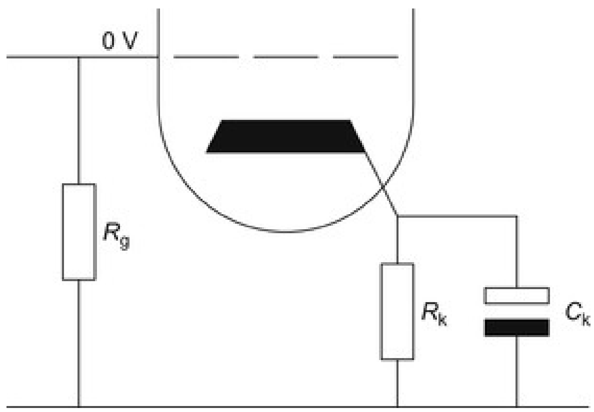


Figure 3.22 Cathode bias using a resistor.

If valve current rises, resistor current also rises, making the cathode more positive with respect to the grid, thus tending to turn the valve off and offering some overcurrent protection. This method of bias has the least sensitivity to variations between valves, making it by far the most popular bias choice. We know I_a and the required V_{gk}, so we simply apply Ohm's law to determine the required cathode resistor.

However, inserting a resistance in the cathode circuit of a single valve common-cathode amplifier creates negative feedback that reduces gain, which might not be acceptable. The traditional solution bypasses the resistor with a capacitor (which is a short circuit at audio frequencies), the cathode is connected to ground at AC, and negative feedback is prevented. It is generally argued that the audio bandwidth extends from 20 Hz to 20 kHz, and that audio electronics should be as nearly perfect as possible within this bandwidth. Because r_k tends to be low (typically <500 Ω), the cathode bypass capacitor needs to be quite a high value, forcing it to be an electrolytic type. Unfortunately, such a capacitor is typically +25%, −15% tolerance so this haphazard time constant should not be allowed to affect any formal filtering action, and its value is usually set to produce $f_{-3\ \mathrm{dB}}=1$ Hz, which is the same as saying $\tau \approx 160$ ms.

When biassing a stage, we make the assumption that the signal voltage is sufficiently small that it does not affect the DC conditions. However, as clipping is approached, the signal voltage at the anode of a triode could be hundreds of volts peak to peak, and the distortion (which contains a DC component) temporarily lowers the average V_a. A secondary effect of this distortion is that it has a DC component that changes the mean anode current.

As an example, a common-cathode triode amplifier was tested. When the generator was muted, $V_a=117.1$ V, but when the stage was driven to a level that produced 5% THD+N, the mean anode voltage fell to 114.2 V, indicating a change in mean anode current. Since this anode current flows through the cathode-bias resistor (which is in parallel with the cathode capacitor), any change in mean anode current is integrated by the cathode CR network ($\tau \approx 160$ ms). When the overload passes, the capacitor takes 5 $\tau \approx 1$ s to recover to 99% of the previous bias point. During this time, r_a (which is dependent on I_a) will change, slightly changing r_{out}. If the circuit feeds a passive equalisation network, r_{out} is inevitably part of the design, so the change in r_{out} causes a temporary frequency response error. Although a minor frequency response error could be considered irrelevant when the amplifier is producing 5% THD+N, a frequency response error that decays to zero over a period of 1 s after overload might not be so acceptable.

The bias shift effect can be observed by monitoring the DC voltage across the cathode bypass capacitor with and without a large sine wave at the anode. This method has the advantage that an ordinary DVM can be used, whereas measuring at the anode requires an instrument that can measure DC accurately in the presence of significant AC.

Ideally, there should never be a shift in the operating point of a valve, whatever the signal level. Provided that the valve is never driven to produce >1% THD, cathode bias is perfectly satisfactory, but if clipping is likely, an alternative bias strategy should be considered.

Grid Bias ($R_k=0$)

If $R_k=0$, the DC component of distortion cannot cause bias shift.

Grid bias from an auxiliary low current negative supply is common in the output stages of Class AB power amplifiers, whereas battery bias is occasionally found in pre-amplifiers (see Figure 3.23).

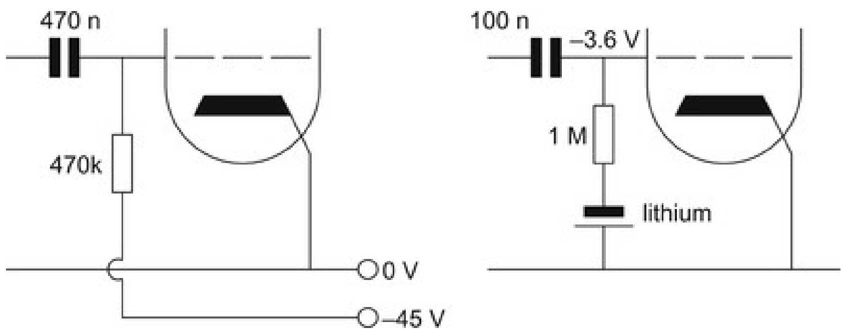


Figure 3.23 Grid bias using an auxiliary power supply, or a lithium battery.

Note that battery bias can be applied either in series or in parallel. In practice, parallel is usually preferred because the output impedance of the previous stage and the battery's series resistor form a potential divider that attenuates battery noise, whereas there is no attenuation of battery noise in series mode.

Grid voltage is fixed, and valve current is determined purely by valve characteristics, so there is no protection against overcurrent, or compensation for changes in valve characteristics with age.

Overcurrent protection is important in transformer coupled stages because the winding resistance of the transformer is negligible and an output valve is almost certainly being operated at maximum anode dissipation. Current from the supply is, therefore, almost unlimited, and a fault is likely to damage an expensive valve quickly, and worse, risks the even more expensive output transformer.

Conversely, in pre-amplifiers or driver stages using resistive or active loads, the valve is typically operated at less than half maximum anode dissipation, and the anode load limits fault current. It is quite conceivable that a fault resulting in maximum current could leave the valve operating well within its limits, and no damage at all would occur.

Rechargeable Battery Cathode Bias (r_k=0)

Rechargeable cells have extremely low internal resistance, so if they are inserted in the cathode path, they do not allow a feedback voltage to appear (see Figure 3.24).

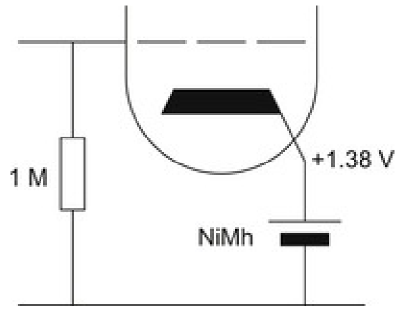


Figure 3.24 Cathode bias using a rechargeable battery operated at trickle current.

Although the diagram shows only one cell, a number of (identical) cells could be connected in series to set the required voltage, although this could become rather bulky. Provided that $I_k \leq C/10$ (C is the cell capacity in A h), the self-heating caused by continuous charging will not damage the cell. However, since the cell is in a valve amplifier, it is probably rather warmer than the battery manufacturer expected, so limiting the current to $C/20$ might be wise. An AA-size nickel metal hydride (NiMH) cell develops ≈1.38 V when charged continuously at 15 mA.

Diode Cathode Bias ($r_k \approx 0$)

Rather than using a resistor, we can use a diode for cathode bias (see Figure 3.25).

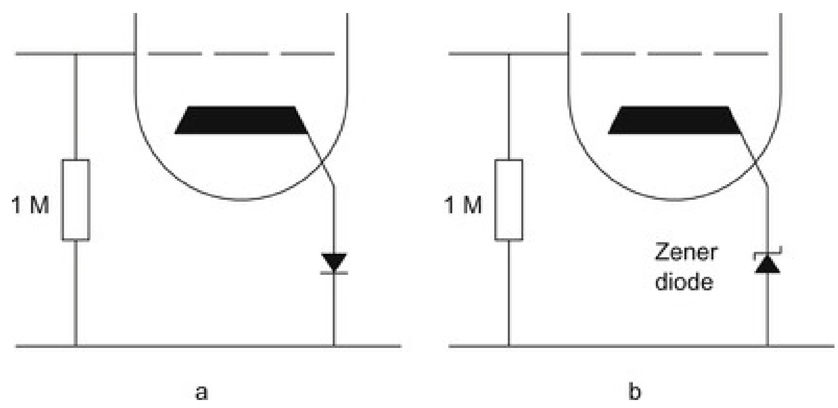


Figure 3.25 Cathode bias with a diode.

The advantage is that a diode's slope (AC) resistance is so much lower than the traditional cathode resistor that we no longer need to bypass it with a capacitor. Although diode slope resistance is low, sometimes it may be necessary to calculate its effect on r_a. Table 3.5 compares forward drops and slope resistances (r_{slope}) for various diodes.

Table 3.5 Comparison of Forward Drops and Slope Resistances of Various Diodes

Table 3.5 Comparison of Forward Drops and Slope Resistances of Various Diodes

Diode type	Forward drop at 10 mA (V)	Typical r_{slope} at 10 mA (Ω)	Pure *R* at 10 mA (Ω)	*R*/ *r*
Silicon (1N4148)	0.75	6.0	75	13
Germanium (OA91)	1.0	59	100	1.7
Infrared LED (950 nm)	1.2	5.4	120	22
Cheap inefficient red LED	1.7	4.3	170	40
HLMP6000 red LED	1.63	3.8	163	43
Cheap yellow/green LED	2.0	10	200	20
EZ81	2.3	195	230	1.18
True green LED (525 nm)	3.6	30	360	12
Blue LED (426 nm)	3.7	26	370	14
EZ80	5.5	485	550	1.13

The column 'Pure *R*' is the value required to achieve the diode forward voltage using a resistor. Thus, the column ' *R*/ *r*' is the ratio by which the diode improves upon a pure resistance – and the larger the number, the better.

The thermionic and germanium diodes barely improve on a pure resistance, so they can be discounted immediately. The winner is the red LED, but it needs to be one of the older less efficient designs that you might find cheap in a junk shop, rather than a modern high brightness type. Alternatively, the more modern (and more readily available) Agilent HLMP6000 LED is slightly better and this red LED is so good that if higher voltages are needed it is better to use a series string of red LEDs than a single, inferior colour LED.

As an example, suppose we needed 3.4 V. Referring to Table 3.5, we see that we could approximate it with a true green LED (3.6 V and 30 Ω) or a pair of HLMP6000s (3.26 V and 7.6 Ω) – the slope resistance of the two red LEDs is a quarter that of the single green LED.

As mentioned in Chapter 1, the author's measurements show that over a range of 0.3–10 mA, the forward drop of a typical HLMP6000 may be predicted from:

$$V = 0.0378 \ln I_{DC(mA)} + 1.5418$$

Differentiating, the slope resistance is:

$$r_{slope(\Omega)} = \frac{37.8}{I_{DC(mA)}}$$

This equation is significant not so much because it allows us to estimate a value for r_{slope}, but because it shows that r_{slope} is inversely proportional to I_{DC}, suggesting that LED bias is best suited in stages passing 5 mA. Nevertheless, it is possible to use LED bias in low current stages if additional current is passed through the LED, perhaps from a resistor connected to the HT, or from a constant current source to a lower voltage source. In this way, an ECC83 requiring 1.6 V but only passing 0.5 mA could be biassed by inserting an HLMP6000 in its cathode circuit and driving additional current through the LED. For a typical HLMP6000 red LED, the current needed to develop the required 1.6 V forward voltage can be found using:

$$i_{mA} = \exp(x)\left(\frac{v-1.5418}{0.0378}\right) = e\left(\frac{1.6-1.5418}{0.0.378}\right) = 4.66\ \text{mA}$$

We already have 0.5 mA from the ECC83, so we need an extra 4.16 mA from somewhere else. If we had a 285 V HT, then a 68 kΩ resistor would do, but it would dissipate 1.2 W, so a 4 W component would be needed. Nevertheless, when the choice is between a 3.2 kΩ cathode-bias resistor (needing a bypass capacitor) and an LED with a slope resistance of 5.7 Ω that doesn't need to be bypassed, then that 68 kΩ 4 W resistor suddenly starts making a lot more sense (see Figure 3.26).

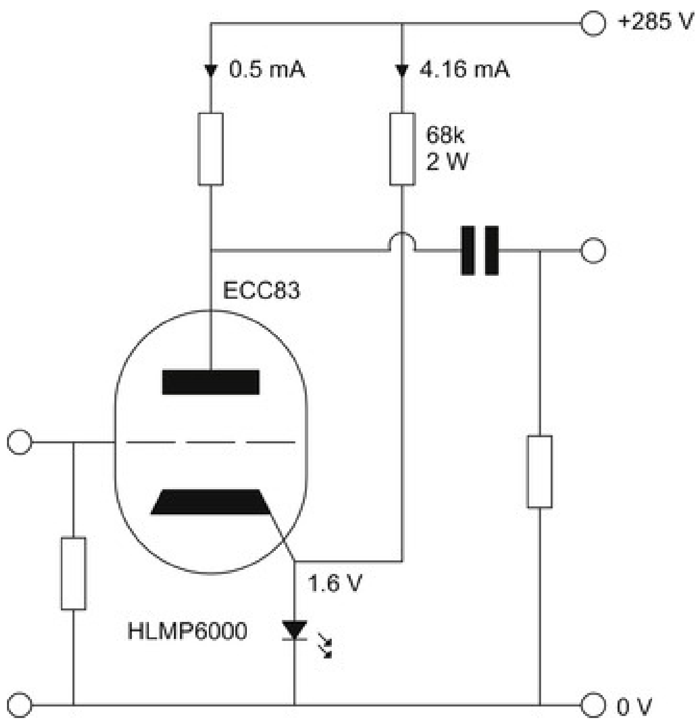


Figure 3.26 Cathode bias with LED using supplementary current.

Reverse bias generally produces more noise in a diode than forward bias, but enables higher reference voltages. Low voltage Zener diodes truly use Zener action, whereas higher voltage diodes actually use the avalanche effect. At 6.2 V, both effects are present, their opposing temperature coefficients cancel and r_{slope} is at a minimum, so 6.2 V Zeners are the most stable. If a stable high voltage reference is required, it is usually better to have a string of 6.2 V Zeners than a single high voltage Zener.

From a DC point of view, diode bias is ideal for the lower valve of a μ-follower or SRPP because I_a is stabilised by the bias arrangements of the upper valve.

Because $r_{slope} \neq 0$, a change in signal current causes a change in the voltage across the diode. The signal current also produces the voltage across R_L, so:

$$\frac{V_{r_{slope}}}{r_{slope}} = \frac{V_{R_L}}{R_L} = i_{signal}$$

Cross-multiplying:

$$V_{r_{slope}} = \frac{V_{R_L} \cdot r_{slope}}{R_L}$$

The significance of this equation is that we have just seen that r_{slope} is inversely proportional to applied current, and because r_{slope} varies with signal current, the signal voltage developed across it must be distorted. Unfortunately, this distorted signal voltage is in series with the input signal because the valve amplifies the *difference* in voltage between the grid and the cathode (see Figure 3.27).

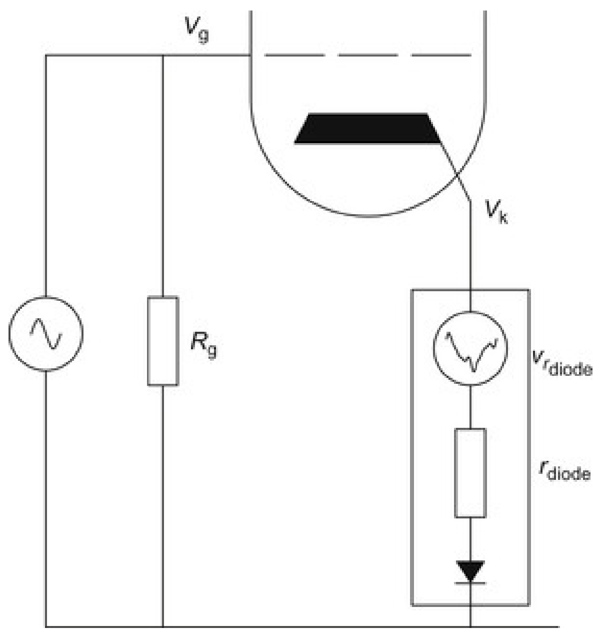


Figure 3.27 Non-linear diode internal resistance adds distortion in series with the source.

However, the equation and the diode curve show us that the distortion added by the diode can be reduced by:

- Avoiding diode bias for I_{diode}<5 mA (because r_{slope} is particularly variable at low currents).
- Minimising r_{slope} by diode choice (red LED).
- Maximising R_L.
- Reducing the output signal voltage V_{R_L}.

These conditions imply that diode bias is best suited to:

- *RIAA input stages*: I_a is high and signal levels are low. Additionally, the stage can recover instantly from clipping due to high voltages at high frequencies caused by dust, etc., on the record.
- *μ-Follower stages*: The active load maximises R_L, and I_a is likely to be high (minimising the variation in r_{slope}).

Constant Current Sink Bias

A constant current sink allows cathode current to be forced to the design value despite valve parameters. However, because a constant current sink is an open circuit to AC, it would cause 100% negative feedback in a single-ended stage, so it must be bypassed with a capacitor (see Figure 3.28).

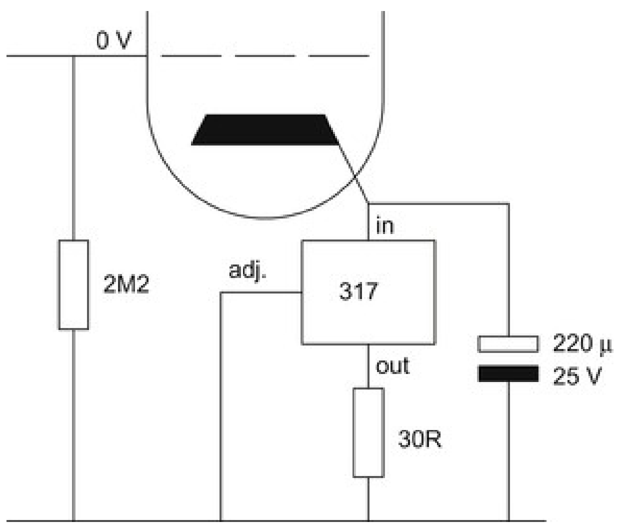


Figure 3.28 Cathode bias using a bypassed 317 constant current sink.

Once bypassed by the capacitor, the AC performance of the constant current sink becomes irrelevant, so a three-terminal regulator such as the 317 becomes perfectly acceptable, and this strategy is quite popular in output-valve cathodes. This is a very rare occasion when DC accuracy is more important than AC performance because accurate matching of DC currents in a push–pull output stage enables us to eliminate the magnetising current that would cause a toroidal output transformer's core to saturate and generate bass distortion. Note also that because the cathode current has been forced by the 317 CCS (and therefore cannot run away), the grid-leak resistor can become rather larger than the maximum specified by the valve datasheet, enabling a smaller coupling capacitor from the preceding stage.

Differential pairs and cathode followers require exemplary AC performance, and DC accuracy is generally a secondary consideration, which was why we took such care when investigating their design at the end of Chapter 2.

Individual Valve Choice

Although a set of anode characteristics having noticeably different spacings between the curves indicates distortion, evenly spaced characteristics do not guarantee low distortion. Ultimately, we must either use valves designed for low distortion, or test valves for distortion.

Which Valves Were Explicitly Designed to be Low Distortion?

Minimising distortion costs money, so when low-distortion valves were designed, they were targeted specifically at the audio market, which included the broadcast, recording and film industries and, of course, the consumer.

In the 1930s, gain was extremely expensive. The idea of deliberately throwing gain away (negative feedback) was treated as heresy, so much so that although Black's jotted notes were witnessed on 18 August 1927, his US patent [5] was not issued until 21 December 1937. As a consequence, low distortion was reliant on valve design and construction, so valves like the 76 were designed to be low distortion. As feedback became more widely accepted, it became cheaper to reduce distortion by sacrificing gain, so the final generation of valves had higher gain, but low distortion became less important.

Low-distortion valves were also required by the telecommunications companies, but not because they were concerned with the fidelity of baseband audio. If we need to provide 1,000 analogue telephone circuits between two cities 10 miles apart, we could lay 1,000 twisted pairs, but a cable containing this amount of wire is expensive and cumbersome to lay. The solution adopted by the telecommunications companies was to modulate each telephone circuit onto an RF carrier with its own frequency – just like different radio stations. One thousand modulated carriers could then be passed down a single (usually coaxial) cable which was cheap and easily laid. All cables introduce loss, and between cities the loss becomes significant, so each cable needed repeater amplifiers at regular distances. One of the many advantages of multiplexing 1,000 telephone circuits onto one cable was that only one repeater amplifier was needed every few miles instead of 1,000, reducing cost. However, any distortion in that amplifier would cause one telephone conversation to crosstalk onto another. Valves designed for use in broadband telephone repeater amplifiers

were therefore required to produce low distortion.
Many of the final generation of valves used a frame-grid, and some, such as the 417A/5842, were explicitly designed for low distortion. Other valves such as the ECC88/E88CC simply benefited from improved production engineering and produce usefully low distortion. Some valves such as the E182CC and 6350 were designed for use in early digital computers, where the most important consideration was long life even with full heater power and no anode current, which tempts the growth of cathode interface resistance.
Finally, some valves were designed and manufactured with a complete lack of regard for distortion.
The problem of field, or vertical, scanning in a television using a Cathode Ray Tube (CRT) was very similar to that of an audio amplifier driving a loudspeaker. Both use a transformer to couple to the driving valve, and the frequency range is similar. However, CRT scan coils must be driven by a controlled current, rather than applied voltage as is conventional for loudspeakers. Unfortunately, the finite (and changing) primary inductance L_p of the small iron-cored output transformer drew a current in addition to the scan coil current, and this meant that the current waveform drawn from the field scan valve was distorted compared with the ideal current required by the scan coils. There were many ways of achieving a compensating distortion, but one was to use the curvature of the I_a/V_a characteristic of a triode. Since transformer L_p was not tightly controlled, the required distortion had to be controllable, so a variable resistor was often inserted in the cathode circuit of the valve to allow adjustment of vertical display linearity.
The crux of the previous argument is that there was no requirement *whatsoever* for the valve manufacturers to produce field scan valves with outstanding or even consistent linearity, since this had to be individually adjusted for each television's CRT. Early field scan valves such as the dual triode 6BX7 show wide variations in distortion (4:1 between best and worst), so they have to be selected for audio use, and the probability of finding a pair of low-distortion valves in one envelope is low, so selecting a pair of low-distortion 6AH4 single triodes would be a much cheaper alternative. Later generation valves such as the ECC82 (also intended for use as a field scan oscillator) benefited from improved production techniques and distortion is extremely consistent from sample to sample; it is consistently poor.

Carbonising of Envelopes

Deketh [6] pointed out that not all electrons accelerated from the cathode/grid interface strike the anode – some miss and collide with the envelope, causing secondary emission. Secondary emission is important because it means that the envelope acquires a negative charge that can distort the flight of electrons from cathode to anode. Deketh considered distortion at high amplitudes in power valves and showed that carbonising the inside surface of the envelope was beneficial because it reduced secondary emission. At the time, nobody was worried about audio distortion at <1%, and Deketh might not have had access to an audio spectrum analyser, so he did not publish distortion results at lower levels. Nevertheless, this author's measurements at +28 dBu (≈19.5 V_{RMS}) show significantly reduced (≈−6 dB) distortion for samples of the 6SN7 having a carbonised envelope compared to clear envelopes.

Deflecting Electrons

Amplifying valves control the flight of electrons by imposing electric fields, but electrons can be deflected by magnetic fields. The Earth's magnetic field is quite weak, so it is unlikely that orienting a valve in any particular direction will affect distortion, but most sheet electrodes are made of nickel, which can easily be magnetised. If the valve was constructed from concentric cylindrical electrodes, magnetic deflection would not matter unless it caused electrons to miss the anode, but box constructions do not have radial symmetry, so horizontal magnetic deflection could influence anode current.

Beam tetrodes with aligned grids are the most susceptible to magnetic fields because vertical magnetic deflection could cause the sheets of electrons to intercept g_2 rather than passing cleanly between the vertically aligned windings. Thus, a magnetic field can change the I_a/I_{g2} ratio, and it would be foolish to suggest that this could not affect distortion. Some years ago, using a coil intended for degaussing television display tubes, the author jokingly degaussed the KT88 (aligned grid beam tetrodes) of a power amplifier, and everyone heard a slight difference.

We should be aware that degaussing requires the magnetic material to be taken to saturation in both directions and then gently taken through ever decreasing hysteresis loops until the residual magnetism is zero. Thus, magnetisation and demagnetisation are achieved by brute force – the author's degaussing coil is 10″ (250 mm) in diameter, consumes 750 VA and is rated only for intermittent use. Applying an audio signal to an amplifier cannot possibly achieve this effect, no matter how exotic the signal may be.

Testing to Find Low-Distortion Valves

Low-noise input stages demand high g_m, and signal levels are so low that distortion is not an issue. To minimise noise, well-designed circuitry amplifies low-level signals once only, and thereafter carries line-level signals.

When designing a power stage, the most important consideration is $P_{a(max)}$, and the consequent DC requirements force distortion to be quite low down on the list of priorities. Further, power stages are used once only to drive the load.

Because of the previous two arguments, low distortion valves are essential for line-level processing (because there is likely to be so much of it), but they need not have outstanding g_m or $P_{a(max)}$. High μ valves might be undesirable if their design assumed the use of negative feedback to reduce distortion. Sadly, most of the low- μ valves were designed for television field scan, so their distortion is distinctly questionable unless individually selected. The remaining valves are medium- μ and have $P_{a(max)}<5$ W.

The 6SN7 is widely accepted as a low-distortion valve, but how well does it justify its reputation? Bearing in mind that valves were assembled by hand and subject to wide production tolerances, is there a 'best' medium- μ valve or manufacturer? This section seeks to answer these questions by reporting on the testing of a selection of medium- μ valves under identical conditions.

The Test Circuit

If we require a low-distortion gain stage, this can be achieved by a single-ended stage with an active load, or a differential pair with resistive loads and constant current sink tail. In short, circuit design can reduce distortion, but only to the point where the valve's *irreducible distortion* takes effect. If we are to select valves for minimum distortion, we should focus on their irreducible distortion, since higher levels of distortion can always be reduced by suitable circuit design. Although this complicates matters by requiring us to measure distortion in a stage deliberately designed for low distortion, it has the benefit of enforcing a level playing field.

If we later use a topology that does not minimise distortion, and valve 'A' sounds better than valve 'B', it is because valve 'A' suits the topology better than valve 'B', *not* because valve 'A' is 'better' than valve 'B'.

As previously mentioned, distortion in a triode amplifier is dominated by the variation of r_a with I_a. Provided that $R_L >> r_a$, the variation of r_a is insignificant, so distortion can be reduced by maximising R_L. In addition, the valve should pass sufficient anode current to place its operating point well clear of the typical bunching of anode curves experienced at low currents. Accordingly, the valves were tested in a μ-follower circuit passing ≈8 mA anode current. In this configuration, the test valve sees $R_L \approx 800$ kΩ, which although far from infinite, can be put into perspective by realising that if a true 800 kΩ resistor were to be substituted, a 6.4 kV HT supply would be required. Heater filaments were fed from a stabilised DC supply adjusted for the correct voltage at the heater pins (see Figure 3.29).

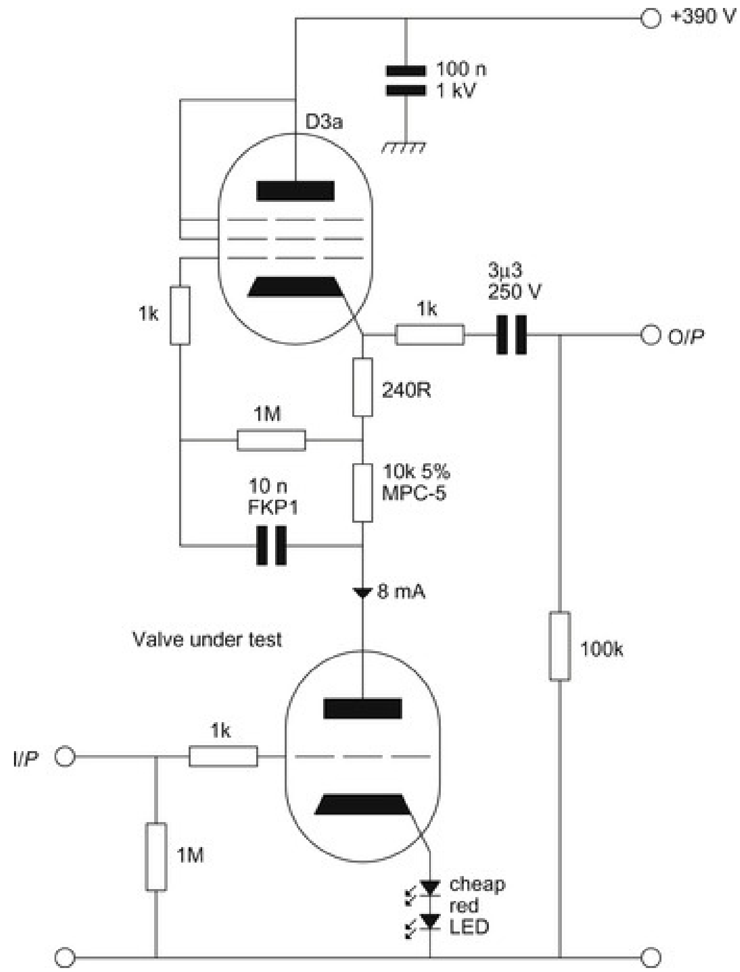


Figure 3.29 Medium- μ valve test circuit.

Audio Test Level and Frequency

Since distortion was expected to be low, the valves had to be tested at a sufficiently high output level to make the distortion easily measurable, yet well below clipping. +28 dBu (≈19.5 V_{RMS}) was found to be a good compromise, so each valve had its input level adjusted to produce precisely +28 dBu at its output. Distortion in all the reported valves is directly proportional to level, so distortion at lower levels can be extrapolated from the test data.

Although initially tested at 120 Hz, 1 kHz and 10 kHz, distortion of the test circuit was found to be completely independent of frequency, so the valves were tested at 1 kHz only. For most of the valves, harmonics beyond H6 were too close to the oscillator distortion residual for reliable measurement, so measurement was only attempted on harmonics up to and including H6.

Test Results

All of the valves tested were 'New Old Stock' (NOS), so the newest valves were at least 30 years old, and the oldest was 58. Since the valves have been out of production for decades, for some types only a few samples were available to the author.

Raw data from the measurements were analysed in a spreadsheet, and broken into different groups as and when significant differences became apparent.

Table 3.6 summarises the results of the 6SN7GT/12SN7GT and its direct equivalents. The number of samples refers to the number of individual triodes tested, not envelopes.

Table 3.6 Comparison of Distortion Harmonics Within the 6SN7 Family

Type	Samples	H2	1 σ	H3	1 σ	H4	1 σ
6SN7GT/12SN7GT	44	−50	3.6	−85	8.4	−96	5.9
7N7	82	−52	3.3	−85	8.6	−97	6.7
14N7	62	−52	3.3	−85	8.6	−97	6.7
Carbonised 6SN7GT	6	**−54**	1.8	**−94**	5.6	–	
Carbonised CV1988	12	**−57**	2.6	−85	7.2	−93	4.2
12SX7GT	12	−50	1.9	−83	3.2	−94	6.0
GEC/Marconi B36	6	−51	2.0	**−90**	8.1	−88	2.0
6J5GT (various)	6	−50	4.1	−82	12.7	−97	3.1
Pinnacle 6J5GT	138	−52	2.6	**−90**	6.7	−96	3.9
RCA 6J5	15	−47	4.8	−84	8.3	−89	7.7
GEC L63	5	−50	1.6	−86	4.4	−89	4.4
7A4	3	−48	0.2	−73	1.6	−93	1.2
1 σ=One standard deviation.							

Table 3.7 normalises distortion to the 6SN7GT/12SN7GT to enable clearer comparison.

Table 3.7 Normalised Comparison of Distortion Harmonics Within the 6SN7 Family

Type	Samples	H2		H3		H4	
		dB	Ratio	dB	Ratio	dB	Ratio
6SN7GT/12SN7GT	44	0	1	0	1	0	1
7N7	82	−2	0.79	0	1	−1	0.89
14N7	62	−2	0.79	0	1	−1	0.89
Carbonised 6SN7GT	6	− **4**	0.63	− **9**	0.35	− **14**	0.2

Carbonised CV1988	12	− 7	0.45	0	1	+3	1.4
12SX7GT	12	0	1	+2	1.26	+2	1.26
GEC/Marconi B36	6	0	1	− 5	0.56	+8	2.5
6J5GT (various)	6	0	1	+3	1.4	−1	0.89
Pinnacle 6J5GT	138	−2	0.79	− 5	0.56	0	1
RCA 6J5	15	+3	1.41	+1	1.12	+7	2.2
GEC L63	5	0	1	−1	0.89	−3	0.71
7A4	3	+2	1.26	+12	4	+3	1.4

Interpretation

The manufacturers claimed that all the preceding valves were electrically equivalent. Nevertheless, there are significant differences between the valves, and useful conclusions can be drawn from Table 3.7.

- Valves with carbonised glass envelopes produce less distortion. Deketh reported that carbonised envelopes reduced distortion at maximum power, but this series of tests suggests that the improvement is proportional to level, and that carbonised envelopes significantly reduce distortion at lower amplitudes.
- The RCA 6J5 (metal envelope) has significantly higher distortion than the 6J5GT, probably due to increased numbers of gas ions resulting from outgassing of the metal envelope causing increased grid current.
- Despite having a clear envelope, the (Russian-made) Pinnacle 6J5GT offers very low distortion – significantly better than any other manufacturer of 6J5GT.
- The Loktal™ base was specifically designed to reduce stray capacitances and inductances by eliminating the glass pinch required by the Octal base, hence the 6SN7GT/12SN7GT has a claimed $C_{ag} \approx 4$ pF, whereas the 7N7 has a claimed $C_{ag} = 3$ pF [7].
- Some valves were selected from the standard production line by their manufacturers. This test shows no significant difference in distortion for the 12SX7 (simply a 12SN7GT selected for transconductance [8] at $V_a = 28$ V) compared to the standard 6SN7GT/12SN7GT.
- The third harmonic distortion of the Loktal™ 7A4 single triode is very disappointing, but as only three samples were available for testing, the results are probably not statistically significant even though the standard deviation was only 1.6 dB.
- The *SN7GT family is available with four different heater filament constructions [9], so Table 3.8 compares the different types.

Table 3.8 Heater Voltages and Currents Within the *SN7 Family

	Voltage (V)	Current (mA)	Heating power (W)
6SN7GT	6.3	600	3.78
8SN7GT	8.4	450	3.78
12SN7GT	12.6	300	3.78
25SN7GT	25	150	3.75

As can be seen from Table 3.8, the heater power is almost identical for all types. Physically, the 6SN7GT has its heaters wired internally in parallel, whereas the 12SN7GT sometimes has them wired in series, but electrode construction is identical, so distortion ought to be similar, so Table 3.9 compares distortion of the 6SN7GT with that of the 12SN7GT.

Table 3.9 Comparison of Distortion Harmonics Between 6SN7 and 12SN7

	Samples	H2	1 σ	H3	1 σ	H4	1 σ
6SN7GT	28	−50	3.5	−83	8.9	−96	5.7
12SN7GT	16	−51	3.8	−87	7.3	−97	6.5
The two valves are very similar – the differences are well within the margins of uncertainty.							

Similarly, we can compare the 7N7 with the 14N7 (Table 3.10).

Table 3.10 Heater Voltages and Current Within the Loctal Family

	Voltage (V)	Current (mA)	Heating power (W)
7N7	6.3	600	3.78
14N7	12.6	300	3.78

Again, we would expect the distortion between the two types to be similar (Table 3.11).

Table 3.11 Comparison of Distortion Harmonics Between 7N7 and 14N7

	Samples	H2	1 σ	H3	1 σ	H4	1 σ
7N7	82	−52	3.3	−85	8.6	−97	6.7
14N7	62	−52	2.4	−88	7.8	−95	6.4

Summarising, the differences between valves having different heater voltages are well within the margins of uncertainty. This is good news because it means that we don't have to use the expensive 6.3 V heater valves, but can use the cheaper and more plentiful 12.6 V heater valves and enjoy the reduction in electromagnetic hum caused by their reduced heater current, although the higher heater voltage causes commensurately more electrostatic hum. The significance of trading electromagnetic hum for electrostatic hum is that although both types can be shielded, earthed aluminium cooking foil is an effective electrostatic shield, whereas effective electromagnetic shielding requires a far greater thickness of iron or expensive mu-metal. Naturally, both problems disappear if we use regulated DC heaters.

A Convention

The author has adopted a convention that will be applied throughout this book. Having established that the 6J5GT, 6SN7GT, 12SN7GT, 7N7 and 14N7 are electrically virtually identical, and that the 8SN7GT and 25SN7GT can be expected to be similar, economy of nomenclature is necessary. From now on, this family will be called the *SN7/*N7 family (to distinguish between the bases and capacitances). Please note that the 6N7 is a *completely* different beast, and bears no relation to a 7N7.

Alternative Medium- μ Valves

Table 3.12 shows possible alternatives to the *SN7/*N7 family.

Table 3.12 Electrical Alternatives to the *SN7/*N7 Family

Type	μ	Samples	H2	1 σ	H3	1 σ	H4	1 σ	H5	1 σ	H6	1 σ
7AF7	16	4	−38	0.3	−62	1.5	−74	0.6	−89	4.2	−91	5.7
ECC82/12AU7/B329	18	28	−37		−56	1.4	−73	3.9	−86	6.6	−96	3.1
E182CC/7199	18	30	−45	1.7	−70	1.5	−92	3.7				
E288CC	20	14	−49	1.3	−69	0.9	−89	5.4	−95	7.2	−96	4.9
37	9	9	−45	0.6	−69	4.9	−87	5.7	−88	10.1	−86	14.2
5687 (various)	16	22	−49	1.1	−72	1.7	−91	3.9				
Philips 5687WB	16	14	−42	2.5	−68	2.8	−92	2.4				
6350	20	26	−44	1.4	−65	2.4	−84	2.4	−98	6.2		
For each valve, μ was read from the manufacturer's anode curves at the 8 mA operating point.												

Table 3.13 allows quick comparison of these alternatives by normalising their distortion to the 6SN7GT/12SN7GT.

Table 3.13 Distortion (normalised to 6SN7GT/12SN7GT) of Electrical Alternatives to the *SN7/*N7 Family

Type	Samples	H2		H3		H4	
		dB	Ratio	dB	Ratio	dB	Ratio
6SN7GT/12SN7GT	44	0	1	0	1	0	1
7AF7	4	+12	4	+23	14	+23	14
ECC82/12AU7/B329	28	+13	4.5	+29	28	+23	14
E182CC/7199	30	+5	1.78	+15	5.6	+4	1.58
E288CC	14	+1	1.12	+16	6.3	+7	2.2
37	9	+5	1.78	+16	6.3	+9	2.82
5687 (various)	22	+1	1.12	+13	4.5	+5	1.78
Philips 5687WB	14	+8	2.5	+17	7.1	+4	1.58
6350	26	+6	2	+20	10	+12	4

The results in Table 3.13 speak for themselves. All of the alternatives are inferior to the 6SN7GT/12SN7GT and produce significantly more H3.

The Loctal 7AF7 dual triode and B9A ECC82 dual triode are particularly ghastly. Perhaps significantly, these valves have an electrode structure that significantly reduced C_{ag} compared to the *SN7/*N7 family (2.3 pF and 1.6 pF versus 4.0 pF). These tests suggest that the measures necessary to reduce C_{ag}

within the electrode structure might adversely affect distortion.

There was a significant difference between the Philips 5687WB and other manufacturers' samples, so this type was isolated. Although H2 and H3 are significantly higher than competing manufacturers, H2 would mostly be nulled if used in a differential pair.

Weighted-Distortion Results

Towards the beginning of this chapter, distortion weighting was suggested as a useful technique, so Table 3.14 mathematically weights the measurements according to CCIR468-2 with a gain offset of −6 dB in order to make the H2 result comparable with an unweighted measurement. Because this particular test was limited to harmonics up to H6 (6.3 kHz is the corner frequency for CCIR468-2) and the distortion was dominated by H2 harmonic, the difference between CCIR468-2 less 6 dB and the Shorter recommendation was only ≈0.1 dB – well below measurement uncertainties, and validating Peter Skirrow's CCIR468-2 single-measurement argument.

Table 3.14 Weighted Distortion Comparison of All the Valves Tested

<table>
<tr><th>Type</th><th>Number of samples</th><th>Weighted distortion (dB)</th></tr>
<tr><td>Carbonised CV1988</td><td>12</td><td>−58</td></tr>
<tr><td>Carbonised 6SN7GT</td><td>6</td><td>−55</td></tr>
<tr><td>Pinnacle 6J5GT</td><td>138</td><td rowspan="2">−52</td></tr>
<tr><td>7N7/14N7</td><td>144</td></tr>
<tr><td>GEC/Marconi B36</td><td>6</td><td>−51</td></tr>
<tr><td>6SN7GT/12SN7GT</td><td>44</td><td rowspan="3">−50</td></tr>
<tr><td>12SX7GT</td><td>12</td></tr>
<tr><td>6J5GT (not Pinnacle)</td><td>6</td></tr>
<tr><td>L63</td><td>5</td><td rowspan="3">−49</td></tr>
<tr><td>E288CC</td><td>14</td></tr>
<tr><td>5687 (not Philips)</td><td>22</td></tr>
<tr><td>7A4</td><td>3</td><td rowspan="2">−48</td></tr>
<tr><td>RCA 6J5</td><td>15</td></tr>
<tr><td>E182CC/7199</td><td>30</td><td>−45</td></tr>
<tr><td>6350</td><td>26</td><td>−44</td></tr>
<tr><td>Philips 5687WB</td><td>14</td><td>−42</td></tr>
<tr><td>7AF7</td><td>4</td><td>−38</td></tr>
<tr><td>ECC82, 12AU7, B329</td><td>28</td><td>−36</td></tr>
</table>

Overall Conclusions

A total of 529 valves was tested, and the results show that the reputation of the *SN7/*N7 family is well justified. Distortion for the dual triodes varied from sample to sample, with a few significant trends visible between manufacturers (probably because manufacturers were well known for buying in another manufacturer's valves and branding them as their own in times of short supply). If individual measurement and selection for low distortion are not possible, then carbonised envelope valves from the *SN7*/N7 family are likely to produce the lowest distortion. If these are not available, then a 7N7, 14N7 or a Pinnacle 6J5GT is likely to be a good choice. Valves with electrode structures shrunk to fit B9A bases are significantly poorer.

Coupling from One Stage to the Next

The most common way of coupling one stage to the next is via a capacitor. A perfect capacitor does not generate distortion. Unfortunately, even a perfect capacitor can greatly exacerbate the distortion generated by valves or transistors.

Blocking

Blocking is an extremely unpleasant mechanism whereby an amplifier mutes for a short time after a momentary overload. In simple terms, blocking is caused by the capacitor that couples to a stage that is overloaded (see Figure 3.30).

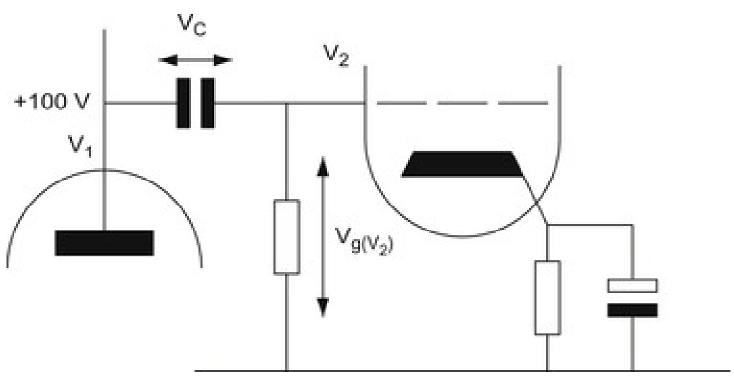


Figure 3.30 Capacitor coupling and blocking.

Capacitor coupling between two stages forms a high-pass filter. In order not to affect the audio, we deliberately design for a low $f_{-3\ \mathrm{dB}}$ frequency using:

$$f_{-3\ \mathrm{dB}} = \frac{1}{2\pi CR};\quad \text{alternatively,}\ f_{-3\ \mathrm{dB}} = \frac{1}{2\pi\tau}$$

Thus, setting $f_{-3\ \mathrm{dB}}=1$ Hz means that $\tau=160$ ms, but the true significance of τ in this context will take a little time to emerge.

Because of Kirchhoff's voltage law, the following must be true at *all* times:

$$V_C = V_{a(V_1)} - V_{g(V_2)}$$

Example: $V_{a(V_1)} = 100\ \mathrm{V}$, and the following grid is tied to ground via a grid-leak resistor, so $V_{g(V_2)} = 0\ \mathrm{V}$, causing $V_C=100$ V.

If we apply an impulse to V_1 so that the anode swings 20 V positively, the grid of V_2 attempts to rise to +20 V, but when it reaches +10 V, V_{gk}=0, and the grid conducts so heavily to the cathode that its voltage is clamped to +10 V. At this instant, the previous Kirchoff equation must *still* be true, so:

$$V_C = 120\ \mathrm{V} - +10\ \mathrm{V} = 110\ \mathrm{V}$$

The capacitor was able to change its voltage almost instantaneously because it charged through the low impedance path of the overloaded grid. When the impulse passes, we can find the grid voltage of the second valve by rearranging the equation:

$$V_{g(V_2)} = V_{a(V_1)} - V_C$$
$$V_{g(V_2)} = 100\ \mathrm{V} - 110\ \mathrm{V} = -10\ \mathrm{V}$$

The grid is at −10 V, but the cathode is held at +10 V by the cathode bypass capacitor, so V_{gk}=−20 V, and the grid has reverted to a high impedance path. More importantly, the valve is now *switched off* and remains so until the grid recovers to 0 V. The only path for the capacitor to change its charge is through the grid-leak resistor, but as we saw earlier, this combination has a time constant of 160 ms. Worse, 5 τ is required for the capacitor to change its charge to within 99% of its final charge, so the grid does not reach 0 V until 0.8 s after the momentary overload.

Recovery from overload is complicated by the fact that once the valve is switched off, there is no cathode current, so the cathode bypass capacitor begins to discharge through the cathode-bias resistor. Although this causes the valve to begin conducting earlier than would otherwise have been expected, it also has to recover from the change imposed by blocking. Thus, a momentary overload has had its effects extended to almost a second.

It might be thought that the severe overload posited to cause blocking is unlikely, but applying global feedback around a power amplifier containing a capacitor-coupled output stage almost guarantees blocking. Suppose that a transient causes clipping of the output stage. Feedback attempts to correct this distortion of the waveform by increasing substantially the drive to the output stage, and the requirements for blocking have been satisfied.

Blocking occurred because a coupling capacitor was allowed to change its charge significantly during an overload. If the capacitor could be eliminated, or moved so that it coupled to a stage that could not be overloaded, the problem would not arise. This technique will be explored in the non-blocking ‘Crystal Palace’ amplifier described in Chapter 6.

Alternatively, the time constant of the coupling capacitor could be reduced, reducing the duration of blocking. The price paid for this solution is low frequency attenuation in a zero global feedback amplifier or increased low frequency distortion in a feedback amplifier. However, if such an amplifier were part of a loudspeaker system having an active crossover (valve treble, semiconductor bass), degraded low frequency performance would be irrelevant. Stuart Yaniger's 'Red Light District' amplifier available at the diyAudio website is a good example of this design philosophy.

Transformer Coupling

Transformers are expensive, but they are essential for connecting loudspeakers to valve amplifiers unless we are prepared to tolerate appallingly low efficiency. Inter-stage transformers offer some unique advantages:

- From the point of view of the primary, if the transformer is used as the anode load, the valve can achieve a much larger signal swing because the anode can theoretically swing to double the HT voltage if necessary. Alternatively, from a loadline point of view, the HT voltage has doubled, yet the signal swing has remained constant, typically halving the distortion compared to a resistive anode load.
- If the transformer steps down by a ratio of 2:1, the stage can produce the same swing as a resistively loaded stage, but with a quarter of the output resistance.
- A push–pull stage allows cancellation of even harmonic distortion.
- From the point of view of the secondary, a centre-tapped winding provides ideal phase-splitting.
- Larger power valves need low grid-leak resistances because of their grid current, so the very low R_{DC} of the secondary is ideal.

Against these advantages we have to weigh the inescapable fact that inter-stage transformers suffer most from transformer imperfections because they operate at high impedances, and single-ended stages force the transformer to pass the (DC) anode current, requiring a gapped core, which reduces bandwidth.

Low Frequency Step Networks

One possibility sometimes seen in power amplifiers is to bypass the coupling capacitor with a resistor to form a step network. Low frequency step networks are best inserted just before a differential pair and may even ease their tail DC arrangements (see Figure 3.31).

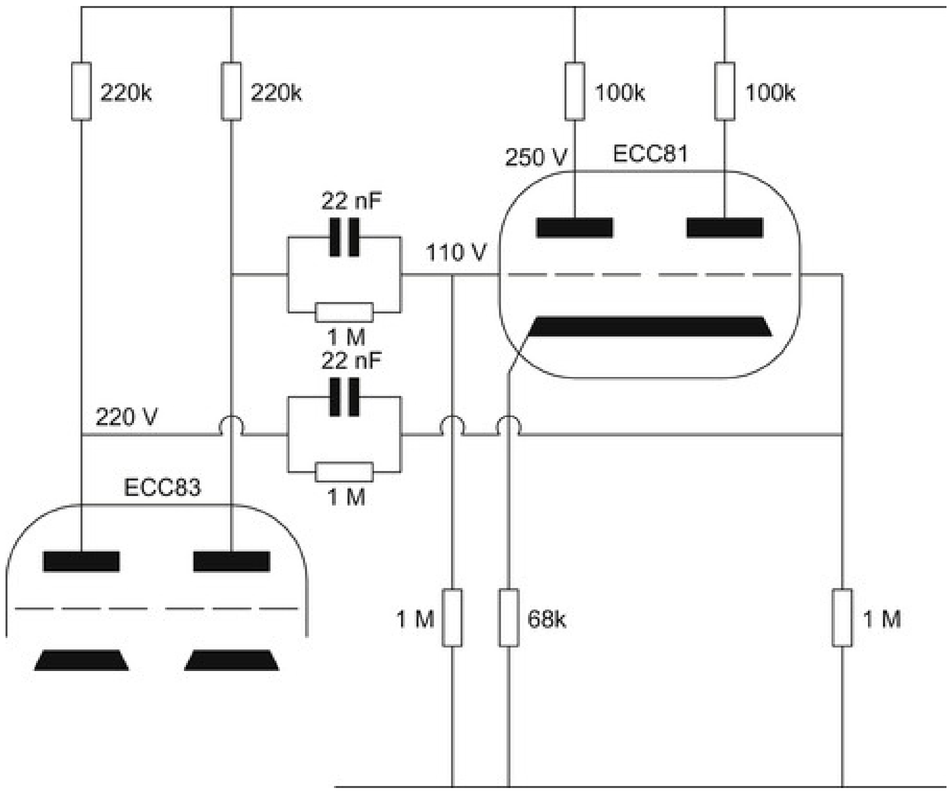


Figure 3.31 A step network to reduce low frequency phase shift.

Note that because 110 V has been applied to the second valve's grids, a simple resistor suffices for cathode bias, although performance would be improved if it were to be replaced by a CCS.

A step network is intermediate between DC and AC coupling. In the example, the step network is a 6 dB DC potential divider bypassed above 15 Hz by the 22 nF capacitors. The significance is that careful design can produce less low-frequency phase shift with a smaller capacitor than pure AC coupling and because reduced low frequency phase shift improves stability, more global feedback can be applied, reducing distortion. Further, blocking becomes less of a problem because the offending capacitor has a discharge resistor directly across it. Unless another low frequency time constant is dominant, step network time constants generally have to be determined by experiment because they are dependent on the (non-constant) primary inductance of the output transformer.

Level Shifting and DC Coupling

In the absence of PNP valves, DC coupled valve amplifiers rely fundamentally on the potential divider (see Figure 3.32).

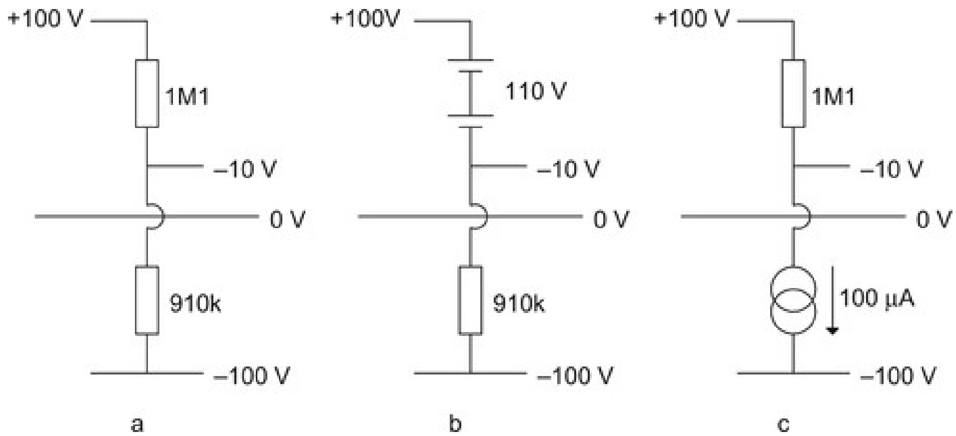


Figure 3.32 The three potential divider arrangements that allow DC coupling.

In Figure 3.32a, we have a simple resistive potential divider to the negative HT. We want to achieve −10 V at the output of the potential divider. Rather than use the potential divider equation, it is easier to set a current through the potential divider and apply Ohm's law to find the required resistances. Our potential divider will steal current from the anode of the preceding valve, so we should minimise this current. If we set the potential divider current to 100 µA, the upper resistor must be:

$$R = \frac{V}{I} = \frac{100\ \text{V} - (-10\ \text{V})}{100\ \mu\text{A}} = \frac{110\ \text{V}}{100\ \mu\text{A}} = 1.1\ \text{M}\Omega$$

Similarly, the lower resistor must be:

$$R = \frac{V}{I} = \frac{-10\ \text{V} - (-100\ \text{V})}{100\ \mu\text{A}} = \frac{90\ \text{V}}{100\ \mu\text{A}} = 900\ \text{k}\Omega$$

The nearest value of 910 kΩ is fine. Unfortunately, not only have we *level shifted* our signal by the required amount, we have also attenuated it. Thinking in AC terms:

$$\frac{V_{\text{out}}}{V} = \frac{910\text{k}}{910\text{k} + 1\text{M}1} = 0.453 = -6.9\ \text{dB}$$

Pure resistive level shifters inevitably attenuate the wanted AC, and this is the price that must be paid for simple DC coupling. A capacitor across the upper resistor would avoid the AC attenuation but create a step network which might not be desirable.

To maintain DC coupling without AC attenuation, we could replace the upper resistor with a battery, to make a Thévenin level shifter (see Figure 3.32b).

Because the battery is a perfect Thévenin source, it is a short circuit to AC, so this level shifter does not attenuate AC. Because 110 V batteries are inconveniently large, we would replace the battery with a Zener diode or neon reference valve. Sadly, both devices must pass a significant DC standing current (typically 5 mA), which makes them awkward to use. Worse, they both add noise.

Our final possibility is to replace the lower resistor with a constant current sink, to make a Norton level shifter (see Figure 3.32c).

There is no reason why we should not make a constant current sink out of bipolar transistors or a pentode. Provided that the sink has $r_{out} >> R_{upper}$, the Norton level shifter does not attenuate AC. However, the noise problem remains. Pentodes and transistors are transconductance amplifiers, which means that they convert their input voltage into an output current. A current sink amplifies its DC reference voltage, and we propose to convert its output current back into a voltage using a high value resistor. In effect, we have constructed a high-gain amplifier that amplifies the noise voltage of the DC reference.

Although Thévenin and Norton level shifters are theoretically usable, their practical problem is noise, and almost all the design effort must focus on reducing this noise to acceptable levels. (Constant current sink tails in differential pairs do not add significant noise because the low impedance load r_k severely reduces their gain, and their noise is applied common mode, so it is mostly rejected.)

Sadly, each of these techniques connects the signal to the negative HT, which may add hum and noise to the wanted signal.

A DC Coupled Class A Electromagnetic Headphone Amplifier

As with all amplifiers facing a difficult load, this circuit was designed working backwards from the output to the input. Be aware that the circuit is presented as a vehicle for investigating the problems of DC coupling – *not* as a detailed exemplar of headphone amplifier design.

A cathode follower is needed to provide a low output resistance. High- μ, high-g_m valves are best suited as cathode followers because the high g_m ensures low r_{out}, and high μ allows plenty of feedback to reduce distortion, so the 6545P is ideal. The cathode is connected to the negative HT via a standard pentode constant current sink using an EL822.

We know that we want to apply DC feedback to stabilise the output of the amplifier at 0 V, and we need our completed amplifier to have a high input resistance so that it does not load volume controls. The ideal input stage is therefore a differential pair, and because we already have a negative HT, it seems churlish not to use another pentode constant current sink (see Figure 3.33).

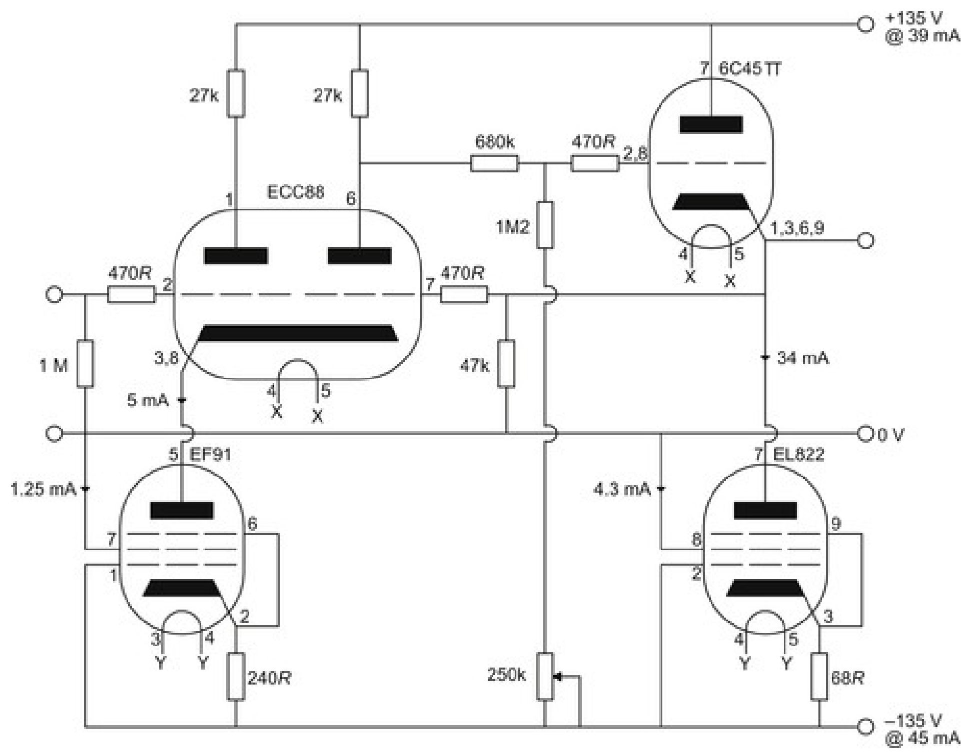


Figure 3.33 A DC-couped amplifier using a potential divider to the negative HT.

Electromagnetic headphones are low impedance devices. Because portable equipment must operate from a 3 V battery (possibly only 1.5 V), headphones designed for portable use are typically 32 Ω, but better quality designs tend towards 200 Ω. Either way, they require considerable current, and are extremely onerous loads for valves. As we increase I_a, P_a rises, so we must reduce P_a by lowering V_a as far as we dare without running into grid current. Setting V_a=135 V keeps us just clear of grid current. Although it is claimed that $P_{a(max)}$=7.8 W, all the other 6C45_ specifications are somewhat optimistic, and the envelope is little larger than an ECC88, so it does not seem wise to operate at the full claimed power. If we set I_a=34 mA, then P_a=4.6 W.

We now need to consider the input differential pair. Because we only have 135 V of HT, we need a valve having good linearity at low voltages. The ECC86 would be ideal, but the ECC88 was available, and a little slithering of loadlines suggested that 27 kΩ load resistors would work well with an anode voltage of 68 V, resulting in a tail current of 5 mA at V_{gk}=2 V.

We are finally in a position to be able to investigate the more significant problems of DC coupling using a pure potential divider.

Each ECC88 anode is at 68 V. The grid of the 6545P cathode follower is at ≈−1.5 V (remember that we need the cathode to be at 0 V), so we must drop 69.5 V. If we were to set 100 μA of potential divider chain current, we would need an upper resistor of 695 kΩ, which is an awkward value. If we choose the nearest preferred value of 680 kΩ, our potential divider current becomes 69.5 V/680 kΩ=102.2 μA.

The lower resistor must drop from −1.5 V to −135 V=133.5 V, so 102.2 μA requires a 1.3 MΩ resistor, which *is* a preferred value. However, we need to consider another factor.

At the calculated anode voltage, the potential divider chain steals ≈100 μA current from the anode current of 2.5 mA. 2.5 mA is a low current for this valve, so sample variations between valves are greatly magnified, and the small stolen current is negligible by comparison. The theft/valve error pulls the output DC in the final circuit away from 0 V, so why not make the divider chain adjustable to compensate for valve variation? Using a 1.2 MΩ fixed resistor in series with a 250 kΩ resistor allows ±10% variation. You could apportion a larger value of variable resistance, but this would make the output DC adjustment more fiddly.

Even if the output DC is carefully adjusted to 0 V, it will drift. We need a means of forcing the output DC to 0 V, and the best way of doing this is to apply negative feedback. We connect our feedback network in *parallel* with the output of the amplifier, but because the output of the feedback potential divider is connected to the other grid of the differential pair, and the differential pair amplifies the *difference* between its inputs, it is in *series* with the input signal. The feedback is, therefore, parallel-derived, series-applied, so it reduces output resistance and increases input resistance. Reduced output resistance is significant because all contemporary electromagnetic transducers rely on r_{source}=0 to produce their designer's intended transient response.

If we are to apply feedback, we must know how much gain is available. 27 kΩ is quite a low anode load resistance for an ECC88, and a loadline predicts a gain of 26.75. The input stage operates as a differential pair, but because only one output is used, we must immediately halve the gain to 13.375. We incur a loss of 0.657 in the level shifter, which reduces the gain to 8.78. Considering the 6545P, at I_a=34 mA, r_k≈25 Ω, so a 32 Ω load incurs a further loss of 0.56, leaving a total gain of ≈5. Thus, even 100% global feedback could produce only $(1+\beta A_0)$=5, or 14 dB, of improvement.

The circuit was tested with 100% negative feedback because this is the most critical condition for stability, yet delivered an exemplary square wave response at 10 kHz into a 200 Ω load (see Figure 3.34).

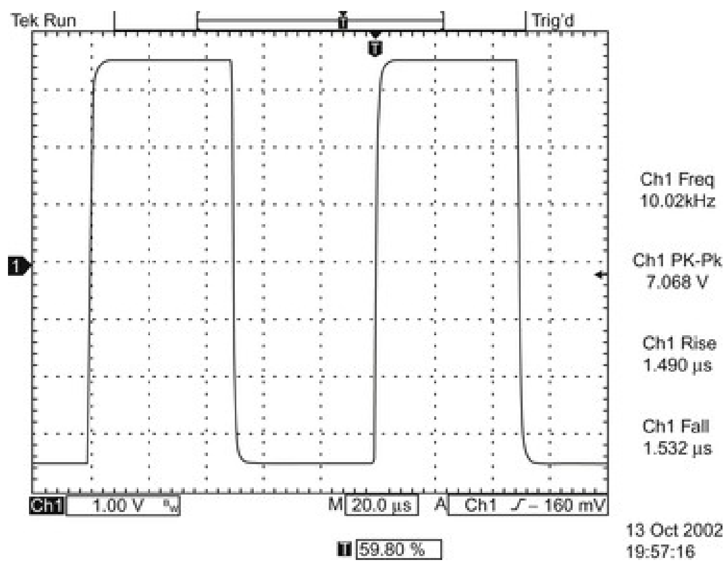


Figure 3.34 Near-perfect 10 kHz square wave response.

The circuit was tested into various load resistances.

Unsurprisingly, Table 3.15 shows that little valves don't like 32 Ω loads and that this circuit can only deliver 8 mW at 0.5% THD+N into a 32 Ω load. The performance improves markedly into a 200 Ω load, whereupon the circuit delivers double the power at <0.1% THD+N and, more importantly, the spectrum of the distortion harmonics became acceptable: H2=−60 dB, H3=−82 dB, H4=−100 dB. Curiously, the author still does not own any electromagnetic headphones of plausible quality, so he was unable to test this amplifier subjectively.

Table 3.15 Headphone Amplifier Distortion Against Load Resistance

	Output in dBu for specified THD+N		
	0.5%	**0.2%**	**0.1%**
100 kΩ	–	–	+20
200 Ω	+16.7	+12.8	+7.8
32 Ω	−3.7	–	–

Eagle-eyed readers having a 6C45_ datasheet will notice that this circuit far exceeds the maximum specified grid-leak resistance of 150 kΩ. Remember that the purpose of the grid-leak resistor is to hold the grid at its correct voltage in the face of grid current. If the resistor is too large, grid current raises the voltage on the grid, reducing V_{gk}, increasing I_a, causing more grid current, until the valve dies. In this circuit, I_a for the 6545P is set purely by the constant current sink, so considerations of thermal runaway caused by excessive grid-leak resistance are irrelevant.

If you need to set up a circuit of this type, it is easiest to break it into two parts before applying feedback. Build the output stage first, short-circuit the upper grid to ground, and adjust the CCS current programming resistor to set correct cathode follower current. Build the differential pair and associated CCS, and adjust the CCS current programming resistor to set required anode conditions. Connect the two stages, and fine-tune the level shifter for 0 V at the output. Finally, close the negative feedback loop.

Using a Norton Level Shifter

As mentioned previously, the Norton level shifter is a high-gain amplifier that amplifies the noise of its reference voltage. The noise problem can be tackled in various ways:

- Reduce the noise produced by the reference. Forward-biassed diodes produce less noise, so red LEDs may be useful. If a reverse biassed higher voltage diode must be used, use 5.6 V because this is a true Zener and quieter than higher voltages (which are avalanche).
- Noise isn't a problem in itself; it becomes a problem when the signal voltage is sufficiently low that the signal-to-noise ratio is compromised. Solution: Don't use Norton level shifters in pre-amplifiers.
- If the noise can be arranged to become common mode, then it can be rejected by a differential pair. This is the most powerful individual technique.

Don't try to obtain all the noise advantage you need from one single technique – it is always better to use a combination of small advantages that sum to a large advantage.

Some years ago, the author picked up 40 N-channel and P-channel MOSFETs from a junk shop, and after passing them through his late lamented curve tracer, he found two reasonably complementary pairs. The hybrid idea promptly resurfaced, coupled with a wish to use some of the more obscure valves in stock. The author had previously been unable to find a use for the wonderfully high g_m (55 mA/V) of the E55L, but realised that it might make a good cathode follower for driving the high-gate capacitance of MOSFETs. A little doodling resulted in a circuit requiring a Norton level shifter (see Figure 3.35).

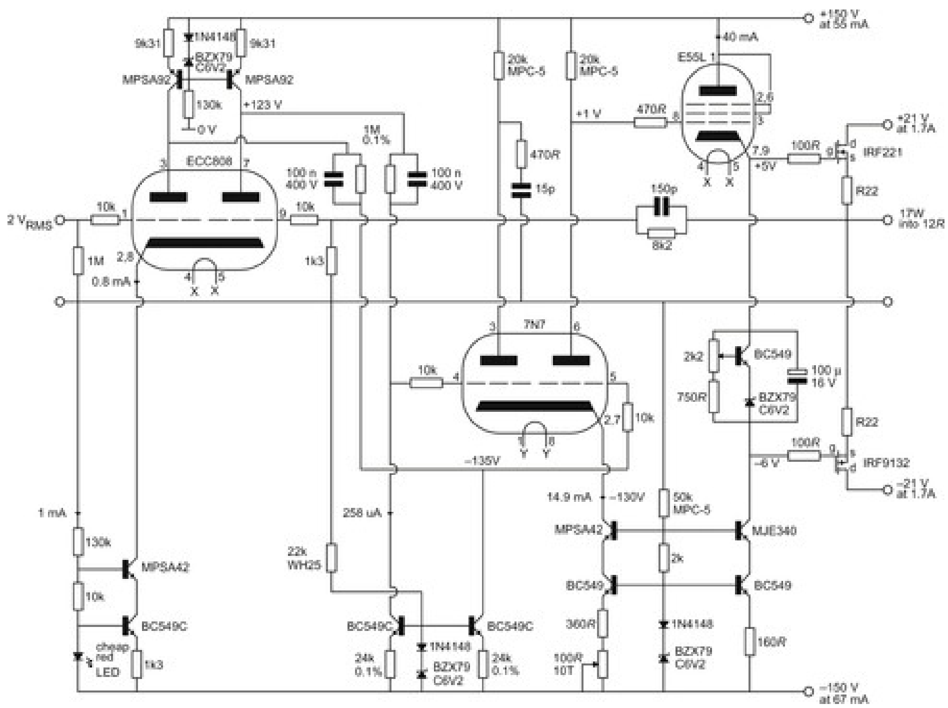


Figure 3.35 DC coupling using a Norton level shifter.

The N-channel MOSFET needs +5 V on its gate to pass the required current of 1.7 A, whilst the P-channel MOSFET requires −6 V, so the V_{be} multiplier between the gates allows these voltages to be applied and output stage current to be set. The E55L cathode follower has a power cascode constant current sink made from an MJE340 and a BC549 as its active load. The 7N7 differential pair has a cascode constant current sink tail that shares the reference voltage of the sink for the E55L. To balance the anode loads, the unused output of the 7N7 has an RC network to ground to simulate the input impedance of the E55L cathode follower. The ECC808 input differential pair is reasonably conventional except that it has constant current source anode loads to improve its linearity when operating from a positive HT of only +150 V.

Because the 7N7 differential pair is directly coupled to the E55L cathode follower, its grids must be at, or close to, the voltage of the negative HT, yet the anodes of the ECC808 are expected to be at +123 V. Thus, the problem is to DC couple the two stages with a minimum of noise.

Arbitrarily selecting 1 MΩ for the upper resistor of the level shifter suggests that the current passing through it will be ≈250 µA. We need to know this current because it is the design current for the constant current sink. If we choose a 6.2 V reference voltage, a 24 kΩ is required. If we treat the level shifter as a common-emitter amplifier, we can find its gain. I_c=250 µA, so g_m=35 I_c=35×0.25=8.75 mA/V. The gain of the amplifier is A_v= g_m·

R_L=8.75×1,000=8,750. However, the amplifier has considerable feedback because of the unbypassed 24 kΩ emitter resistor, so we can use the feedback equation:

$$A_v = \frac{A_0}{1 + (R_e/R_C) \cdot A_0} = \frac{8{,}750}{1 + (24/1{,}000)8{,}750} = 41.5$$

Alternatively, now that we know that the gain before feedback is likely to be large, we can simply use the approximation:

$$A_v = \frac{R_C}{R_e} = \frac{1{,}000}{24} = 41.7$$

The significance of this exercise is that a larger reference voltage reduces the gain of the amplifier because of the high value of R_e that it enforces. Although a 1.6 V red LED would be perhaps 9 dB quieter than a 5.6 V Zener, it would enforce 14 dB more gain in the CCS, making it 5 dB noisier in this example.

If the level shifters share the (potentially noisy) DC reference, its noise is amplified identically, so it is presented to the differential pair as common mode noise, which it can reject. To ensure that the noise remains common mode, the emitter resistors and 1 MΩ must be matched, so 0.1% tolerance types are necessary. At high frequencies, the common-mode rejection of the differential pair deteriorates, but if we bypass the 1 MΩ resistors with capacitors to form a step network, the level shifter's noise attenuation rises with frequency, tending to compensate for the differential pair's falling CMRR.

Distortion and Negative Feedback

Negative feedback reduces distortion. All engineers agree on this, but shouting begins immediately afterwards regarding the character and subjective effect of the new distortion. Baxandall [10] investigated the problem in 1978 and originated the graph that has been reproduced elsewhere with little or no explanation (see Figure 3.36).

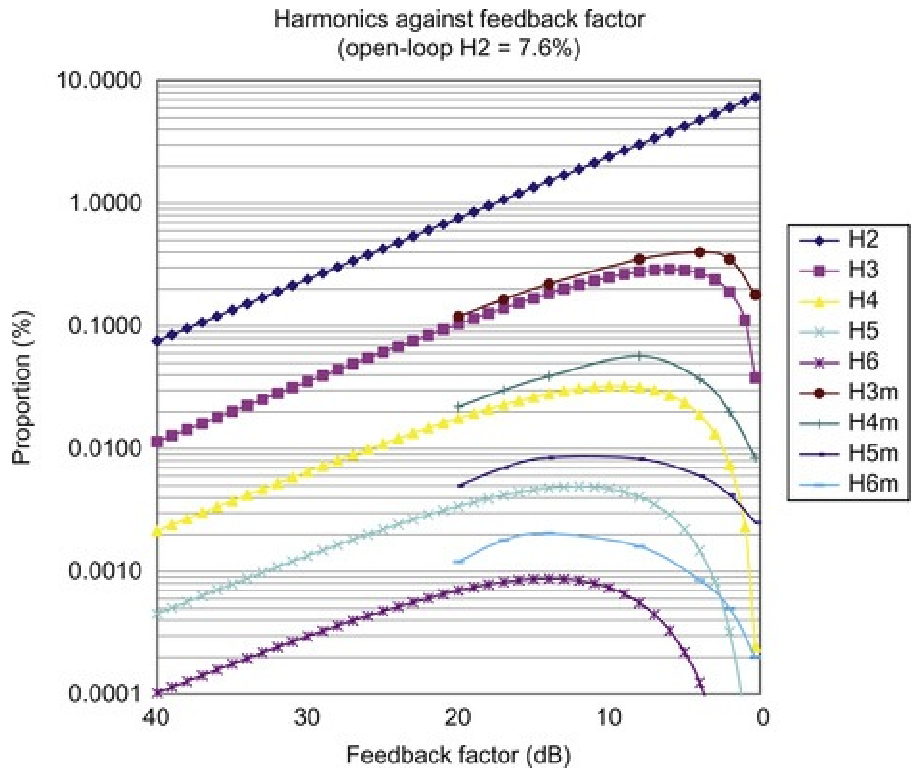


Figure 3.36 The ubiquitous Baxandall feedback and distortion graph.

Baxandall's original graph tested the hypothesis that if negative feedback was applied to a distorting amplifier, then the feedback signal would itself be distorted, generating harmonics of distortion harmonics, and that the amplitudes of these new harmonics could be predicted if the amplifier's transfer characteristic was simple and known. The chosen amplifier was a resistively loaded 2N5458 common source JFET because theory indicated that a JFET's distortion should be dominated by H2, and to ease the measurement problem, it was tested at a sufficiently high-signal amplitude to force 7.6% H2. As can be seen, the agreement between measurement and theory is good, but deteriorates as harmonic order increases because the fundamental assumption that the JFET produces only H2 becomes increasingly invalid at these distortion levels. It is vital to realise that the graph is the result of the need to set test conditions that would produce measurable data and whilst it vindicates the hypothesis, it does *not* represent real-world conditions and therefore almost all conclusions and extrapolations based on this raw graph are wrong.

Despite the previous damning statement, Baxandall's hypothesis is useful because, as we saw earlier, triodes produce distortion dominated by H2 – and all that is necessary is to change the equation's starting conditions (level of open-loop H2) to a more reasonable value (see Figure 3.37).

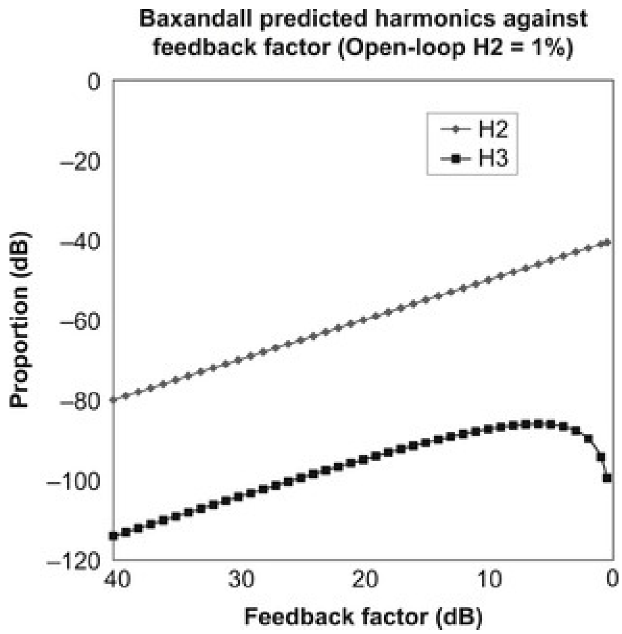


Figure 3.37 The effect on Baxandall's graph of reducing open-loop distortion to 1%.

A triode amplifier might produce 1% distortion at the expected signal amplitude, dominated by H2, and the model shows that if 20 dB of negative feedback were to be applied, H2 would drop in amplitude by 20 dB (as we should expect), and feedback would produce H3 at an amplitude of −95 dB with respect to the fundamental. There are audio sources with noise/distortion floors better than −95 dB, but the most common digital source (CD) is 16 bit and this is its noise floor. In short, what Baxandall's hypothesis *really* showed was that if the distortion spectrum after feedback is not to be questionable, open-loop distortion must be 1% or better.

Although it is difficult to design a power amplifier that produces <1% THD at full power dominated by H2 before applying feedback, designing small-signal stages to this standard is trivial. Thus, if we needed a line stage with a gain of 6 dB (perhaps to make up the 6 dB loss caused by inserting a balance control), we could load a triode from the *SN7*/N7 family with a constant current source to achieve an open-loop gain equal to μ (20), buffer it with a cathode follower (to avoid the feedback resistor loading the gain stage and increasing distortion), then apply 20 dB of feedback (see Figure 3.38).

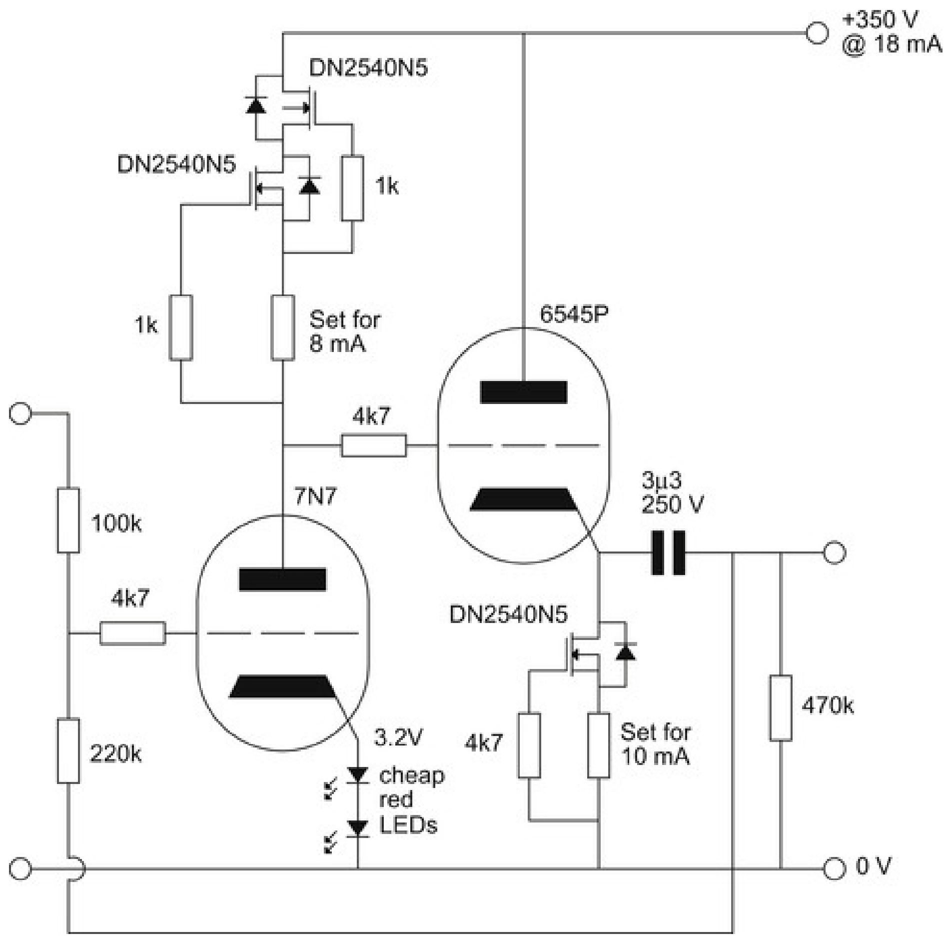


Figure 3.38 This common cathode triode with CCS load and cathode follower buffer has low open-loop distortion – ideal for the addition of negative feedback.

Referring to the measurements made earlier in this chapter, we would expect open-loop H2 at +28 dBu (19.5 V_{RMS}) to be better than −50 dB, but at the more reasonable level of 2 V_{RMS}, we could expect −70 dB (0.03%). Using 0.03% as the starting value in Baxandall's equations predicts that H2 drops by 20 dB to −90 dB, and that the H3 produced by feedback will be entirely negligible at −156 dB. In practice, we know that the gain triode and cathode follower will produce small amounts of H3 and higher harmonics, but these will each be reduced by the feedback, and we can calculate their amplitudes (Table 3.16).

Table 3.16 Calculated Distortion Harmonics of CCS-Loaded *SN7*/N7 Before and After 20 dB of Negative Feedback

	H2 (dB)	**H3 (dB)**	**H4 (dB)**
Before feedback	−70	−105	−110
After 20-dB negative feedback	−90	−125	−130

Summarising, negative feedback only degrades the distortion spectrum if distortion before feedback is >1%, otherwise it pushes all harmonics further into the noise floor.

Note, however, that there was a price to be paid for the low distortion; the grid circuit now has a 100 kΩ series resistor that generates noise and therefore degrades the signal-to-noise ratio of the compound stage.

Carbon Resistors and Distortion

Carbon has much higher resistivity than manganin (resistance wire), so a shorter path gives the same resistance. Since resistors are commonly constructed as a helix, higher resistivity means fewer turns and reduced inductance, so we use carbon resistors for valve and FET grid/gate stoppers.

However, not only are carbon resistors noisy, but they also produce distortion. Bailey (better known for 'transmission line' loudspeakers and long-haired wool) warned in an article on designing an audio generator with less than 0.02% distortion [11] that 'It is important that only wire-wound resistors are used in the filter, as carbon types are non-linear and can produce distortions of up to 0.5%. This non-linearity is often overlooked, but where low levels of distortion are being dealt with, the results may be very serious.'

A quick test of a 100 kΩ carbon resistor in a potential divider revealed resistor distortion at −81-dB H2 and −71-dB H4. Whilst not as bad as Bailey's example, the fact that a single resistor can produce measurable distortion is distinctly alarming.

Nevertheless, carbon is still a valid choice for grid/gate stoppers because the distortion is caused by the resistor's non-linear voltage to current relationship, and if there is no grid/gate current through the resistor, there can be no series voltage developed across it and there can be no distortion.

Noise

In this section, I will define noise to be strictly uncorrelated random events possibly due to the granularity of a current that is the sum of individual electrons over time. More simply, we will include thermal noise and its relatives, but exclude man-made interference. To maximise dynamic range, we must minimise noise.

In audio, there are three main sources of noise:

- Resistances
- Amplifying devices
- DC references.

Noise from Resistances

We saw in Chapter 1 that all resistances produce a noise voltage:

$$V_{\text{noise}} = \sqrt{4kTBR}$$

where

k=Boltzmann's constant ≈1.381×10^{-23} J/K

T=absolute temperature of the conductor ≈°C+273.16

B=bandwidth of the following measuring device

R=resistance of the conductor.

This leads directly to the well-known requirement that low-noise amplifiers should minimise the value of their resistances to minimise noise.

Noise from Resistive Volume Controls

At some point in the amplifying chain, we need a means of controlling volume, with variable attenuation from 0 dB perhaps up to 60 dB. All resistive volume controls can be reduced to a potential divider, so at the same time that the control is attenuating the signal, it is also generating noise. If we assume that the source resistance feeding the volume control is negligible, as far as noise is concerned, the two resistors of the potential divider are in parallel, so we can find that parallel resistance and calculate its noise voltage (over a 20 kHz bandwidth). Because the volume control attenuates, the noise generated at the output of the volume control must be compared with that attenuated signal. If we assume a peak input level of 2 V_{RMS}, we can determine the signal to noise ratio of a 100 kΩ volume control at various attenuations (see Figure 3.39).

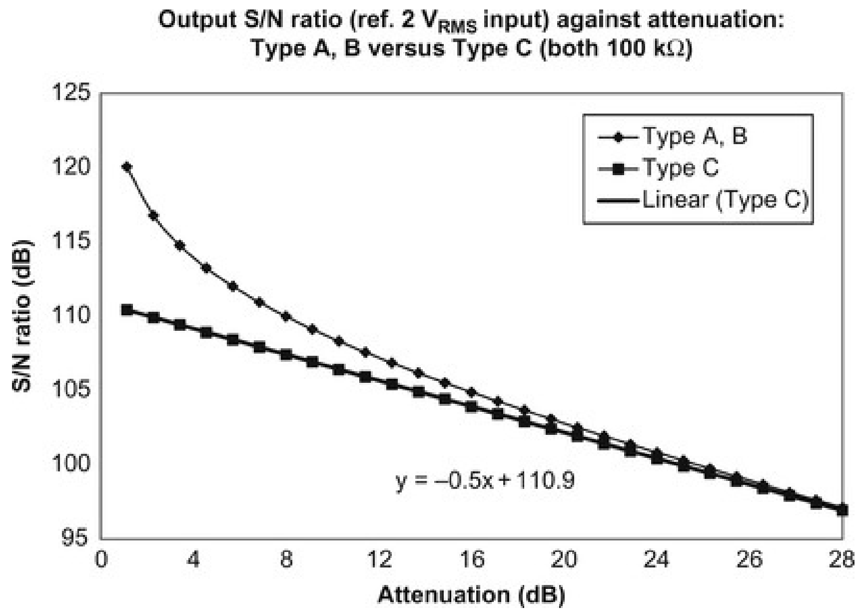


Figure 3.39 Volume control signal to noise changes with setting.

We will investigate detailed design of volume controls in Chapter 7, but suffice to say that a Type C attenuator has a fixed value series resistor with variable shunt, whereas the Types A and B can vary both series and shunt arms. Note that for the specified values of input voltage and resistance, as attenuation increases, all three types tend towards a signal to noise ratio of 111 dB minus half the value of attenuation (in dB). More importantly, if we reduce the input signal from 2 V_{RMS} to 1 V_{RMS} (−6 dB), we immediately reduce the signal to noise ratio by the same amount. Now imagine that we have an amplifier with an input sensitivity of 400 mV$_{RMS}$ and we connect it to a digital source (2 V_{RMS}) via a volume control. At maximum volume (before the amplifier overloads), we must have 14 dB of attenuation, so we have limited our maximum signal to noise ratio to 97 dB purely because of noise in the volume control.

If we require the same signal to noise ratios at 400 mV$_{RMS}$ that we had at 2 V_{RMS}, we must reduce our volume control resistance from 100 kΩ to 5 kΩ, and this is why semiconductor electronics typically has this value of volume control. Alternatively, we could state that since valve electronics generally needs a 100 kΩ volume control, it must be at a point of 2 V_{RMS} sensitivity. (As an example of noisy design, the Leak Stereo 20 power amplifier and Point One Stereo pre-amplifier combination placed a 100 kΩ volume control at a point of 125 mV$_{RMS}$ sensitivity, resulting in a maximum signal to noise ratio purely due to the volume control of 87 dB.)

Noise from Amplifying Devices

There are two distinct noise sources within a triode and, like a transistor, their relative amplitude determines the optimum source resistance. Internal shot noise due to the granularity of anode current can be considered to be a voltage source at the grid (usually expressed as the value of resistance that would produce that noise voltage). Remembering that the noise produced by a resistor is:

$$V_{\mathrm{n(R)}} = \sqrt{4kTBR}$$

And that the equivalent noise resistance of a triode is approximately:

$$R \geq \frac{2.5}{g_{\mathrm{m}}}$$

Substituting we get:

$$V_{\mathrm{n(g)}} \geq \sqrt{\frac{10kTB}{g_{\mathrm{m}}}}$$

Thus, voltage noise referred to the grid is proportional to the inverse square root of g_{m} and this is the reason why quiet RIAA stages for moving coil cartridges require high- g_{m} valves such as E810F, D3a and EC8010.

The granularity of (unwanted) control grid current forms a current source, and a valve's optimum source resistance is therefore:

$$R_{\mathrm{source}} = \frac{V_{\mathrm{n(g)}}}{i_{\mathrm{n(g)}}}$$

where

$V_{\mathrm{n(g)}}$=noise voltage at the grid

$i_{\mathrm{n(g)}}$=noise current from the grid.

Although this equation predicts the optimum source resistance for a given input stage, it doesn't guarantee minimum noise – after all, we could quadruple both voltage and current noise and obtain the same optimum source resistance with a noisier amplifier. However, what the equation does tell us is that high source resistances require low grid current noise. We can see this intuitively by observing that any grid noise current must pass through the source resistance and develop a voltage across it, so the smaller the current, the smaller the voltage. Note that although reactances cannot *produce* noise, Ohm's law guarantees that driving a noise current through the inductance of a moving magnet cartridge converts grid current noise into a noise voltage. Nevertheless, in a noise source context, even moving magnet cartridges are *not* high impedance and the grid current noise of a competently manufactured valve ought to be insignificant even

with a high inductance moving-magnet cartridge.
However, capacitive transducers are unquestionably high impedance sources, and for audio that means condenser microphones. The typical large (20 mm) capsule microphone popular for lead vocals has a source capacitance of ≈30 pF and, for such a source, it is not grid voltage noise that must be minimised but grid current noise.

Grid Current Noise and the Poisson Distribution

Before investigating how to minimise grid current noise, we must make a very brief foray into statistics. Valve control grid current is so low that we are concerned with the exact number of electrons that leave or enter the grid circuit. If we were to consider 1 µs slices of time, and each 1 µs slice contained exactly the same number of electrons, then there would be no audible noise current because it is the variation (AC component) that we hear, not the DC component.
The valve manufacturers will always have designed to minimise grid current noise, and therefore the probability of a given physical process resulting in an electron that contributes to grid current is low, but thermal agitation ensures that there is a large number of tests, so their statistical result has a Poisson distribution. The significance is that a Poisson distribution is completely described by its mean value, and:

$$\sigma = \sqrt{\mu}$$

where

σ=standard deviation (from the mean)

μ=mean.

This equation is the Ohm's law of statistics and that square root relationship is ubiquitous in noise calculations.
For audio purposes, the mean is the DC component of a current (or voltage) and the standard deviation is the AC (noise) component. Thus, if we want to minimise grid current noise (standard deviation), we must minimise the (unwanted) grid current. Note that this argument does not apply to the valve's internal shot noise because we need the anode current in order to modulate it, so it is the *ratio* between the standard deviation and the mean that must be minimised.

Electrometers and Grid Current

Whereas a DVM might measure currents down to fractions of a microampere, an electrometer is an instrument designed specifically to measure very small currents, typically ranging between 1 fA (10^{-15} A) and 1 nA (10^{-9} A). When investigating electrometer input valves, Dagpunar [12] showed that grid current could be split into broad categories, firstly determined by whether electrons were arriving (positive grid current) or leaving (negative grid current). Given that each positive ion arriving at the grid surface requires one or more electrons to leave the grid to discharge it, arriving positive ions contribute to negative grid current.

Positive grid current is due to:

$\boldsymbol{I_1}$

Electrons leaving the cathode but intercepted by the grid.

Negative grid current is due to:

$\boldsymbol{I_2}$

Photoelectric and thermionic electron emission leaving the grid. Positive ions arriving from the heated cathode (atoms of cathode-emissive material). These sources produce a fixed number of electrons/ions within the valve's grid/cathode region, so V_{gk} simply determines how many are collected, leading to a saturation value irrespective of V_{gk}.

$\boldsymbol{I_3}$

Positive gas ions caused by the direct impact of accelerated electrons during their flight to the anode. Positive gas ions caused by the Bremmstrahlung (braking radiation) and X-rays produced by the anode due to arriving electrons being abruptly decelerated. Positive ions of anode material dislodged or emitted from the anode heated by electron bombardment.

$\boldsymbol{I_4}$

Secondary electrons emitted from the grid due to the impact of positive ions.

$\boldsymbol{I_5}$

Surface leakage over glass or insulation inside and outside the valve. This is assumed to be a resistive component, directly proportional to V_{gk}. Brimar [13] noted that the most common cause of noise variation between valves of the same type was grid current noise caused by surface leakage currents across mica spacers and stray fibres or lint between electrodes. In a similar vein, the external surfaces of electrometer valves must not be touched with bare fingers as traces of sweat will cause surface leakage currents.

When represented graphically these idealised categories sum to produce the familiar grid current curve with its crossover point typically at $V_g \approx -1$ V (see Figure 3.40).

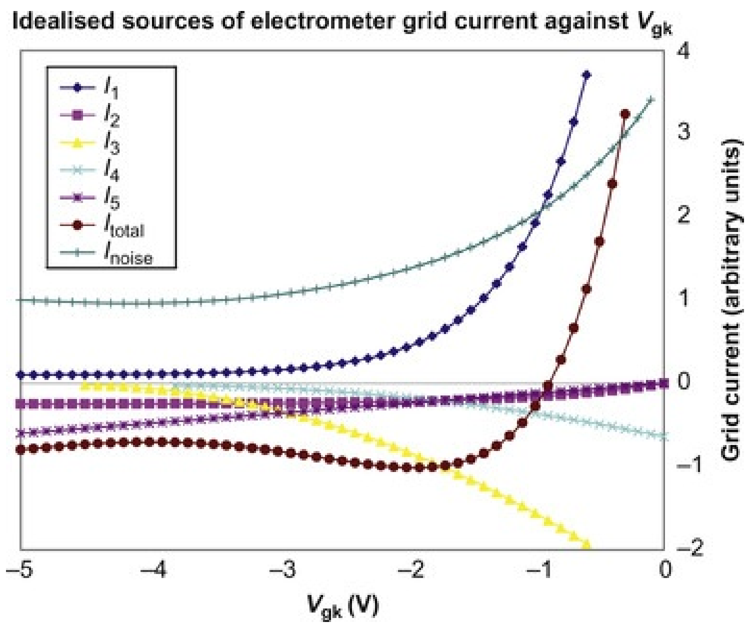


Figure 3.40 Idealised sources of grid current and their effect on grid current noise.
(After Dagpunar [12].)

Each individual physical noise process was earlier assumed to have a Poisson distribution, implying that the standard deviation (noise) was equal to the square root of the mean. But we have seen that grid current is due to a number of sources, so total grid noise current must be the power summation of individual noise sources, and this is why grid current noise does not fall to zero at $I_g=0$, but has a very broad minimum at the inflection of I_g ($V_{gk}=-4$ V in this model). Note that any change in individual noise sources (and particularly leakage) will completely change the noise current curve.

Photoelectric emission from the grid is easily minimised by operating the valve in darkness, but it is currently fashionable to mount a bright LED in the valve base that shines directly into the valve's internal structure. Whilst this is clearly the worst possible photoelectric scenario, any application that is sufficiently insensitive to hum to allow the valve to be on display is also insensitive to the increased grid current noise due to LED illumination. However, remember that LEDs are easily capable of tracking a 100 Hz/120 Hz hum waveform and its harmonics, so it would be unfortunate to prove that an illuminated valve was sensitive to optical hum – if you must have a light show, use clean DC for the LED.

Prior to the production of valves designed specifically for electrometers, Gillespie [14] investigated the possibility of using the Mullard EF37 in an electrometer, and found that there were three practical ways of reducing the grid current of a given valve (ranked in order of decreasing effectiveness and categorised according to Dagpunar):

$\boldsymbol{I_3}$

Reducing anode voltage reduces the *energy* involved in a collision between an accelerated electron and a gas molecule. Gillespie stated that gas/ion current varies with the square of accelerating voltage, so electrometer valves typically operate with an accelerating voltage of only 10 V (whether provided by the anode or screen grid).

$\boldsymbol{I_3}$

Biassing the grid further negative reduces anode current and therefore the *number* of collisions between electrons and gas molecules, thereby reducing gas/ion current.

$\boldsymbol{I_2}$

Reducing heater dissipation reduces grid electron emission and cathode ion emission. (Dedicated electrometer valves avoid indirectly heated cathodes because the unavoidable temperature drop across the heater/cathode insulation requires the heater needs to be hotter, increasing photoelectric emission.)

Gillespie found that because grid electron emission and cathode ion emission are second-order effects, reducing heater dissipation was only worthwhile if all other sources of grid current had been minimised and the valves had been selected for low grid current. Only one valve in six passed Gillespie's low grid current selection process, so when Mullard adopted it, such EF37s were released as type ME1400.

Despite the previous caveats, the reduced heater dissipation technique has erroneously entered audio folklore as a general solution for low noise, yet only a condenser microphone head amplifier using a valve previously selected for low grid current would be likely to achieve a noise reduction from judicious reduction of heater dissipation. Unfortunately, reducing heater dissipation drastically reduces g_m, increasing grid voltage noise, so it seems unlikely that an

improvement of more than 1–2 dB could be gained.
We can summarise grid current noise by stating that for a competently manufactured valve:

- Grid current noise is only significant with high source impedances and for audio, which means condenser microphones.
- Grid current (and its noise) is highly sample dependent because it is usually dominated by surface leakage currents, but a quiet valve is likely to have a very broad minimum well away from the crossover voltage.

The crucial significance of this investigation into grid current noise and making the distinction between a valve's internal shot noise and its grid current noise is that whereas the magnitude of the internal shot noise can be predicted reasonably accurately by fundamental physics, the magnitude of grid current noise is wholly dependent on the vicissitudes of production engineering.
It must be remembered that valves were/are assembled by hand, requiring considerable manual dexterity and attention to detail. Experience shows that manufacturing processes dependent on skilled manual labour require not only a motivated workforce adroitly directed by a good engineer/manager, but also, more importantly, a great deal of practise to determine the exact production technique necessary to meet the required standard, and even in the golden age, valves needed to be selected when low grid current was required. The valve market is far smaller now than in its heyday, so economies of scale can no longer be applied, making the production engineering problem far greater. In short, making valves is high technology dependent on mass production, and it is unrealistic to expect small runs of contemporary valves to achieve the low grid current noise or consistency of selected golden age valves.

Noise in DC references

We occasionally need a DC reference voltage. The most obvious example is the reference in a regulated power supply, but we might also need a reference to enforce operating conditions in an amplifier. We will investigate the DC reference in a regulated supply because it produces some initially counter-intuitive (but very useful) results.
The typical series regulator is nothing more than a non-inverting power amplifier amplifying the DC reference. As an example, if we had a 317 IC regulator with a reference voltage of 1.25 V and needed an output voltage of 15 V, we would choose external resistor values that caused the amplifier to have a gain of 12. But this means we have also amplified the noise voltage of the DC reference by a

factor of 12, and this is why the 317 datasheet specifies the output noise voltage as a percentage of output voltage.

If we need a high voltage, it seems sensible to start with a high-voltage DC reference and amplify it only a little rather than start with a low voltage and suffer higher noise amplification. The key factor is the noise produced by the reference. Sadly, not all the gas DC reference datasheets specified their noise voltage, and information on how that measurement was made was sparse. Clearly, it was time for the author to make some measurements.

How the Author's DC Reference Noise Measurements Were Made

The references were all operated at a current of 5 mA set by a cascode DN2540N5 JFET constant current source powered from a filtered DC supply. Noise was measured by an MJS401D audio test set coupled by an input transformer to improve rejection of mains hum and a capacitor coupled to protect that input transformer (see Figure 3.41).

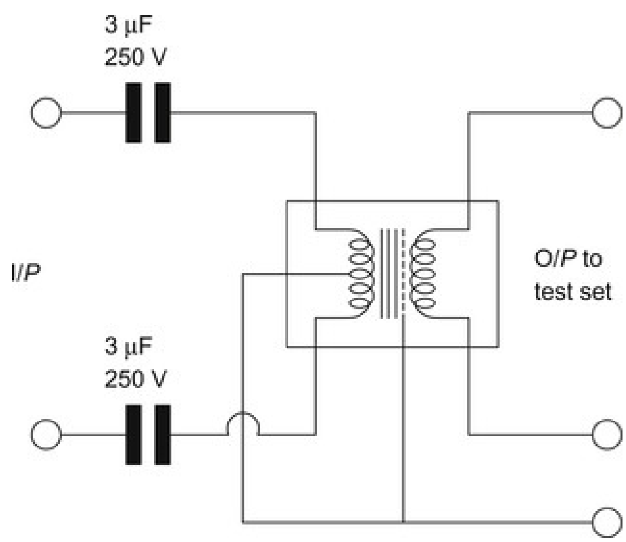


Figure 3.41 Measuring microvolts of noise on a DC supply without damaging test equipment.

The test set was set to RMS rectifier in order to make the results comparable with datasheets. Unless specified otherwise, bandwidth was limited by the MJS401D's maximally flat 36 dB/octave 22 kHz low-pass and 18 dB/octave 22 Hz high-pass filters. This bandwidth is wider than the 10 Hz to 10 kHz bandwidth commonly used by device manufacturers, but remembering that noise is proportional to the square root of measurement bandwidth:

$$\text{Correction factor} = \sqrt{\frac{10{,}000 - 10}{22{,}000 - 22}} = 0.6742$$

which amounts to −3.4 dB, and enables comparison of measurements. As an example, the author measured 138 μVRMS of noise over a 22 kHz bandwidth from a batch of 85A2s, which corresponds to 93 μV_{RMS} over a 10 kHz bandwidth, but the Mullard 85A2 datasheet claimed ‘of the order of 60 μV_{RMS}’, which is 4 dB quieter. (Mullard didn’t explicitly state ‘RMS’, but stating that the noise was equivalent to a 22 MΩ resistor implies RMS.) There are four possibilities for this 4 dB discrepancy:

- The devices have deteriorated over the decades and become noisier.
- The devices needed to be operated for 3 min to achieve their lowest noise.
- The Mullard measurement used a different rectifier (perhaps average reading but calibrated RMS on sine wave) plus smoothing.
- Mullard determined 85A2 noise by comparison. A noise meter could have been switched between the 85A2 and a selection of resistors, and when the two readings were equal, the noise deemed to be that of a 22 MΩ resistor.

The first possibility seems unlikely, and although the second is true for DC stability, no noise change was observable over 2 h. The third possibility could account for 1 dB (switching from RMS to average reading gave a 1 dB reduction on the author’s MJS401D). However, the fact that the noise is stated to be equivalent to that of an E6 standard value resistor suggests that this was the method used, and it could be that the noise measurement on the resistor produced higher noise than should be expected from the square root of 4 $kTBR$. One distinct possibility is that any grid current noise from the noise meter’s input amplifier would develop a significantly higher noise voltage across a 22 MΩ resistor than across the typical 300 Ω slope resistance of a gas reference, leading to a falsely small value of resistor being needed for equal noise, and thus a perceived lower gas reference noise.

There is always a danger when measuring small signals that the measurement is actually of interference such as mains hum or ADC sampling noise from DVMs. As a check, a Tektronix 2213 analogue oscilloscope and a TDS3032 digital oscilloscope were connected to the monitoring output of the audio test set. If the audio test set was measuring pure noise, then setting the digital oscilloscope to average over 512 samples should average the noise nearly to zero and no coherent waveform should remain. Mains hum was the most likely source of interference, so the digital oscilloscope was triggered from AC line to ensure that averaging revealed mains hum if present.

Similarly, there is a danger that the test set’s own noise could be confused with external noise. With no input, MJS401D self-noise measured with 22 Hz to 22 kHz filter and RMS rectifier was −116 dBu.

Gas Reference Noise Measurements

Sadly, the author doesn't have very many gas references, so the statistical validity of the following measurements is somewhat questionable, although sample variation was very low at <1 dB for all devices.

Table 3.17 shows that there is very little difference in the voltage-referred noise between devices, although a 108 V reference appears to be a quieter choice than the ubiquitous 85 V 85A2.

Table 3.17 Comparison of Noise for Various Gas Reference Valves

	Number of samples	Average noise (dBu)	Reference voltage	Voltage-referred noise (μV/V)	Voltage-referred noise (dB)
75C1	3	−74.5	75	1.46	−0.9
85A2	9	−75.0	85	1.62	0
85A1	1	−74.0	85	1.82	+1
QS1215	1	−75.5	90	1.45	−1
QS95/10	2	−72.7	95	1.67	+1.3
VR105/30	4	−74.5	105	1.39	−1.3
CV286	4	−74.3	108	1.38	−1.4
0B2	6	−74.3	108	1.38	−1.4
QS150/15	10	−69.2	150	1.78	+0.8
0A2	5	−69.8	150	1.67	+0.3

Voltage-referred noise is the measured noise voltage divided by the reference voltage and is thus an absolute indicator of the noise the device would contribute when used in a typical regulator. Voltage-referred noise in dB is referenced to the 85A2.

Variation of Gas Reference Noise with Operating Current

We have previously seen that noise is due to the granularity of current, so we should expect the noise of a reference to be proportional to the inverse square root of operating current. A Mullard 0B2 was tested over a 4–26 mA range (see Figure 3.42).

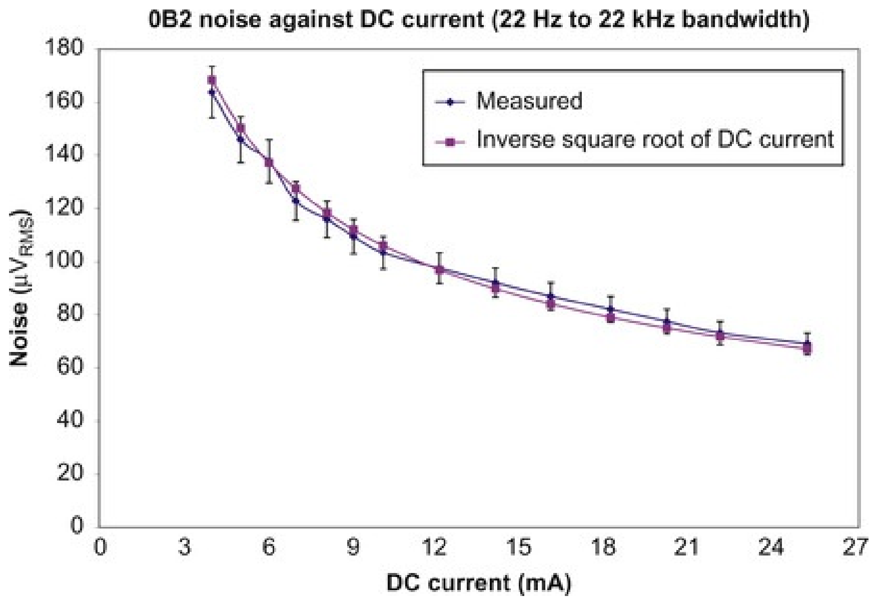


Figure 3.42 0B2 gas reference: 22 Hz to 22 kHz noise against operating current.

The uncertainty of the noise measurements was estimated to be ±0.5 dB, so this uncertainty has been added to each measurement as an error bar. An inverse square root relationship passes comfortably through the error bars, and although a flatter coefficient ($I^{-0.475}$) gives a better fit, inverse square root is quite good enough when choosing operating current.

To sum up, the 108 V references seem the quietest, and noise is inversely proportional to the square root of operating current. Having investigated gas references, the author realised that (for completeness) some semiconductor reference measurements ought to be made.

Semiconductor Reference Noise Measurements and Statistical Summation

Since the series of gas reference tests had been referred to the 85 V 85A2, it not only seemed sensible to compare semiconductors at a similar voltage, but it was expected that the noise of such a high voltage reference would be easier to measure. Thus, 16 BZX55 Zeners having a nominal voltage of 5.6 V were connected in series to produce an 84 V composite Zener. Using the same test set-up as before, this composite Zener produced noise at −101 dBu, which is 29 dB quieter than the 85A2.

A 29 dB noise improvement is so enormous that it immediately questions the measurement technique, so the 85A2s were immediately checked and their noise confirmed. 5.6 V is a quiet voltage still within true Zener action, but as Zaphod

Beeblebrox [15] complained, ‘leaves us a very large improbability gap still to be filled’. The improbability gap is that the noise of each individual Zener is uncorrelated with the others, so noise powers must be summed and the noise of the composite Zener is *not* 16 times the noise of each individual Zener, but only four times. This technique is the dual of reducing current noise by connecting n devices in parallel to reduce their noise by $\sqrt{n}$.

If the $\sqrt{n}$ hypothesis is true, connecting an even larger number of Zeners in series should produce a predictable noise advantage, and this was tested. Thirty-five 5.6 V Zeners were connected in series to form a composite 195 V Zener. If the $\sqrt{n}$ hypothesis is true, the noise advantage produced by this new reference compared to the old should be:

$$\frac{V_{n_2}}{V_{n_1}} = \sqrt{\frac{n_2}{n_1}} = \sqrt{\frac{35}{16}} = 1.479 = 3.4\ \text{dB}$$

Had the $\sqrt{n}$ hypothesis been untrue and the noise sources correlated, we would expect increasing the reference voltage from 84 V to 195 V to increase noise by 7.3 dB (195/84), but the hypothesised 3.4 dB statistical noise advantage should reduce this to 3.9 dB. With both composite Zeners measured at 10 mA, the 84 V composite Zener produced noise at −103 dBu, whereas the 195 V composite Zener produced noise 4 dB higher at −99 dBu, confirming the $\sqrt{n}$ hypothesis.

Variation of Zener Reference Noise with Operating Current

We saw earlier that the 0B2 gas reference produced noise inversely proportional to the square root of its operating current, and the same statistics should apply to the composite Zener. After careful optimisation of the test set-up, uncertainties at these measurement levels were reduced to an estimated 0.7 dB (unavoidably 0.2 dB higher than obtained in the gas reference tests) (see Figure 3.43).

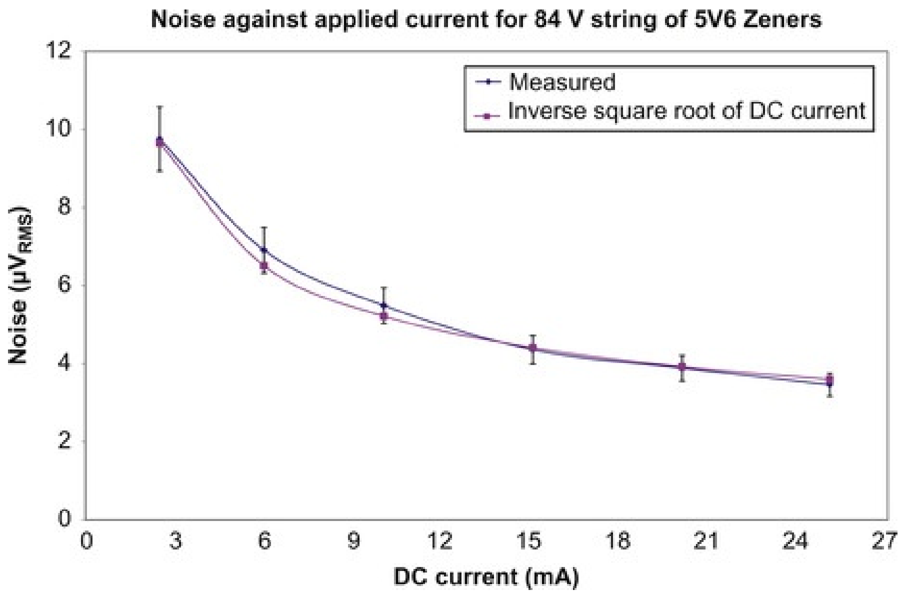


Figure 3.43 BZX55 Zener: 22 Hz to 22 kHz noise against operating current.

As can be seen, an inverse square root curve once again passes through the error bars, confirming the hypothesis. But this result is more significant than before because a 5.6 V BZX55 (500 mW) has a current rating of 89 mA, so we could safely pass a much higher current than we could through a gas reference. 25 mA would not be an unreasonable current, reducing the 84 V composite Zener noise to −107 dBu, or 3.5 μV $_{RMS}$ measured over a 22 Hz to 22 kHz bandwidth.

Noise of the Composite Zener Compared to a 317

The 317 IC regulator has been mentioned in passing, and Linear Technology's 317 datasheet [16] states that a 317 produces noise (10-Hz to 10 kHz bandwidth, RMS rectifier) equal to 0.001% of output voltage, so if we needed a 195 V supply we would expect it to produce 1.95 mV $_{RMS}$ of noise. By comparison, the 195 V composite Zener passing 10 mA produced 8.7 μV of noise over a 22 Hz to 22 kHz bandwidth. Adjusting for bandwidth, this makes the 195 V composite Zener 50 dB quieter than the 317. Admittedly, the usual speed-up capacitor added to the 317 reduces its noise gain at high frequencies, but it must have full gain at DC, and to avoid destroying transient response, the speed-up capacitor only becomes effective >100 Hz, implying ≈100 Hz noise bandwidth and therefore a 20 dB improvement compared to the 10 kHz noise bandwidth, reducing the noise advantage of the composite Zener to 30 dB.

As an aside during the semiconductor DC reference tests, the author tried an 84 V composite device made up of seven BZX55 12 V avalanche diodes, and at 10 mA it produced noise at −75 dBu (comparable with the gas DC references but

28 dB worse than the 5.6 V version). To be fair, the 12 V devices suffered a 3.6 dB statistical disadvantage compared to the 5.6 V, but it does seem that the 12–15 V region is a particularly noisy avalanche voltage. A single ZPY56 56 V device was tried at 6 mA, and this produced noise at −82 dBu, which is significantly better than the gas DC references, but not nearly as good as the composite devices using 5.6 V Zeners.

Red LED Noise

Finally, the author measured a string of HLMP6000 red LEDs. These LEDs are reputed to be very quiet, so 105 were soldered together to form a 167 V reference. Noise did not vary smoothly with current (although it was repeatable), perhaps suggesting a questionable soldered joint. Nevertheless, the results normalised to 84 V suggested that the LEDs were better than 6 dB quieter than the composite Zener using 5.6 V devices. However, whilst hand-soldering 35 wire-ended Zeners in series on a tag strip is perfectly practical (if a little tedious), hand-soldering 105 LEDs isn't, restricting the use of LEDs to low-voltage references, although a surface mount 'pick-and-place' machine followed by wave soldering would undoubtedly have no trouble reliably making a high-voltage low-noise (but rather expensive) reference.

References

[1] Shorter, DL, The influence of high-order products in non-linear distortion, *Electron Eng.* **April** (1950) 152–153.

[2] Henderson, FE, *Introduction to valves*. 2nd ed. (1943)Wireless World.

[3] Wright A. The tube pre-amp cookbook. 2nd ed.; 1997.

[4] Reduction of distortion in vacuum tube circuits. US patent _1,970,325; 1934.

[5] Black HS. US patent _2,102,671.

[6] Deketh, J, *Fundamentals of radio-valve technique*. (1949)Philips Gloeilampenfabrieken; pp. 439–444. Reprinted by Audio Amateur Press, 1999.

[7] Technical manual Sylvania radio tubes. 6th ed. Emporium (PA): Sylvania Electric Products; 1946.

[8] Electron tubes, receiving types 12SN7GT and 12SX7GT. MIL-E-1/63B. USA military specification sheet; 4 March 1954.

[9] Mikolajczyk, P; Paszkowski, B, *Electronic universal Vade-Mecum*. 2nd ed. (1994)Wydawnictwa Naukowo-Techniczne; p. 442.

[10] Baxandall, PJ, *Audio power amplifier design – 5*. (1978)Wireless World;

53–6.

[11] Bailey, AR, Low-distortion sine-wave generator, *Electron Technol.* **February** (1960) 64–67.

[12] Dagpunar, SS, Electrometer valves, *Mullard tech commun* **66** (August) (1963) 194–199.

[13] Brimar application report VAD/508.4 6BS7; January 1952.

[14] Gillespie, AB, *Signal, noise and resolution in nuclear counter amplifiers*. (1953)Pergamon Press; pp. 33–7.

[15] Adams D. The Hitchhiker's Guide to the Galaxy. First broadcast on BBC Radio 4; 15 March 1978.

[16] Linear databook. Section 4, p. 139. Linear Technology; 1990.

Recommended Further Reading

- Metzler, B, *Audio measurement handbook*. (1993)Audio Precision; As should be expected from a manufacturer of audio test equipment, this little book is excellent on purely analogue techniques, but its age precludes recent digital advances.
- Low level measurements. 4th ed. Keithley; 1992. At first glance, this little book might not seem relevant to audio, but the techniques needed to measure small currents accurately are invaluable for minimising audio noise.
- Dove S. Designing a professional mixing console. Studio Sound; September 1980 to February 1982. This down-to-earth series of articles is an excellent tutorial on how to maximise dynamic range written from the perspective of a manufacturer of broadcast desks.
- Fletcher T Balanced or unbalanced? Studio Sound; December 1981. As above.
- Fletcher T, Dove S. Development of a digitally-controlled console. Studio Sound 1983; October. This article includes a particularly interesting circuit dubbed the 'Superbal'.
- Van der Ziel, A, *Noise: sources, characterization, measurement*. (1970)Prentice-Hall; Noise appears to have been Van der Ziel's life work, and this is one of his shorter books that summarises his previous work. Covers noise in bipolar transistors, FETs and valves. If you buy only one of his books, choose this one.
- Self D. Audio power amplifier design handbook. 5th ed. Newnes. The solid state designer's bible to the H C Lin power amplifier – all its problems are

examined in microscopic detail and design solutions offered.
• Duncan, B, *High performance audio power amplifiers*. (1997)Newnes; A somewhat more general treatment of solid state power amplifiers primarily from a professional audio (studio, stage) viewpoint.

Chapter 4. Component Technology

In Chapter 2, we began to design simple circuits, which will later be combined to form complete systems. In doing this, we calculated values for components. We now need to know how to specify voltage or thermal ratings of components.

Correct specification of individual components is extremely important. An underspecified component may fail prematurely and cause further damage, whereas an overspecified component may waste money, which could have made an improvement elsewhere. To be able to specify components correctly requires knowledge of the stresses that will be applied to the component (electrical, thermal and mechanical) and of the imperfections of that breed of component. (No components are perfect, although some are more equal than others).

Much has been said about the ‘sound’ of components, particularly capacitors. This has caused such polarisation of the engineers versus audiophiles debate that rational speech has only rarely been heard. This is curious, since there are well-known physical imperfections in components, and it seems reasonable to suppose that they could have an influence on sound quality. On the other hand, although components are not magical, there *are* purveyors of snake oil.

This chapter will help you to avoid the more obvious pitfalls, but it is *not* a substitute for detailed manufacturers’ data sheets and the application of intelligence.

Resistors

Preferred Values

So far, we have calculated resistor values and then picked the nearest preferred value. These preferred values are derived from an exponential function, so they are known as *E series* (E6, E12, E24 and E96), whose values are given in Appendix A. Each series denotes the number of different values in a decade.

For example, E6 contains the values 1, 1.5, 2.2, 3.3, 4.7 and 6.8, making a total of six values per decade. If we now consider that we will probably need values from 1 Ω to 10 MΩ, then this is seven decades, and we will need 43 different values (10 MΩ is the start of a new decade). For a complete set of E24 (the most commonly used range), we would need 169 different values.

The E series is loosely related to the tolerance of the component, so 20% tolerance components are E6. The reason is that the upper limit of tolerance of one value just meets the lower limit of tolerance of the next highest value, so

there are no gaps in the range.
The argument begins to fall down when we look at E24, since 1.3+5%≠1.5–5%, but the E96 series is aligned more closely.

Heat

Resistors convert electrical energy into thermal energy. The amount of energy converted per second is the power, and this determines the temperature rise.
A signal resistor is unlikely to be a problem, but an anode load resistor could dissipate significant power. We can easily calculate the power dissipated as V^2/R, and select an appropriate component. This is not actually quite as easy as it sounds, and there is plenty of scope for getting it wrong. Resistor manufacturers typically specify power ratings at a component temperature of 70 °C (158 °F).
If your equipment operates at a typical domestic temperature of 20 °C (68 °F), then the *internal* temperature must be higher than this, since the equipment is consuming energy and is not 100% efficient. An average internal temperature of 40 °C (104 °F) would be quite likely, whilst temperature of areas of localised heating (hotspots) could be considerably higher. If you are fortunate enough to live somewhere warmer, then an external temperature of 35 °C (95 °F) might not be unusual, and the internal temperature would rise accordingly.
We can lose heat only from a higher temperature to a lower one, and we can make a useful electrical analogy.

Temperature difference ΔT (°C) is equivalent to potential difference.

Power dissipation q (W) is equivalent to current.

Thermal resistance R_θ (°C/W) is equivalent to electrical resistance.

From this, we can derive a thermal 'Ohm's law':

$$\Delta T = R_\theta \cdot q$$

This tells us that a given thermal resistance will create a greater temperature rise above ambient as more power is dissipated. Resistor specifications give a value for the thermal resistance R_θ, but this value assumes that the flow of air to cool the resistor by convection is *not restricted*.
In practice, we often mount the resistor on a Printed Circuit Board (PCB), which considerably restricts the flow if the board is mounted horizontally. Even mounted vertically, there may still be large components, such as capacitors, that block the air flow.
Combining the arguments of restricted air flow and high ambient temperature, it is not generally advisable to operate resistors at more than one-third of their 70

°C rating unless you are able to do a detailed thermal analysis. Even with this proviso, a resistor operated at one-third of its rating will be significantly warmer than its surroundings, so if it changes its temperature, we should expect its electrical parameters to change too. And they do.

Electrical resistance varies with temperature in accordance with the temperature coefficient of the resistor, generally given in parts per million change of value per degree Celsius. This may sound small, but a 30 °C rise in temperature can cause a significant change in value. Therefore, if we have gone to the expense of using 0.1% resistors in a critical part of a circuit, we should not allow *any* significant power to be dissipated in them if we want their value to remain substantially the same. Maximum dissipation of one-eighth of full-rated power would not be unreasonable. In addition, we should ensure that the resistor is not heated by other components.

Resistors are available in two main types: metal film resistors and wirewound resistors. Despite their recent minor cult status, carbon film resistors are an anachronism and will not be considered, as their tolerance and noise specifications are so very poor, although they are useful as grid-stoppers. (Carbon resistors have reduced inductance compared to metal film resistors because the higher resistivity of carbon means that a given resistance can be made with fewer turns.)

Metal Film Resistors

The control of the quality of the materials and processes in the manufacture of metal film resistors determines their performance, so it is worth detailing their construction.

The process starts with the individual insulating ceramic rods onto which the resistive film is to be deposited. These rods must have a smooth surface, as excessive surface roughness varies the thickness of the resistive layer and causes discontinuities that produce electrical noise. Although the ceramic material is chemically inert, it may have picked up surface contamination, such as grease or packaging materials, so this is burnt off by passing the rods through an oven at a temperature >1,000 °C.

Whilst the rods are still hot, they are transferred to a drum in batches of up to 50,000 at a time. The drum is in a high vacuum sputtering system, which is effectively a large valve. An electron gun fires sufficiently high velocity electrons at the nickel–chromium anode (known as the target) that surface molecules are dislodged to form a nickel–chromium vapour. Tumbling the rods in the drum causes the vapour to deposit evenly upon them. The duration in the drum determines the thickness of the film, and this is the first process that determines the resistance of the final resistor.

The thickness of the applied film affects the noise of the final resistor, with thinner films being noisier than thicker films. If the nickel–chromium alloy contains impurities, it will become more granular, and this also causes noise. If the adhesion of the film to the ceramic rod is poor, the film will lift, causing noise, instability and open circuit failure.

A simple nickel–chromium film cannot achieve a temperature coefficient of 5 ppm unaided, but proprietary techniques can improve this by adjusting the chemistry of the film if necessary.

End caps are fitted next, to allow connection to the resistive film. These end caps are an interference fit onto the rods, and their precise fitting is critical. If they are too tight, then as they are pushed on, they will damage the film, but if they are too loose, then they will not make a good contact. Either of these defects causes noise in the finished resistor. Because the end cap is of a dissimilar metal to the resistive film, the interface between the two is a thermocouple which generates a thermal Electro Motive Force (EMF), but because the EMFs at each end of the resistor are in opposition, their DC cancels. Unfortunately, any AC component of this EMF does not cancel, so the cap material must be carefully selected.

Commonly, ferrous end caps are used, but some manufacturers, such as MEC Holsworthy, use non-magnetic end caps on their Holco range, and it is suggested that this may be a contributory factor to their good sound. Many components use steel-cored leads because the poor thermal conductivity (compared to copper) reduces the temperature rise in associated wiring. The author is sceptical that the signal currents in resistors could be adversely affected by the magnetic properties of the end caps or wires, but is much more prepared to believe that steel might make a poorer contact than copper and that the thermocouple effects at each end cap (which cancel in theory) might not cancel perfectly in practice. Whatever the reason, end caps or leads of any component can easily be checked with a small magnet.

Sputtering is not a particularly precise process, and the resistance of the films typically has a spread of ±10–20%. Now that the end caps have been fitted, it is possible to measure this resistance and grade the rods into batches. The purpose of this is to ensure consistency of helixing (see shortly) and hence of the performance of the product.

Although the rods now have a resistive element, it is quite low resistance, and this must be increased. This is done by cutting a helix through the film from one end cap to the other in order to lengthen the resistive path whilst making it narrower. If there are more turns to the helix, this makes a longer, narrower path, and the resistance of the final resistor is proportionately higher, so resistor manufacturers call this parameter *gain*, and we will return to this later.

Traditionally, the helix was cut by a diamond-edged circular saw whose depth of cut was critical. If the cut was too shallow, then the resistive film would be incompletely removed, leaving traces of material bridging adjacent turns of the helix. If the cut was too deep, the saw would be damaged on the ceramic rod, and subsequent resistors would be cut poorly. Either defect caused noise in the finished resistor.

The modern technique is to use a YAg laser to cut the helix, which produces a narrower, more precise cut, but even this process is not without pitfalls. If the energy directed by the laser is insufficient, the resistive layer is incompletely burnt away, causing bridging. If the energy received from the laser is too great, then the resistive film at the edges of the cut becomes disrupted and has an uneven edge. Both defects cause noise.

As the gain of the resistor rises, the track narrows, causing edge imperfections to become proportionately more significant. This is reflected by manufacturers' published noise performance, which shows that the excess noise generated by film resistors rises for values >100 kΩ. This effect is particularly noticeable for resistors of low power rating because their smaller physical size demands a higher gain for a given value.

Film resistors also have a maximum *voltage* rating which is independent of their power rating, but is determined by the maximum allowable potential across the gap between adjacent turns of their helix. As the applied voltage rises, it becomes more likely that *tracking* (intermittent voltage-dependent conduction) will occur across the gap due to imperfect removal of the film in the gap. Taken to its extreme, a sufficiently high applied DC causes arcing between turns of the helix and permanently damages a film resistor. When using film resistors as anode loads, it is not sufficient simply to ensure that the power dissipation is satisfactory, the voltage rating must also be checked. Typically, higher power components have higher voltage ratings and lower excess noise.

At much lower voltages, tracking is partly responsible for the inclusion of an excess noise specification for the resistor, which is typically given in terms of μV of noise per volt of applied DC. For minimum noise with film resistors, the applied DC across them should be minimised. Rarely specified, a typical value for this excess noise is 0.1 μV/V for values <10 kΩ, but typically doubles for each decade rise in value beyond 10 kΩ, so a 100 kΩ anode resistor might be expected to produce 0.2 μV/V. That 100 kΩ resistor would probably have 150 V_{DC} across it, resulting in 30 μV of noise due to the DC, but if there were also

1 V of signal, this would create 0.2 μV of noise. By this means, applying a *signal* voltage across a film resistor generates a signal level dependent noise or *modulation noise*. Since amplifiers contain many resistors, modulation noise could conceivably rise above the thermal noise floor in a very low noise amplifier, but be masked by a poorer amplifier.

Laser cutting of the film resistor produces a precise tolerance resistor, which then has tinned copper wire leads welded to its end caps before being coated with an insulating protective epoxy film onto which the value is marked.

It will be seen that almost every process can cause noise if carried out incorrectly, so resistor manufacturers routinely measure noise or third harmonic distortion as a means of quality assurance. Unfortunately for audio designers, their noise measurement generally uses a 1 kHz bandwidth filter centred on 1 kHz rather than a 20 Hz to 20 kHz audio band filter. Nevertheless, the figure is a useful guide to product ranges from a given manufacturer.

The resistor need not have end caps and leads fitted. Surface mount resistors have their ends plated with a silver–palladium alloy. When soldering surface mount resistors, it is essential to use a silver-loaded solder to prevent the silver leaching out from the plating and reducing solder ability.

Metal film resistors are commonly available in E24 values from 10 Ω to 1 MΩ, although values up to 50 GΩ are now available off the shelf, and even 500 GΩ is available (at a price).

Power (Wirewound) Resistors

Power resistors are generally wirewound, with 50 W components being readily available, but ratings up to 1 kW are possible. Resistance values cover as many decades as metal film resistors, but the maximum value available is typically 100 kΩ.

Again, the process begins with a ceramic rod as a former for the resistive element, but this time resistance wire or tape is wound onto the rod and welded to the end caps, to which leads are then welded. Smaller components (<20 W) are then coated with a ceramic glaze to prevent movement of the windings and also to seal the component. Larger components may have screw terminal end caps and be fitted into an aluminium extrusion to conduct the heat from the resistive element to an external heatsink. However, high-value resistors require many closely spaced turns of fine resistance wire, so the possibility of tracking between adjacent turns defines a voltage rating which can easily override the power rating.

Ageing Wirewound Resistors

Ageing Wirewound Resistors

Scroggie [1] pointed out that because the resistance wire is wound under tension to ensure a consistent wind, this sets up strains within the wire that relieve with time, causing the resistor's value to change. He further suggested that the process could be accelerated by heating the resistors in an oven to 135 °C for 24 h. The author tested the theory by measuring his entire stock of aluminium-clad wirewound resistors, leaving them in the kitchen oven on its minimum setting for a day, allowing them to cool slowly in the oven, and measuring them again. Even a 3½ digit DVM was able to show significant differences; resistors more than four years old showed no change, but the newest resistors changed by up to 0.5% in value. It therefore seems sensible to age wirewound resistors intended for anode loads in differential pairs *before* matching.

Noise and Inductance of Wirewound Resistors

Because the film resistors' resistive element is a thin track, they develop excess noise proportional to the DC voltage drop across them (≈0.1 μV/V). By contrast, surface imperfections of the resistance wire in a wirewound resistor form a very small proportion of its cross-sectional area, and excess noise is virtually non-existent, making them ideal as anode loads in low-noise pre-amplifiers.

Wirewound resistors are wound as a coil, and even though $\mu_r \approx 1$ for the alumina core (making it comparable with an air core), all coils have inductive reactance which might conceivably be significant compared with their resistance.

The resistance of a conductor is:

$$R = \frac{\rho l}{A}$$

where

ρ=resistivity of conductor

l=length of conductor

A=cross-sectional area of conductor.

But the wire is of circular cross-section, and the area of a circle is:

$$A = \frac{\pi d^2}{4}$$

Substituting we get:

$$R = \frac{4\rho l}{\pi d^2}$$

To make resistors cheap, the resistance wire is wound onto standard-sized cores. To ensure efficient heat transfer to the surroundings, and to reduce the possibility of hotspots, the core is completely covered with one layer of wire from end to end with an infinitesimal gap between turns. The number of turns of wire to completely fill a core of length *C* is:

$$n = \frac{C}{d}$$

The length of this wire is:

$$1 = \pi n D = \frac{\pi C D}{d}$$

Substituting into the resistance equation, π cancels, and the resistance achievable by a single-layer wirewound resistor is:

$$R = \frac{4\rho C D}{d^3}$$

Simplifying:

$$R \propto \frac{1}{d^3}$$

Inductance is proportional to n^2, and since n is proportional to $1/d$:

$$L \propto \frac{1}{d^2}$$

As observed earlier, it is the *ratio* of *L* to *R* that is important, not the absolute value, so:

$$\frac{L}{R} \propto \frac{1/d^2}{1/d^3} \propto d$$

This result is very significant because it shows us that L/R rises as we use thicker wire, so we should only expect low value wirewound resistors to possess significant inductance. This theory was tested on a component analyser, which produced models for a selection of wirewound resistors. Because the resistors were aluminium clad, transformer action to the shorted turn might be expected to reduce inductance, but later dissection of a WH25 type showed that the coil diameter was half the internal cladding diameter, implying loose coupling and insignificant transformer action (see Figure 4.1).

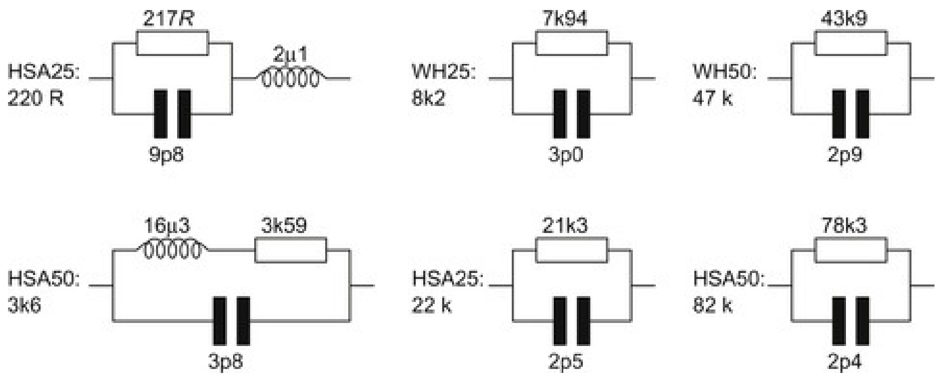


Figure 4.1 Equivalent circuits of practical wirewound resistors.

As can be seen from the models, measurement confirms the theory by showing that only low resistance wirewound resistors had measurable inductance. Besides deriving models, each resistor was swept from 100 Hz to 100 kHz whilst measuring phase deviation from a perfect resistor. Only the 220 Ω resistor showed a measurable deviation (0.2°).

All of the resistor models required a small shunt capacitor, and once the resistor values were typical of anode load resistances, this shunt capacitor settled to a value of 3 ± 1 pF, a value commensurate with the strays that one would expect to find in a practical circuit.

Summarising, the inductance of wirewound resistors is entirely negligible even at 220 Ω, but, as predicted, inductance is more observable at low resistances than high resistances.

Some resistors are deliberately wound to minimise inductance. One way to do this is to take the centre of the wire to be wound and begin winding in the middle of the resistor until the ends of the wire reach the end caps. The significance of this Ayrton–Perry winding is that one coil of wire is effectively wound in the opposite direction to the other and their mutual inductance tends to cancel. Fortunately, the author's stock of vitreous enamelled resistors included a variety of 1 k2 resistors, with some wound to the Ayrton–Perry methodology (easily identifiable because this manufacturer used a brazed connection directly to the terminating wires rather than press-fitted end caps) (Table 4.1).

Table 4.1 Inductance of 1 k2 Wirewound Resistors

Power rating (W)	Inductance (μH)
2.5	6
4	18
6 (Ayrton–Perry)	9

From the previous model, we should expect a physically larger resistor to have higher inductance, and this is true when comparing the 2.5 W resistor with the 4 W resistor, but the Ayrton–Perry winding of the even larger 6 W type significantly reduces inductance from what might be expected. Note that even 18 μH is entirely trivial compared to 1k2 and causes only 0.5° of phase shift at 100 kHz.

The wirewound resistor to be avoided at all costs is the old-fashioned wirewound rheostat occasionally found lurking at the back of a cupboard in an old physics laboratory. With a power rating of 200 W or more, these beasts seem ideal as power amplifier dummy loads, but their typical ≈1″ (25 mm) core diameter and length of ≈4″ (100 mm) mean that their inductance becomes significant at 1 kHz, and if their DC resistance is used in a V^2/R power calculation, the assumed output power of an amplifier will be wrong. As an example, the author used one of these devices while measuring the output power of his 'Scrapbox Challenge' single-ended amplifier for the third edition, and measured 6.8 W when 6 W was expected. The reason for the error was that the very slightly higher impedance load allowed the amplifier to swing a higher voltage than expected, and this error was magnified by the V^2 term in the power calculation.

Non-Inductive Thick Film Power Resistors

The hybrid technology that was initially used to make wide bandwidth amplifiers by printing resistors and tracks directly to a ceramic substrate and then adding surface mount transistors has been borrowed to make non-inductive power metal film resistors. At the lower end (5 W) of the range, the resistors are the size of a postage stamp and are cooled by convection, but higher dissipation types are now available either in transistor packages (TO-220) or in moulded plastic packages having a footprint identical to that of the traditional aluminium-clad resistors, but with non-magnetic terminations. Not only is their inductance dominated by the leads soldered to them, but stray capacitance also tends to be even lower than the typical 0.5 pF of small metal film resistors and certainly lower than the 3 pF typical of wirewound resistors.

General Considerations on Choosing Resistors

Tolerance

- Is the *absolute* value important? If the resistor is part of a network that determines a filter or equalisation network, then we need a close tolerance (perhaps even 0.1%) component to minimise frequency response errors.
- *Matching*: Is the component part of a pair? Anode loads in differential pairs should be matched, and so should corresponding components in filter

networks for each stereo channel.

Heat

Will the resistor be heated by other components? Will its value change? Will this matter?

Voltage Rating

- Is the voltage rating of the component adequate, even under conditions of maximum signal? (Grid-leak resistors for low- μ power valves, such as 845, might need to consider this factor.)
- Will the DC voltage drop across the resistor develop an unacceptable level of excess noise? If so, a wirewound or a bulk foil type should be considered.

Power Rating

Is the power rating of the component adequate under all conditions? Could the (varying) *audio* signal heat the resistor sufficiently to change its value and cause an error? If a power component is required, what provision has been made to lose the heat that this component will generate? Will it heat other, sensitive components?

Capacitors

Capacitors store charge. This charge is stored in the electric field between two plates having a potential difference between them. If there is no potential difference between the plates, then there is no stored charge, and the capacitor is said to be discharged.

Capacitors for electronic circuits are made of two fundamental components: a pair of conducting plates and an insulating material called the *dielectric* that separates them. In its simplest form, a capacitor could be a pair of parallel plates separated by a vacuum.

The Parallel Plate Capacitor

Unsurprisingly, the capacitance of a parallel plate capacitor is proportional to the area (A) of the plates and inversely proportional to the distance (d) between the plates. We should expect this, since if we move the plates an infinite distance apart, they can no longer ‘see’ one another, and a plate on its own is not much of a capacitor. If the charge is stored between the plates, then it is reasonable to suppose that the interposition of any material between the plates will affect the capacitance. We can formalise these arguments by combining them into a proportionality:

$$C \propto \frac{Ak}{d}$$

where

C=capacitance

A=area of plates

k=relative permittivity of interposed dielectric

d=distance between plates.

To calculate real values in electronic units, we must add some fudge factors to generate the equation:

$$C = \frac{A \cdot \varepsilon_0 \cdot \varepsilon_r}{d}$$

This equation looks a lot more impressive, but ε_0 is simply a fudge factor to make the real world fit into our system of units and is known as 'the permittivity of free space'; it has a measured value of ≈8.854×10^{-12} F/m. ε_r (also known as '*k*') is the *relative* permittivity of the material we insert as the dielectric, compared to the value for a vacuum, and is always >1.

A quick calculation using this equation shows that a parallel plate capacitor in a vacuum (although air is almost identical) with plates 1 m^2, separated by 10 cm, would have a capacitance of 88.5 pF. If we are going to make practical valve amplifiers, we are going to have to do something about the size of this capacitor.

Reducing the Gap Between the Plates and Adding Plates

One obvious method of increasing capacitance is to reduce the gap between the plates, so typical commercial capacitors use gaps of 5 μm or less.

Another method is to add more conductive plates in the form of a stack with alternate plates connected together. This almost doubles capacitance over what might at first be expected because we now use *both* sides of each plate (except for the two outermost plates). This form of construction is used for silvered mica capacitors and for stacked film/foil capacitors (see Figure 4.2).

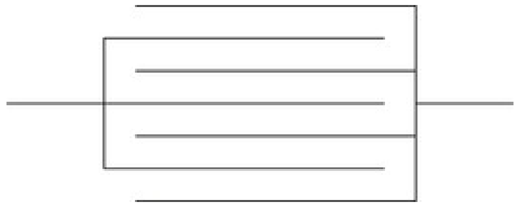


Figure 4.2 Cross-section of general form of parallel-plate capacitor.

Cutting squares of dielectric and plates and assembling them to form a capacitor are an expensive business, so most capacitors are constructed by winding two long strips of plates and dielectric together to form a cylinder, and then connecting a wire to each plate.

The Dielectric

Maintaining a precise air gap of 5 μm between a set of plates would be virtually impossible, so an insulating spacer is needed. This insulating dielectric will have $\varepsilon_r > 1$, which further reduces the physical size of the capacitor for a given value of capacitance.

Unfortunately, we gain this increase of capacitance at the expense of other parameters, and so we should investigate these. The dielectric has three important properties: relative permittivity ε_r, *dielectric strength* and *dielectric loss*.

Relative permittivity ε_r has been mentioned earlier and is effectively the factor by which the capacitance of a capacitor is increased by the insertion of the new dielectric.

Dielectric strength refers to the maximum field strength, measured in volts per metre, that can be applied to a given insulator before it breaks down and conducts. It is this limit that sets voltage ratings for capacitors.

Dielectric loss refers to how closely the dielectric approaches a perfect insulator at voltages *below* breakdown. One way of specifying this loss is to measure the leakage current, in μA, that flows when the maximum rated voltage is applied across the capacitor – this method is typically used for aluminium electrolytic and tantalum capacitors. Film capacitors are typically rather less lossy, and so the *insulation resistance* or *leakage resistance* of the capacitor may be specified. Dielectric loss may be different from AC to DC, and so a more useful measurement is to measure *tan* δ, which is the ratio of the total resistive component of the capacitor to the reactive component at a specified frequency or frequencies. Note that tan δ does not distinguish between the parallel leakage resistance of the dielectric or any series resistance, such as lead or plate

resistance.
Lead and plate resistance are collected together as one term and are known as the *Effective Series Resistance* (*ESR*). In components such as high-capacitance electrolytic capacitors for power supplies, or cathode bypasses, the ESR is highly significant, since it may be an appreciable fraction of the total impedance of the capacitor. In power supplies, significant currents flow in the reservoir capacitors, which cause self-heating of the internal structure. For this reason, a parameter is quoted that is very closely linked to ESR, and this is *maximum ripple current* (see Chapter 5).
Both the leads and the plates have series inductance. For a modern capacitor, series inductance measured as close to the capacitor body as possible typically ranges between 10 nH and 100 nH. We can now draw a simple equivalent circuit for a real capacitor (see Figure 4.3).

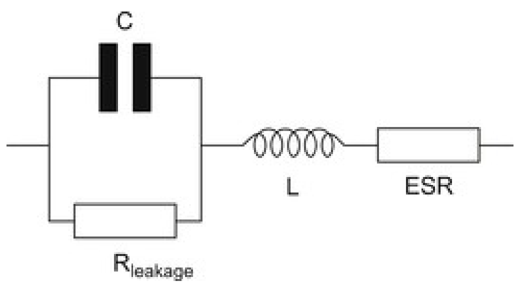


Figure 4.3 Basic equivalent circuit of practical capacitor.

It is immediately apparent that we are dealing with a resonant circuit, and for electrolytic capacitors, this self-resonant frequency is often specified in the manufacturer’s datasheets, and we will return to this later.

Different Types of Capacitors

With the various ways of making the plates or the dielectric, there are many combinations of capacitor construction available (see Figure 4.4).

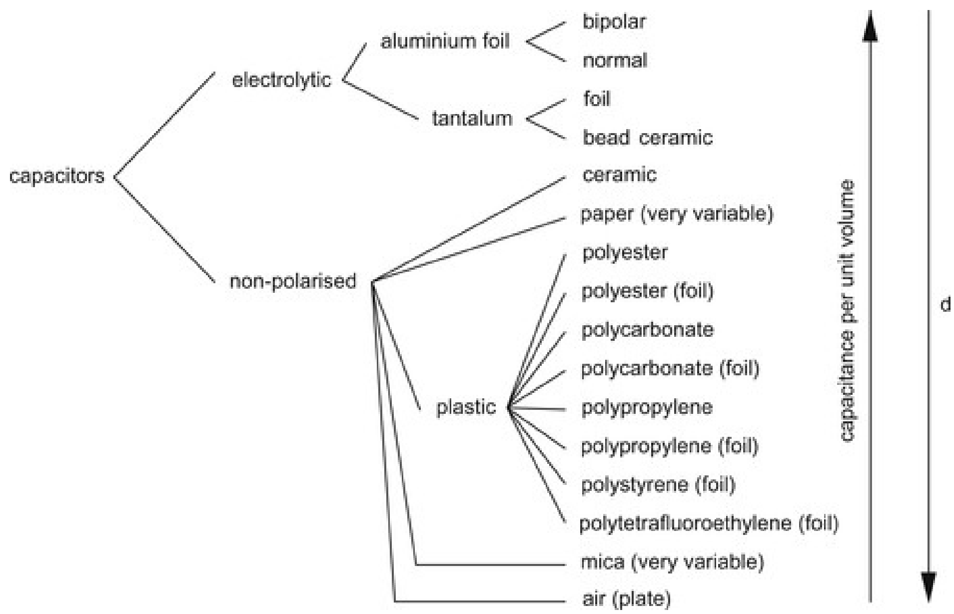


Figure 4.4 Comparison of different types of capacitor.

This tree of capacitors shows the various possibilities available. The first branching is between polarised and non-polarised capacitors. A polarised capacitor would be damaged by having DC applied in the reverse direction. Non-polarised capacitors branch into their plate construction, self-supporting plates, foil or a surface coating of metal sputtered directly onto the dielectric. The final branchings deal with the dielectric, and although some dielectrics are represented in both categories, others are not, due to their manufacturing impossibility.

Broadly speaking, the more nearly perfect capacitors are at the bottom of the tree, whilst high capacitance per unit volume capacitors are at the top of the tree. This can be further generalised by observing that high quality capacitors tend to be physically large for their value of capacitance.

Air Dielectric, Metal Plate ($\varepsilon_r \approx 1$)

These capacitors are invariably constructed as trimmer or variable capacitors with sets of intermeshing semicircular rigid plates and are primarily used in radio frequency (RF) circuits, although they are occasionally useful in audio. Because of the difficulty of supporting plates that are very closely separated, air dielectric capacitors have low values of capacitance and are not usually larger than 500 pF. They generally have ≈10:1 range between maximum value (vanes fully meshed) and minimum value (vanes fully separated). Possible audio uses include:

- The typical 2×365 pF tuning capacitor from an old radio across the inputs of a moving-magnet RIAA stage allows any cartridge to be optimally loaded by the pre-amplifier.
- RIAA stages with the equalisation in the feedback network may require sufficiently low-value capacitors that they are within the range of air types.
- ≈50 pF for trimming equalisation capacitors to their exact value.

Valve short-wave radios often contain many smaller ‘beehive’ trimmer capacitors, and although they might not be the exact maximum value required for your particular application, the silver-plated brass vanes are soldered to their supports, so they can easily be desoldered to reduce the maximum value as necessary (see Figure 4.5).

Figure 4.5 Selection of variable air-spaced capacitors. Note that the right-hand trimmer has its vanes disengaged so that they can be seen.

Plastic Film, Foil Plate Capacitors (2< ε_r<4)

This is the most important class of capacitors for use in valve amplifiers, as we will use these for coupling stages and also for precise filters. The better dielectrics are very nearly perfect, and manufacturers usually describe their imperfections in terms of the value of tan δ or dissipation factor ‘ *d*’:

$$\tan \delta = \frac{R}{X_C}$$

where

R=sum of resistive losses expressed as parallel resistance

X_C=capacitive reactance.

There appears to be a strong correlation between the subjective sound quality of capacitors and their value of ‘ *d*’, with low‘ *d*’ capacitors being subjectively superior.

The significance of ‘ *d*’ in engineering terms is not simply that the capacitor has a leakage resistance across it, but that the capacitor is actually a ladder network of capacitors, separated by resistors, that extends indefinitely (see Figure 4.6).

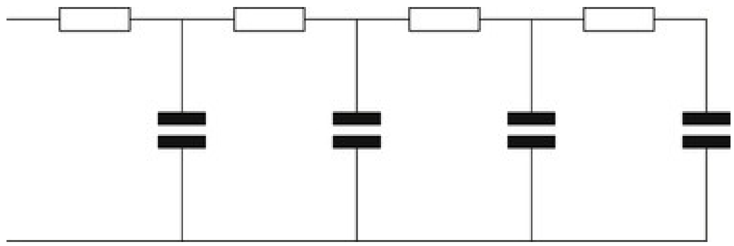


Figure 4.6 Equivalent circuit of practical capacitor to model dielectric absorption.

If we were to charge a capacitor whilst monitoring its terminal voltage with a voltmeter of infinite resistance and then discharge it by briefly short-circuiting it, we would expect the capacitor voltage to remain at 0 V. However, we actually see the voltage rise from 0 V the instant that the short circuit is removed. This is because we discharged the capacitor that is 'near' to the terminals, but other capacitors were isolated by series resistors and were not discharged. Removing the short circuit allowed the undischarged capacitors to recharge the 'near' capacitor and the voltage at the capacitor terminals rose. This effect is known as *dielectric absorption*, and it is more pronounced as the value of ' *d*' rises.

Applying a pulse to a capacitor is equivalent to instantaneously charging and discharging the capacitor, so any voltage left on the capacitor at the end of the pulse is distortion. Music is made up of a series of transients, or pulses, and it may be that dielectric absorption is one cause of capacitor 'sound'.

Film/foil capacitors are constructed by laying four alternate layers of dielectric and foil which are then wound into a cylinder. Guiding these four layers whilst winding the capacitor tightly is not a trivial task, and is partly responsible for the higher price of these capacitors. The foils are wound slightly offset to one another (known as extended foil construction) so that one end of the cylinder has an exposed spiral of foil that is one plate, whereas the exposed spiral at the other end is the other plate. Each spiral is then sprayed with zinc or tin–zinc alloy to connect all points of that spiral together, and because this puts all of the foil in parallel, this greatly reduces series inductance and resistance. A further reduction of series inductance and resistance can be obtained by reducing the aspect ratio (can length divided by diameter) because this simultaneously shortens the series paths and increases their parallel number, but the narrower film strip requires more turns for a given capacitance, increasing the time taken to make each capacitor.

Because of the low melting point of polystyrene, traditional small (<100 nF) polystyrene capacitors made a single tab contact halfway along the foil, although larger capacitors (470 nF) made two tab contacts to reduce series inductance. However, LCR appear to have circumvented the polystyrene melting problem and make contact over the entire extended foil (see Figure 4.7).

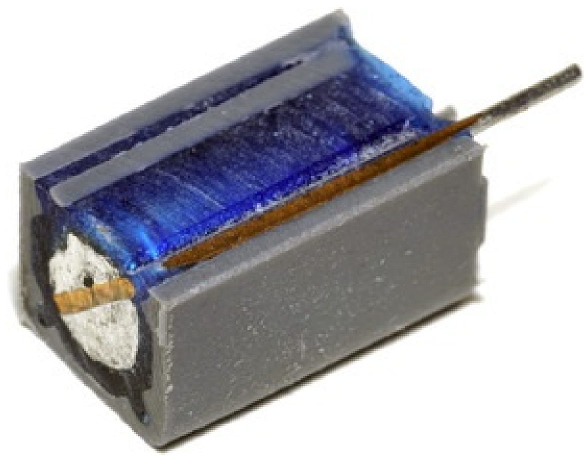


Figure 4.7 Polystyrene capacitor.

Polystyrene capacitors often have one terminal of the capacitor delineated by a red or yellow band or a dot. This does not mean that they are sensitive to polarity, but that the marked end is connected to the outer foil. This is significant because one end of the capacitor may be connected to a less sensitive part of circuitry than the other. For instance, if a small polystyrene capacitor was used as part of an active crossover network and connected as a series coupling capacitor (high-pass filter), then the outer foil should be connected to the source to reduce induced hum. Alternatively, if one end were connected to ground, then the outer foil should go to ground to reduce stray capacitance to active signals (strays to ground rarely cause problems, but the Miller effect can cause other strays to be significant).

Although polytetrafluoroethylene (PTFE) is an excellent dielectric, there is an abrupt phase change at 19 °C that causes a 1% change in volume. If the material is unconstrained, the change in volume is seen as a change in capacitance. When constrained, the attempt to change volume causes a change of capacitor charge because PTFE is piezo electric. The effect is negligible in most applications, but the temperature-induced change in charge caused by a PTFE feed-through insulator compressed into a hole in a conductive chassis was enough to ruin the performance of a prototype electrometer. Fortunately, the heating effect of nearby valves is likely to prevent the 19 °C phase change being observable even in a high impedance audio application such as a condenser microphone. PTFE is also triboelectric (scraping it with a conductor transfers charge), so tapping PTFE insulated wires in a high-gain amplifier causes a disturbance. As a minor point of interest, PTFE sleeved wire is silver plated not because the improved conductivity compared to copper would reduce skin effect at high frequencies (HFs), but for the much more prosaic reason that tinning would melt at the high temperature needed to extrude PTFE.

Metallised Plastic Film Capacitors

Because of the difficulty of winding the layers of dielectric *and* foil, most film capacitors are made by sputtering one side of the film with a layer of aluminium up to 12 μm thick to form the plate. This makes the capacitor easier to make, and a higher capacitance per unit volume is obtained because the plate is so much thinner, but ESR is usually higher than for a foil capacitor. Since ESR for a plastic capacitor is only significant at very high frequencies when it becomes comparable with the reactance of the capacitor, this may not be a problem. However, *foil* capacitors are generally described by their manufacturer as being more suitable for high frequency pulse applications because their lower ESR reduces self-heating (I^2R).

If there is granularity due to impurities in the film of a metal film resistor, this generates excess noise, and film resistors are always noisier than wirewound resistors. Since the plates in a metallised film capacitor are also produced by sputtering, it is not unreasonable to suppose that they will suffer from the same quality control problems – with the difference that capacitors are not routinely tested for modulation noise. Foil capacitors have traditionally been preferred in audio, possibly for this reason.

Recent metallised capacitors are almost indistinguishable from foil capacitors. The author compared a 100 nF 2 kV Vishay Roederstein MKP 1845 metallised polypropylene to a 100 nF 1 kV LCR foil polypropylene, and after allowing for the different working voltages (and therefore dielectric thicknesses), the two had almost identical losses and ESR.

Metallised Paper Capacitors (1.8< ε_r<6)

Metallised paper was the traditional dielectric for capacitors in classic valve amplifiers, and depending on the paper and its impregnant, the performance ranged from poor to tolerable. Unfortunately, if the seals of the capacitor are less than perfect, humidity enters and the capacitor becomes electrically leaky. The author once bought a Leak Stereo 20 power amplifier having paper-coupling capacitors, every one of which had gone leaky.

Because paper capacitors are inherently self-healing, they are widely used in the power generating industry. In the event of an overvoltage spike, the paper breaks down at its weakest point and the metallisation at that point is vaporised, thus preventing a short circuit and catastrophic failure.

It has been known for a century [2] that paper capacitors have capacitance that falls with frequency, so although they might be suitable as coupling capacitors, they should not be used for equalisation (especially avoid RIAA because the wide separation of time constants exacerbates response errors due to errors in component values).

Silvered Mica Capacitors (Muscovite Mica, ε_r=7.0)

This was the traditional small-value capacitor used for RF circuitry, or for audio filters where excellent stability of value was important. Mica is a crystalline rock that can be easily cleaved into fine sheets, which are then coated with silver, and a stacked construction gives low inductance.

Since mica is a natural material, it is subject to all the accompanying vagaries of inconsistency. There are various types of mica, but muscovite makes the lowest loss capacitors. Although muscovite mica and polystyrene have comparable losses ($0.001< d<0.0002$), dielectric absorption is 80 times worse for muscovite mica than polystyrene [3], so polystyrene is generally preferred for audio.

Most of the world’s mica has been mined, making the remainder expensive and of variable quality, so silvered mica capacitors have now been almost eclipsed by polystyrene, which, in turn, is being superseded by the very slightly inferior polypropylene.

Ceramic Capacitors

These have no place in the path of analogue audio!

Until now, the dielectrics that we have seen have had ε_r<10, but ‘high- *k*’ ceramic capacitors can achieve ε_r (or k)≈200,000! Commonly, ceramic capacitors are made up of barium or strontium titanate, both of which are piezoelectric materials. This means that they generate a voltage when mechanically stressed (these materials were the basis of ceramic cartridges used by inferior ‘music centres’ for playing vinyl records).

Ceramic capacitors excel as high frequency bypasses in digital or heater circuitry where their poor stability of value and low ‘ *d*’ are irrelevant.

Electrolytic Capacitors

These capacitors are polarised. Reverse biassed, they form quite a good quality short circuit, and damage the driving circuitry, whilst the capacitor expires to the accompaniment of heat, smoke and noxious fumes. Aluminium electrolytic capacitors may even explode and shower the surrounding circuitry with soggy paper and aluminium foil, causing further damage.

Some people have religious convictions against using electrolytic capacitors, but with all their faults they are still useful components, and our choice of design is severely restricted if we refuse to use them. Most of the faults ascribed to electrolytic capacitors relate to inappropriate usage.

Electrolytic capacitors take high capacitance per unit volume to the limit, and they do so by attacking all parts of the parallel plate capacitor equation. The gap between the plates is minimised, surface area is maximised, and $\varepsilon_r \approx 8.5$ for aluminium oxide, as opposed to $\varepsilon_r \approx 3$ for the plastic films. The principle of operation is broadly similar for all types, so only the aluminium type will be described in detail.

Aluminium Electrolytic Capacitors ($\varepsilon_r \approx 8.5$)

The aluminium foil of one plate is anodised to form an insulating layer of aluminium oxide on the surface (≈1.5 nm/V of applied polarising voltage), and it is this micro-thin layer that is the dielectric of the capacitor. Since anodising is an electrochemical process, and the aluminium oxide is an insulator, it follows that there must be a maximum thickness of aluminium oxide that can be produced before the insulation of the layer prevents deeper anodising, implying a maximum possible working voltage of ≈600 VDC. Unfortunately, a practical problem intervenes somewhat earlier. Aluminium oxide is a very brittle insulator, but the electrolytic capacitor is formed as a roll, so there is a danger of the rolling process cracking the oxide and causing leakage paths. Because of this, the practical maximum working voltage of electrolytic capacitors tends to be ≈500 VDC, and old >450 V capacitors should be looked upon with grave suspicion. Because of the fragility of the brittle dielectric, we should be careful not to dent electrolytic capacitors by overzealous tightening of capacitor clamps.

Although by anodising the aluminium foil we have both a plate and the dielectric, we still need the other plate. We could use another piece of aluminium foil pressed tightly to the first, but any gap between the two foils would negate the advantage of the micro-thin dielectric. The second plate is therefore made of thin soggy paper or simply a gel, which because it is wet makes perfect contact with the anodised surface of the first plate, and this is the electrolyte from which the component derives its name. The electrolyte is not a particularly good conductor of electricity, and so a second aluminium foil is laid on top of the electrolyte to allow a low resistance plate to be made.

We now have two aluminium foils separated by electrolyte that can be rolled into a cylinder to make our capacitor. If, before anodising the aluminium foil, we had etched the surface of the aluminium foil, this would roughen the surface and greatly increase the surface area on a microscopic level. Since the electrolyte plate is in perfect contact with this surface, we have dramatically increased the area of the plates of the capacitor, and the capacitance rises accordingly.

Unfortunately, the electrolytic capacitor has its disadvantages. Electrolyte resistance is significant, so deep etching of the foil increases the resistance from the bulk of the foil to the extremities that form the plate, and we can expect the highest capacitance per unit volume components to have a higher ESR. Not only do these tortuous paths into the nooks and crannies increase resistance, but they have also limited current carrying ability before heating significantly and causing the electrolyte to evaporate. Compact capacitors therefore not only have high ESR, but also have a low ripple current rating, although Sanyo's OS-CON range of capacitors uses an organic semiconductor electrolyte that significantly reduces ESR. Interestingly, the best use for OS-CONs is not in analogue electronics but in digital electronics, bypassing supply rails adjacent to noisy digital chips (see Figure 4.8).

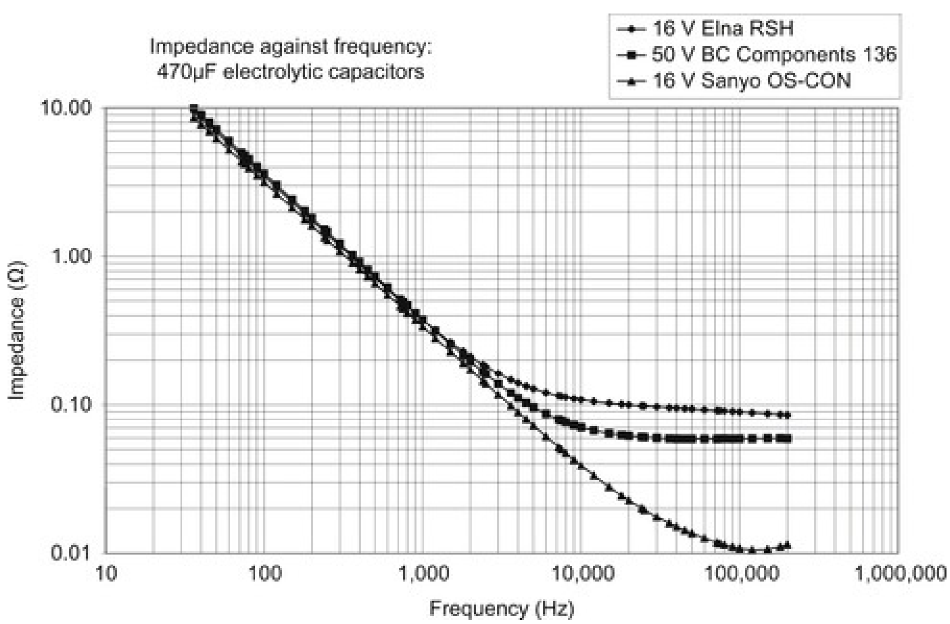


Figure 4.8 Capacitor impedance falls with frequency to its ESR.

Spraying molten zinc onto the entire spiral of an extended foil capacitor connected all parts of each plate together, which was equivalent to an infinite number of connections, and minimised series inductance. This technique is not possible with an electrolytic capacitor, and we are forced to connect to the plate at a single point using a welded foil tab, but we do have a choice as to where to position that tab along the foil. The worst place to position the tab would be at one end of the foil because that would give maximum inductance to the opposite end of the foil. Conversely, placing the tab in the middle of the foil gives equal inductance to each end, which would be in parallel, and therefore halved.

Given the electrolytic capacitor's single tab connection, series inductance and resistance can be reduced by increasing the aspect ratio (can length divided by diameter) to obtain the required capacitance because this makes the unrolled plate nearer to a square than a narrow strip. For a given capacitance, the strip and square must have equal area, but inductance is proportional to length and falls with width, so a square is the optimum shape.

Although the manufacturer aims to minimise series inductance and thus inductive reactance ($X_L = 2\pi fL$), for a large capacitor, X_C is also low, so even a small series inductance becomes significant. Electrolytic capacitors typically have series inductance ranging between 10 nH and 100 nH depending on can size and shape, with older types tending to have higher inductance. As a consequence of this series inductance, larger capacitors always have a lower resonant frequency, which may be as low as tens of kHz (see Figure 4.9).

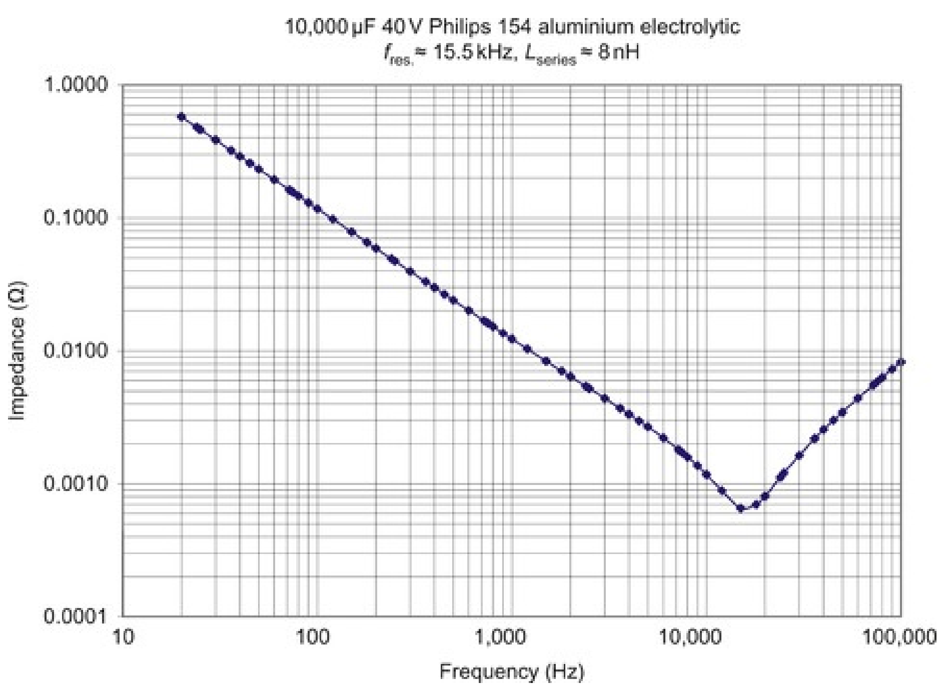


Figure 4.9 This 10,000-µF capacitor may have its resonant frequency within the audio bandwidth, but it is still an effective short

circuit to AC.

Note that although this capacitor has a resonant frequency of 15 kHz, it is still a very adequate short circuit (8 mΩ) at 100 kHz. Limitations of the author's test equipment (Hameg 8118) mean that the graph suggests an overly optimistic ESR of 0.6 mΩ, but the series inductance (derived from resonant frequency and capacitance) is reliable.

Electrolytic capacitors are lossy. When they are first manufactured, a polarising voltage is applied, and this causes an anodising current to flow through the capacitor, forming the aluminium oxide layer on the plate. Once this oxide layer has been formed, very little leakage current flows. However, over time, this micro-thin layer is corroded by the electrolyte and needs to be re-formed. Provided that the capacitor always has DC applied across it, the capacitor balances itself by always passing the minimum necessary anodising current to maintain the oxide layer for the applied DC voltage.

The author selected a number of 470 µF electrolytic capacitors having as widely different technologies as possible that were known to have been stored uncharged for at least 10 years and tested them by applying a constant voltage from a low-noise supply and measuring charging current at 1-s intervals whilst logging the result directly to a spreadsheet. All the capacitors produced a very similar characteristic curve (see Figure 4.10).

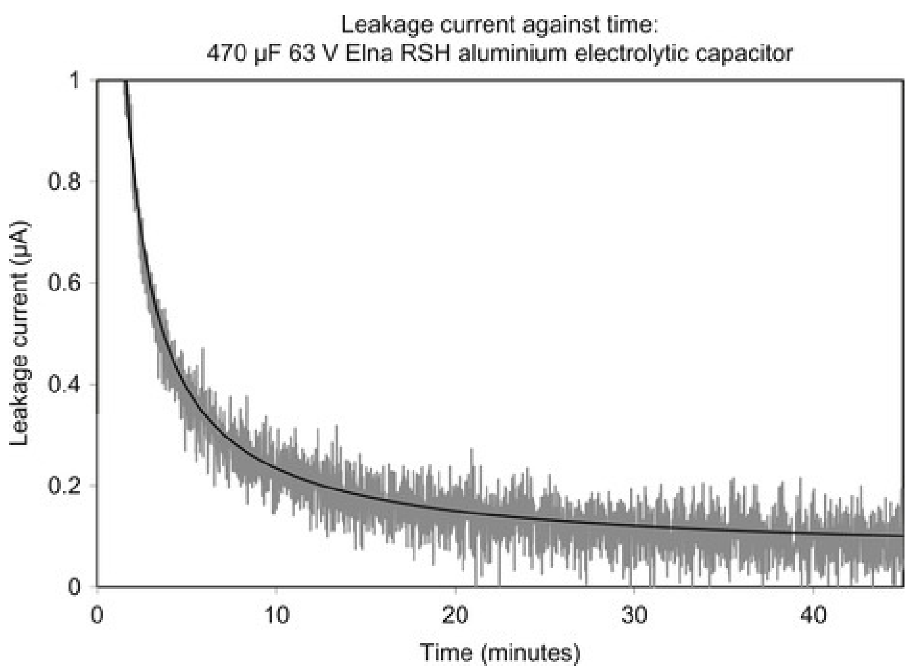


Figure 4.10 Electrolytic capacitor 'leakage' current decays with time.

The measured current is clearly the sum of a slowly decaying DC term and a pure AC noise term, so the author experimented to find a model that would fit the decaying DC term. The ideal model would fit the decaying DC term so well that when subtracted from the measured results, the remainder would be pure noise (no DC component) (see Figure 4.11).

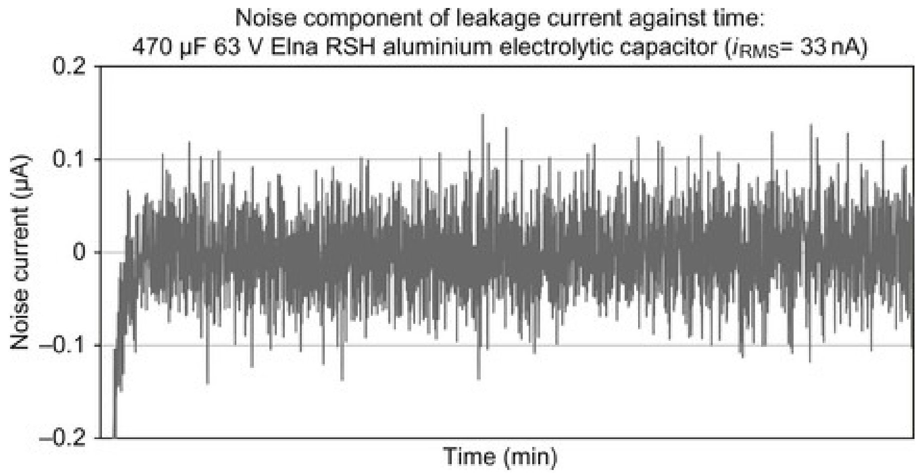


Figure 4.11 Once the decaying current is subtracted, pure noise remains.

Given that the remainder *is* a close approximation to noise, the relationship used to fit the decaying DC term can be used with confidence. More importantly, the decaying DC term can be deemed to be the anodising current referred to earlier – and this hypothesis was verified by repeating the experiment with a 120 μF 400 V polypropylene capacitor and noting that the DC term decayed in a few seconds rather than many minutes (see Figure 4.12).

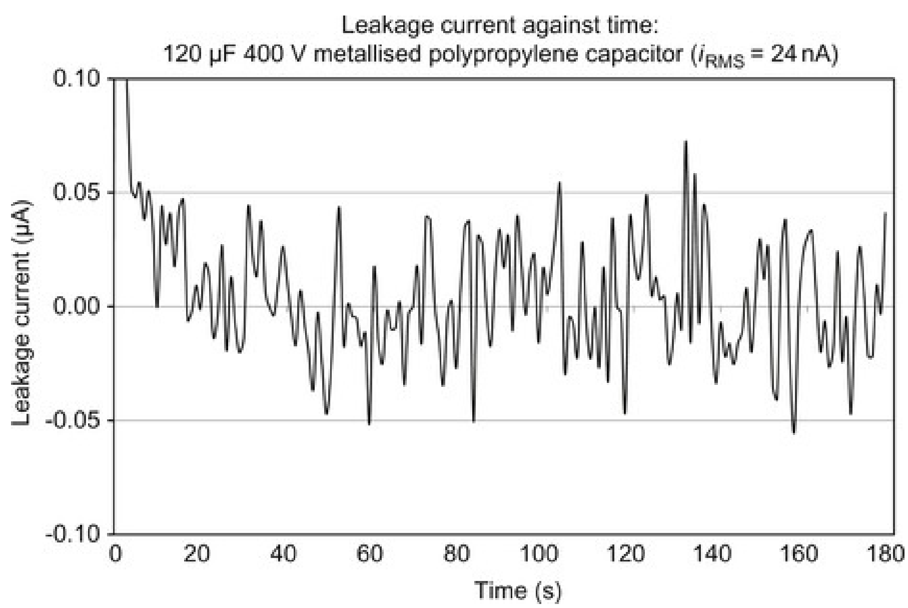


Figure 4.12 A polypropylene capacitor shows no decaying current.

Electrolytic capacitor anodising current can be fitted using an equation of the form:

$$I_{\text{anodising}} = \frac{a}{t^b} + c$$

where

t=time in minutes

a, b and c=experimentally determined constants.

The author's measurements suggest that a typical aluminium electrolytic requires at least 45 min after voltage is applied before the anodising current decays to its steady-state value. Capacitor manufacturers neglect the (very variable) time-dependent term and refer to anodising current as leakage current, and then specify its final value (' c' in the anodising current equation) in terms of the CV product of the capacitor. An electrolytic in good condition has a leakage current comprised of the DC anodising current plus a typical AC noise current of ≈50 nA $_{RMS}$.

Given that the model for anodising current was good, a comparison of the modelled results was made (see Figure 4.13).

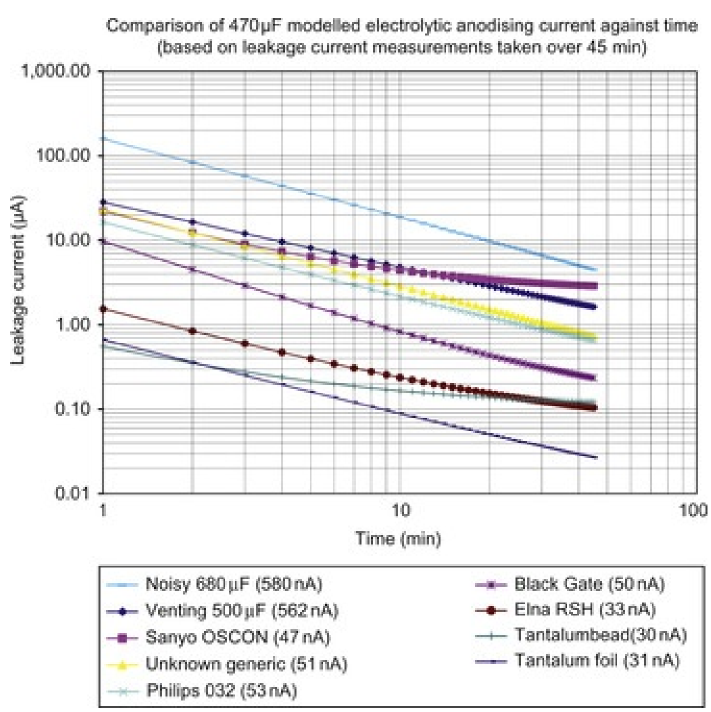


Figure 4.13 Comparison of different electrolytic capacitor 'leakage' currents.

The current following the descriptor in the graph is the RMS noise current calculated from the measurements, and it is notable that the capacitors with the highest leakage current displayed the highest noise current and vice versa – if low noise is required, low leakage current should be specified.

Unsurprisingly, the two capacitors that were known to be noisy also had the highest anodising currents, and the unknown generic capacitor displayed predictably mediocre performance. The Sanyo OS-CON had a surprisingly high anodising current; perhaps the organic semiconductor electrolyte that allows its very low ESR is a little more corrosive, and its slightly higher noise current of 53 nA $_{RMS}$ suggests that its use as a cathode bypass in a low-noise stage might be unwise. The added voodoo of the Black Gate capacitor (noise current, 50 nA $_{RMS}$) didn't prevent its performance being surpassed by the Elna RSH.

Once equipment is switched off, when power is reapplied, a higher than normal anodising current must flow until the oxide layer has been re-formed. The longer the elapsed time without bias volts applied, the greater duration and amplitude this initial anodising current must be, and there is a danger that it will cause serious heating of the electrolyte. As the electrolyte is heated, it evaporates more readily, and the resulting gas may build up sufficient pressure to cause the can to explode. Because of this, it is wise to use a Variac to gently apply power to equipment containing electrolytic capacitors that has lain idle for some time.

Modern capacitors have safety pressure seals to vent the gas via a rubber bung in the base of the component (large capacitors), or the aluminium can may be deliberately weakened at the top with a series of indentations that allow controlled rupturing for the gas to escape (small capacitors). Either of these occurrences signifies the demise of the component, but they do prevent damage to other components, with the bonus of a simple visual inspection of their health. The leaky 500 μF capacitor not only had a measured noise current of 560 nA $_{RMS}$, but also was visibly vented (see Figure 4.14).

Figure 4.14 Note the bursting pimple that warns of a leaking electrolytic capacitor.

Gentle heating evaporates the electrolyte through the seals of the capacitor (no seals can be perfect), and as the quantity of electrolyte falls, it makes less and less contact with the nooks and crannies of the etched plate, so the ESR rises and capacitance falls. Thus, ESR can sometimes be a useful guide to the health of an electrolytic capacitor, and dedicated ESR meters are available (Peak Atlas in the UK and Dick Smith kit in Australia), but more significantly, they are useful for giving a value of ESR that can be fed into PSUD2 (see Chapter 5), and although they are not quite as accurate as a genuine swept frequency component bridge, they are far cheaper.

Electrolyte evaporation causes electrolytic capacitors to be heat sensitive, so capacitor life doubles for every 10 °C drop in temperature.

Applied voltage also affects electrolytic capacitor life. Without a bias voltage, the oxide layer cannot be re-formed and is gradually corroded by the electrolyte, causing the capacitor to become leaky. This was a well-known fault in analogue sound mixers using symmetrical + and − supplies with operational amplifiers coupled by electrolytic capacitors, which then had little or no polarising voltage.

Provided that there is a polarising voltage, operating electrolytic capacitors below their maximum rated voltage increases their life significantly:

$$\text{Life}_{(\%\ \text{of rated})} \geq \left(\frac{V_{\max}}{V_{\text{applied}}}\right)^5 \times 100\%$$

Using this relationship, we see that operating an electrolytic capacitor at 87% of its rated voltage doubles its life. It is wise not to read too much into this formula, since we could easily use it to predict a lifetime measured in centuries by lowering the operating voltage sufficiently. A good engineering rule of thumb is that, if possible, electrolytic capacitors should be operated at two-thirds of their maximum voltage rating, giving a theoretical eight-fold increase in life expectancy, which is probably at the limit for which the formula is valid. Fortunately, all of the capacitor manufacturers offer detailed datasheets that enable far more accurate lifetime predictions to be made taking into account such factors as ambient temperature, ripple current and applied voltage.

Many classic valve amplifiers had electrolytic capacitors with more than one component concentrically wound in a single can. The outer capacitor was marked with a red spot, and in an amplifier using cascaded RC smoothing, this capacitor should be connected to the most positive potential. The logic for this is that the highest potential has the greatest ripple voltage, and as there is no field within a conductor, this ripple is not coupled to subsequent stages. Wiring the capacitors in the wrong order causes excess hum.

Historically, electrolytic capacitors had poor tolerance on their capacitance (typically +100% to −50%). Although modern types are typically ±10%, we should *never* use an electrolytic capacitor in a position where its value could not be safely doubled or halved without upsetting operation of the circuit.

There is a class of aluminium electrolytics available for use at AC that is known as *bipolar*. These capacitors used to be ubiquitous in loudspeaker crossovers because they were so much cheaper than plastic-film capacitors of comparable capacitance. Their construction is effectively two electrolytic capacitors back to back (see Figure 4.15).

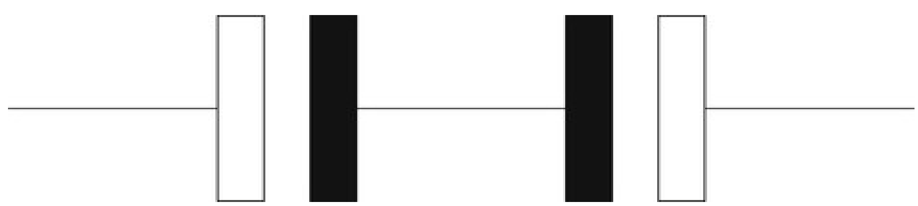


Figure 4.15 The bipolar electrolytic capacitor.

There is no constant polarising voltage, and each individual capacitor has to be twice the value of the final capacitor. Defects are thereby multiplied by a factor of four over the normal unipolar electrolytic capacitor, so their performance is poor.

Tantalum Electrolytic Capacitors ($\varepsilon_r \approx 25$)

The increased relative permittivity significantly reduces component volume compared to aluminium electrolytic capacitors ($\varepsilon_r \approx 8.5$). Tantalum foil capacitors have two additional advantages directly related to the increased chemical inertness of the tantalum oxide layer. Firstly, ESR can be reduced by using a lower resistance electrolyte that would have corroded aluminium foil. Secondly, because the oxide layer is more inert, leakage current is reduced. However, tantalum is expensive, whereas aluminium electrolytic capacitors are improving all the time.

Tantalum bead capacitors are only available in low voltages and are often specified by semiconductor manufacturers for bypassing voltage regulators or logic chips, but there is a very large variation in ESR between types – an ESR meter is essential here. Unfortunately, they are only available in low values (rarely >100 μF) and are often not large enough to be used as cathode bypasses. Tantalum bead capacitors have extremely limited ripple current capacity, so they're not a good choice for power supply coupling where switchers are involved. When tantalum bead capacitors fail (they have zero tolerance to reverse polarity), they tend to fail dead short circuit, and the consequential damage tends to be spectacular. They are expensive.

The main justification for using tantalum capacitors is that (unlike aluminium electrolytic capacitors) they have almost indefinite shelf life, a useful property for rarely used instruments that must be ultra-reliable, but less useful for audio.

Variation of Capacitance with Frequency

Some plastics are *polar*; this does not mean that the capacitor can be damaged by reversing the polarity of any applied DC, but that at a molecular level within the dielectric, there are permanently charged electric dipoles, similar to the magnetic dipoles in a magnet. Under the influence of an external electric field, these dipoles attempt to align themselves to that electric field. By contrast, non-polar dielectrics have very much smaller losses and are very nearly perfect at audio frequencies. Almost all dielectrics with ε_r>2.5 are polar (Table 4.2).

Table 4.2 Comparison of Capacitor Dielectrics and '*d*'

Dielectric	Common name	ε_r	Typical 1 kHz '*d*' (100-nF capacitor)	Polar?
Polytetrafluoroethylene	PTFE, Teflon™	2.1	0.00001	Y[a]
Polystyrene		2.6	0.00020	N
Polypropylene		2.2	0.00010	N
Polycarbonate		3.2–3.0	0.00090	Y
Polyethylene terephthalate	PET, polyester	3.2–3.9	0.004	Y

[a]Although PTFE *is* polar, the balanced dipoles of its polymer chain make it appear non-polar.

Because work must be done each time a dipole is flipped, capacitors having polar dielectrics suffer frequency-dependent losses that are reflected in capacitance that varies significantly with frequency (see Figure 4.16).

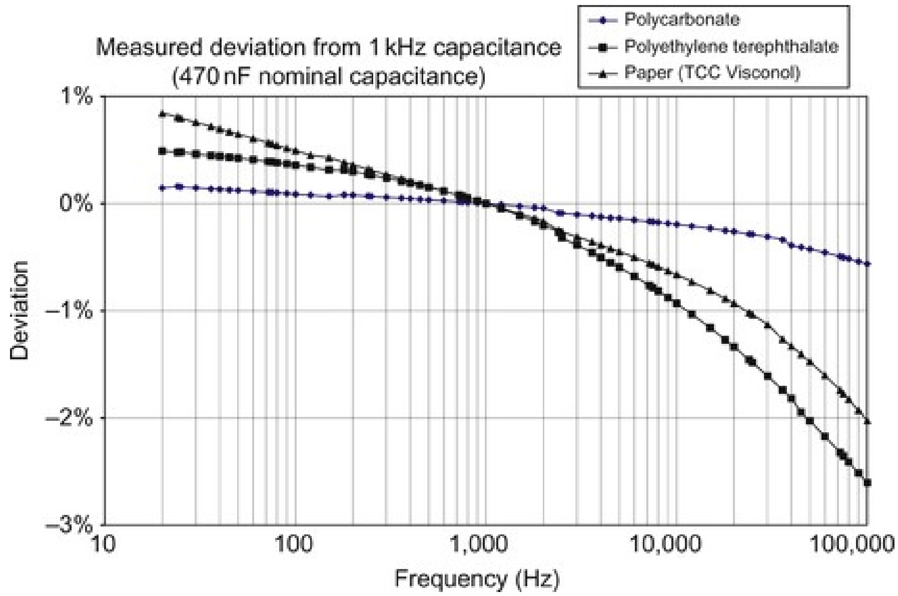


Figure 4.16 Capacitors having a polar dielectric do not have constant capacitance with frequency.

As can be seen, the use of paper or polyester capacitors in valve amplifiers can only be justified in areas such as power supply decoupling where their capacitance deviation with frequency is irrelevant. Although not shown on the graph, electrolytic capacitors are even worse, and their 1 kHz capacitance is typically −10% compared to their DC capacitance.

Imaginary Capacitance

Since permittivity has a real and imaginary component, capacitance must also have real and imaginary components. If we measured the series real component of a capacitor's impedance over a range of frequencies, we would expect to see a low and constant value equal to the capacitor's ESR, but what we actually see is a low frequency loss component due to the imaginary capacitance that falls to the capacitor's ESR at high frequencies (see Figure 4.17).

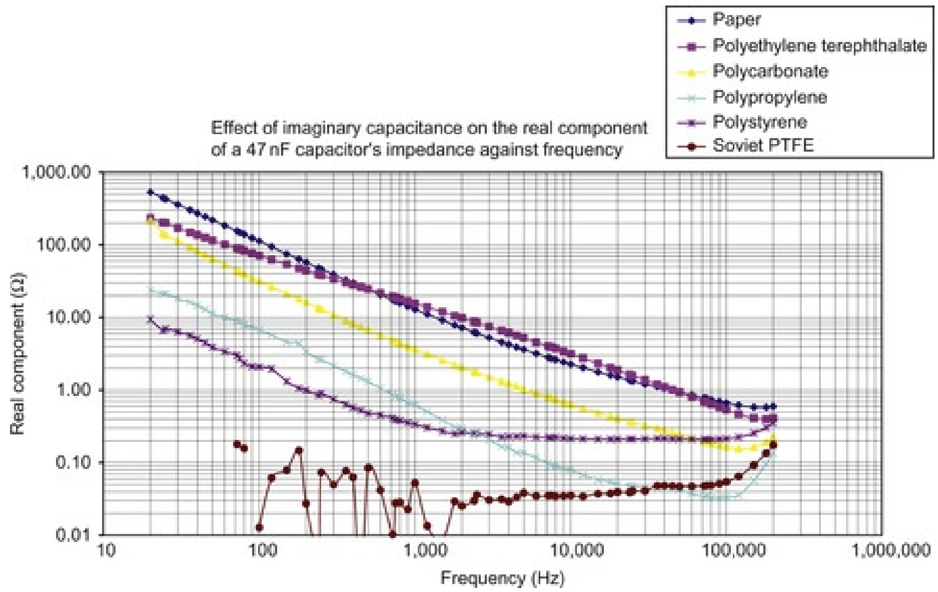


Figure 4.17 Imaginary capacitance causes a frequency-dependent loss at low frequencies.

Imaginary capacitance is in series with the real capacitance and is ideally much larger (see Figure 4.18).

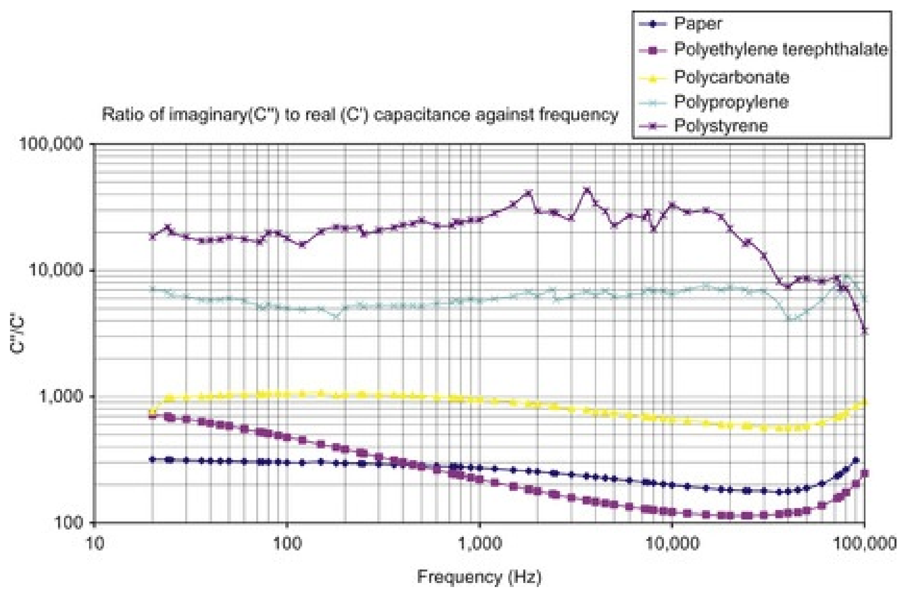


Figure 4.18 Imaginary capacitance should ideally be large and constant with frequency.

PTFE is not shown on the imaginary capacitance graph because it is simply too good for the author's bridge and even the data for polystyrene are questionable. Nevertheless, it can be seen that not only do the better dielectrics have a higher ratio of imaginary to real capacitance, but it is nearly constant over the audio bandwidth. More significantly, it should be noted that the measurement of imaginary capacitance correctly shows polystyrene to be a superior dielectric to polypropylene, whereas the 1 kHz comparison of ‘ *d*’ in Table 4.2 implied quite the reverse.

General Considerations in Choosing Capacitors

Voltage Rating

- Will the voltage across the capacitor change polarity, or is it simply a varying DC? If the voltage across the capacitor is AC, then conventional electrolytic capacitors are eliminated.
- Will the capacitor have an acceptable predicted lifetime with the applied DC voltage plus the maximum expected signal voltage (V_{pk}, not $V_{average}$)?
- Could the capacitor withstand the maximum possible High Tension (HT) voltage? If not, what arrangements have been made to ensure that its rated voltage is never exceeded?

Capacitance Value

- Is the *absolute* value important? If it is part of a filter or equalisation network, then we need a close tolerance component whose capacitance does not vary with frequency, and only air, PTFE, polystyrene, polypropylene (now available in ±1%) or silvered mica will do. Polycarbonate would once have been a marginal candidate, but now that it is obsolete no longer need be considered.
- *Matching*: Is the capacitor part of a pair, such as coupling capacitors in a push–pull amplifier, or the corresponding component in the other stereo channel? If it is, then the capacitors should be matched if possible.
- Each family of capacitor dielectric is only available in a limited range of values, and if we need 330 μF, then only an electrolytic can provide this value at a sensible price and size.

Heat

Will the capacitor become warm? Will the consequent change in value matter? In general, capacitors should not be operated at more than 50 °C (because the resistance of all insulators falls with increasing temperature), but even this ambient temperature could reduce the life of an electrolytic capacitor unacceptably.

ESR

ESR

The reservoir capacitor in a capacitor input supply has to pass significant ripple current that causes self-heating, raising the internal temperature above ambient – this is why electrolytic capacitors have a ripple current rating.

Leakage and ' d'

- Is leakage important? A cathode bypass or HT smoothing capacitor may be allowed to pass a small leakage current. A grid coupling capacitor may not be allowed to be leaky under any circumstances.
- Is this component important for final sound quality? Capacitors in the obvious signal path are important, but the signal *current* has to return through the HT supply, so HT smoothing and bypass capacitors are equally important. Smoothing capacitors for bias circuitry may be less important if there is no audio signal on them.

Microphony

All capacitors are microphonic to greater or lesser extent. The reason for this is very simple. Suppose that we have stored a fixed charge on a capacitor:

$$Q = CV$$

The capacitance of a parallel plate capacitor is:

$$C = \frac{A \cdot \varepsilon_0 \cdot \varepsilon_r}{d}$$

Combining these equations, and solving for *V*:

$$V = \frac{Qd}{A \cdot \varepsilon_0 \cdot \varepsilon_r}$$

Since Q, A, ε_0 and ε_r are constants, if we vary the spacing between the plates, the voltage across the capacitor *must* change. This principle is the basis of the capacitor microphones used in studios and the ubiquitous electret microphone found in portable recorders.

The principle is reversible, and varying the applied voltage across a capacitor alters the attractive forces between the plates, and if the plates are free to move, this causes vibration. This is the basis of the electrostatic loudspeaker.

It might be thought that the plates of a plastic film capacitor would be sufficiently tightly wound that no movement was possible, but the author once built a stabilised HT power supply where the output bypass capacitor whistled loudly at ≈2 kHz. The circuit was thereby diagnosed as unstable even before the oscilloscope was ready!

The problem of capacitor microphony can be tackled in three ways, listed in descending order of desirability:

- Avoid using capacitors. To a limited extent, this is feasible.
- Isolate capacitors from vibration. Capacitors carrying low-level signals will be proportionately more sensitive to microphony than those carrying high-level signals. Pre-amplifiers are therefore most susceptible, and it is well worth isolating them from vibration. This is easily done at the design stage, but is much harder once built.
- Capacitors are physical objects, so they have mechanical or acoustical resonances. If we excite these resonances, we should expect them to be audible in the same way that striking a tuning fork produces an audible note. If we mechanically damp the capacitor by gluing it to another surface, then these resonances will be reduced. Provided that the capacitor can survive the heat, the soft glue used by hot-melt glue guns is ideal.

There is no reason why we should not use all three methods in combination if it seems that microphony is likely to be a problem. A good test for microphony is to tap each component with a plastic pen (to avoid shock risk) with the equipment turned on whilst listening to the loudspeaker. The results may surprise you!

Bypassing

All modern capacitors (whether plastic or electrolytic) have series inductance of between 10 nH and 100 nH, so their impedance becomes inductive as frequency rises. Early capacitors had much higher series inductance, so it was worthwhile bypassing a large capacitor with a small one in order to maintain low impedance at high frequencies, but modern electronics operates at higher frequencies, forcing capacitor design to improve, rendering the technique superfluous.

Wires have inductance, and although this is not really linear with length, 0.75 nH/mm is a good working approximation, so there is no point in spending money on a filter capacitor having only 10-nH series inductance if it is then connected to the circuit via a pair of wires 100 mm long (adding ≈150-nH inductance). We may not be able to connect the filter capacitor directly between the output transformer HT tap and the cathode returns of the output valve(s), but we can, and should, connect a low-ESR bypass capacitor between those points using the shortest wires possible (to minimise its series inductance) (see Figure 4.19).

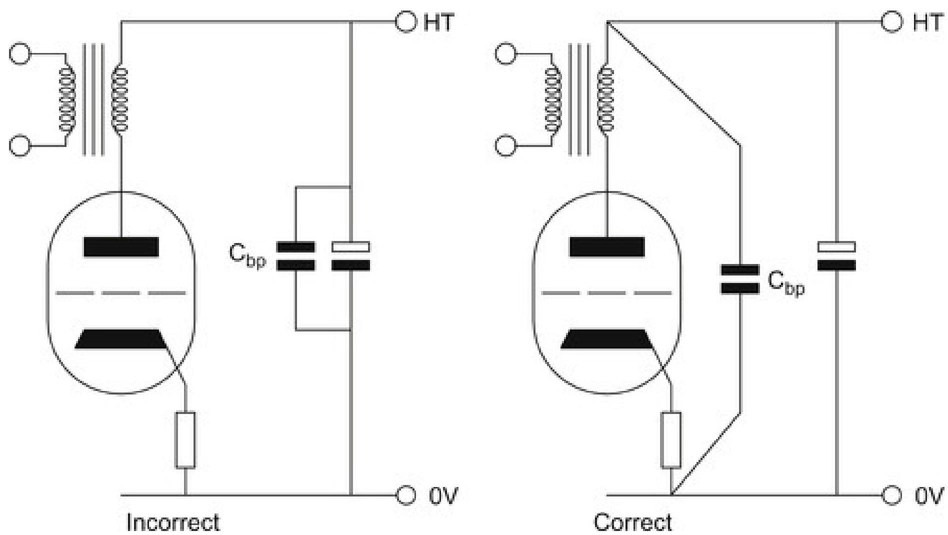


Figure 4.19 Connection of a bypass capacitor.

Magnetic Components

Magnetic components include transformers and inductors. Transformers may be signal transformers, such as output transformer and moving coil step-up transformers, or they may be mains transformers. Inductors may be the small-signal inductors used in filters, or they may be the power chokes in an HT supply.

Magnetic components are easily the least perfect *passive* components (resistors, capacitors and inductors/transformers), and for this reason many designers shun them. This is unfortunate because it seriously restricts design choice.

Inductors

Inductors store energy in the form of a magnetic field. Any wire passing current generates a magnetic field, so it must possess inductance. We can deliberately increase this inductance by winding the wire into a coil, whilst placing the coil around an iron core increases inductance still further. These proportionalities can be expressed by:

$$L \propto \frac{\mu_0 \mu_r A N^2}{l}$$

where

L=inductance

μ_0=permeability of free space=$4\pi\times10^{-7}$ H/m

μ_r=relative permeability of the core magnetic material

A=magnetic path cross-sectional area

l=magnetic path length

N=number of turns.

Relative permeability is the magnetic analogue of relative permittivity that we met earlier and has a value of 1 for air and ≈5,500 for iron. The magnetic path length is the length through the core back to the starting point, and the cross-sectional area of the magnetic path is simply the cross-sectional area of the core, so it ought to be easy to derive a useful equation for calculating inductance.

Unfortunately, μ_r varies hugely with flux density, the path length is easily affected by air gaps, and some flux escapes the core. We will look at each of these problems in more detail in a moment, but suffice to say that we cannot often accurately calculate the inductance of a coil. We are forced to make an informed guess, add a few turns, measure the inductance under the actual operating conditions, and then remove turns until the desired inductance is achieved.

A curve that appears in virtually every discussion of magnetic materials is the *B*/*H* curve. This is a plot of the relationship between applied magnetising force and resulting magnetic flux. For our purposes, we need only note that μ_r is proportional to the gradient of the curve and that as the gradient changes with level, so must μ_r (see Figure 4.20).

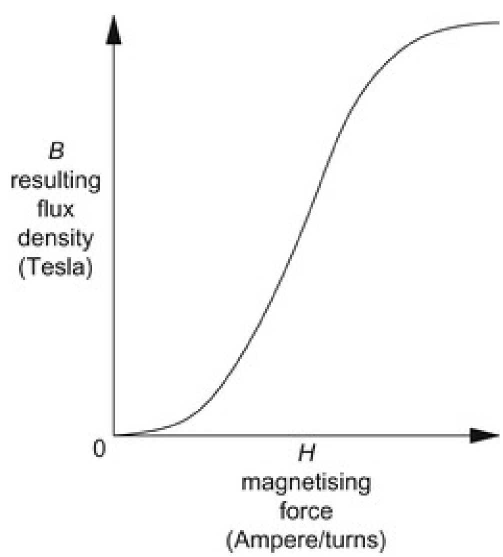


Figure 4.20 *B*/ *H* curve: Non-constant slope implies non-constant μ.

Air-Cored Inductors

We can completely avoid the problem of non-constant μ_r with level by not using a magnetic material in the core. *Air-cored* inductors have constant inductance with applied signal level, and do not therefore cause distortion, making them popular in high quality loudspeaker crossover networks. Determining the magnetic path area is more difficult since this theoretically extends to infinity, whilst the path length is similarly awkward to define. Nevertheless, formulae have been produced for various core geometries, and a particularly useful set of formulae for optimum (lowest) resistance air-cored copper wire coils based on a paper by A.N. Thiele [4] follows:

$$R = 8.01 \times 10^{-3} \sqrt[5]{\frac{L^3}{d^8}}$$

$$N = 10.2 \sqrt[5]{\frac{L^2}{d^2}}$$

$$c = \frac{d\sqrt{N}}{0.841}$$

$$l = 0.188\sqrt{Lc}$$

where

R=resistance in Ω

L=inductance in μH

d=diameter of wire in mm

N=number of turns

c=core radius (see Figure 4.21)

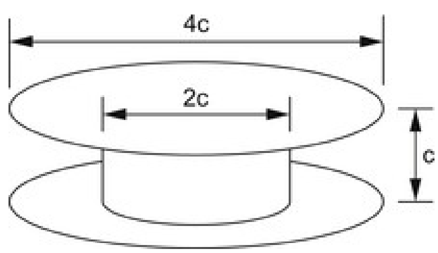


Figure 4.21 Relative bobbin dimensions for air-cored coils using Thiele formulae.

l=length of wire in m.

The formulae are given in this modified form because wire is available only in a range of standard diameters, and the resistance of the coil is not usually critical. If the resistance is different from that wanted, then a different wire diameter can be tried. The best way to use these equations is to drop them into a spreadsheet and experiment. If the spreadsheet has standard wire gauges, so much the better.

Experimentation soon reveals that air-cored inductors have significant resistance or that they are very large. This problem of resistance is common to all inductors, and is one of their imperfections. Air-cored inductors are useful not only for loudspeaker crossovers, but also for the output filters of Digital to Analogue Converters (DACs), where the resistance is far less of a problem.

It should be noted that because of practical considerations (winding efficiency, variable wire diameter, etc.) the formulae *cannot* give exact answers, and it is therefore wise to design 5% oversize, and then remove turns whilst measuring the inductance with a component bridge.

Many component bridges use a 1 kHz internal oscillator. When measuring air-cored coils, the inductive component can easily be swamped at low frequencies by the relatively high resistance, causing the bridge to give misleading results. If it is possible to feed such a bridge from an external source of AC, it should be fed with the highest frequency that the bridge manufacturer allows (typically 20 kHz), and this will allow sensible measurements to be taken.

Gapped Cores for AC Only

One way to reduce resistance without suffering gross distortion is to use a coil with a magnetic core that has an air gap. A *gapped* core significantly increases inductance over an air-cored coil, but because the air gap is a large magnetic reluctance (analogous to magnetic resistance) in the path of the magnetic flux, it swamps the variation in permeability of the much smaller magnetic resistance of the core, and inductance is more nearly constant. As the gap becomes larger, inductance falls, and if the gap is infinitely large, then we are back to an air-cored coil. This technique was used for many years by the BBC Research Department for inductors in passive loudspeaker crossovers.

We may unintentionally make an inductor with a gapped core. Many ferrite cores used for small inductors are supplied as two mating halves that fit around the core once wound. Dust on the mating surfaces causes an air gap, and if the cores are gently squeezed together whilst inductance is measured, a significant increase in inductance will be seen.

Gapped Cores for AC and DC (Power Supply Chokes)

If an inductor has to pass DC, it is essential that the DC current should not saturate the core, since this would drastically reduce the inductance. Iron-cored inductors passing DC are invariably gapped in order to maintain their inductance up to their rated maximum current.

However, a power supply choke must accommodate a different set of compromises to a loudspeaker crossover choke. Specified inductance must be delivered at the rated DC current, but inductance variation over the rest of the range is not only permissible but also expected. Remembering the *B*/ *H* curve, we can expect inductance to fall as DC current rises and saturation approaches (see Figure 4.22).

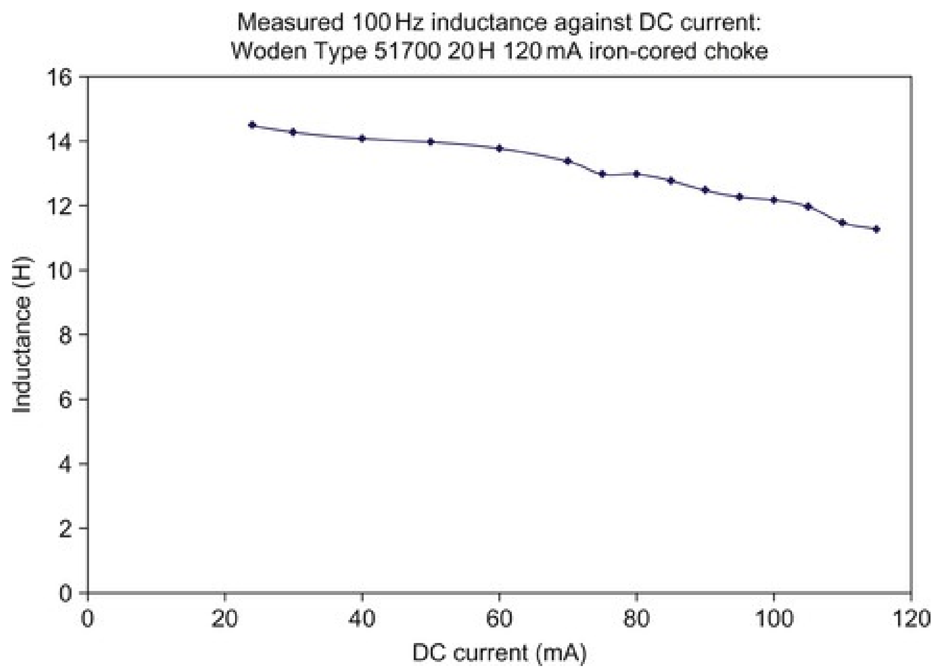


Figure 4.22 Real inductors have inductance that falls with applied DC current.

Not only does measured inductance fall with DC current, but inductance also changes with AC excitation. Most component bridges only provide between 100 mV_{RMS} and 1 V_{RMS} excitation, but power supply choke manufacturers expect their products to see rather more hum voltage across the products, and they design for those conditions, so a conventional component bridge might see a smaller inductance (see Figure 4.23).

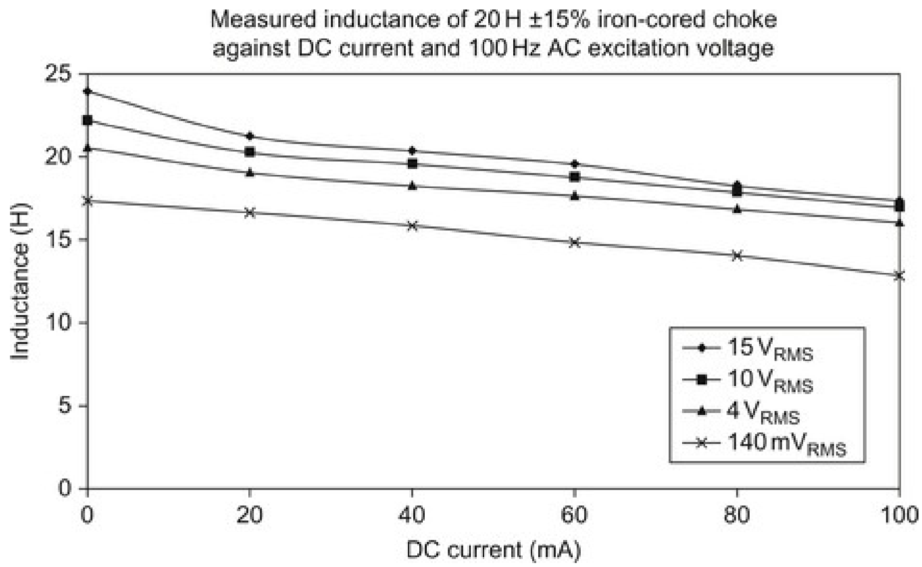


Figure 4.23 Real inductors have inductance that changes with applied AC excitation.

Note that the choke only meets its specified inductance when a representative AC excitation is applied – beware the misleading results produced by low-level AC excitation.

Self-Capacitance

If a coil is made of many turns of wire and there is a potential between different turns and layers of turns, then we must expect the inductor to have capacitance in parallel with its inductance (see Figure 4.24).

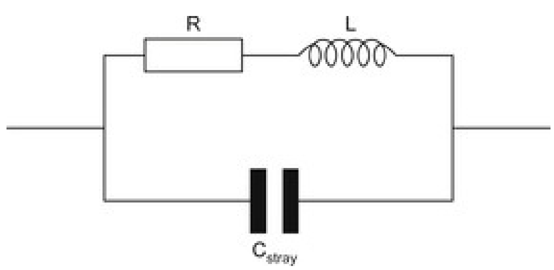


Figure 4.24 Equivalent circuit of practical inductor.

We now have our familiar resonant circuit, which means that as we rise above the choke's self-resonant frequency, the capacitor begins to reduce the choke's impedance and reduce its filtering effect (see Figure 4.25).

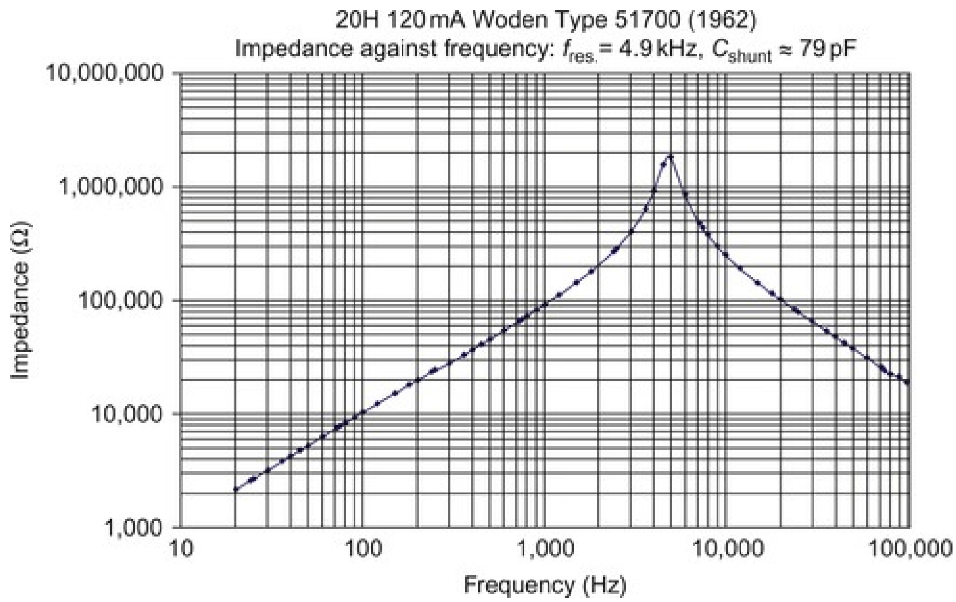


Figure 4.25 Impedance of a choke against frequency.

Note that although the filtering effect declines past the resonant frequency of 4.9 kHz, the impedance is still 100 kΩ at 20 kHz, and in combination with a 68 μF electrolytic capacitor having an ESR of 0.5 Ω, this would theoretically give 105 dB of attenuation. Obviously, the higher the self-resonant frequency, the better it will be, and the best way to determine this frequency is to measure phase (see Figure 4.26).

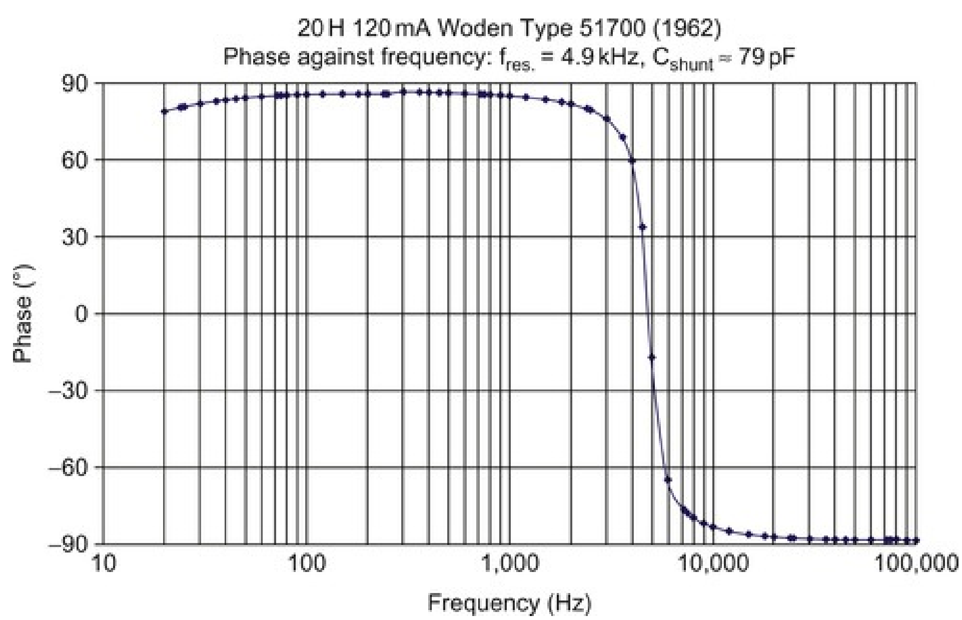


Figure 4.26 Phase changes abruptly at resonance.

As Figure 4.26 showed, phase changes very sharply at resonance, and the easiest way to measure phase is to arrange a circuit that generates a Lissajous figure on an analogue oscilloscope (see Figure 4.27).

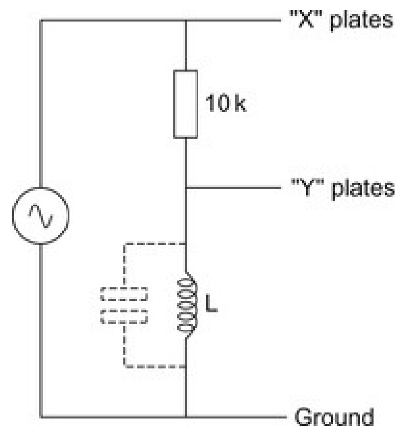


Figure 4.27 Connection to an analogue oscilloscope for a Lissajous figure to find the self-resonant frequency of inductor.

The oscilloscope is used in *XY* mode, and as the frequency of the oscillator is varied and approaches the choke's self-resonant frequency the Lissajous figure changes from an ellipse to a straight line. The resonant frequency can now be read directly from the oscillator or measured using a counter (even cheap DVMs now offer frequency measurement to sufficient accuracy for this measurement).

Transformers

In a perfect transformer, the magnetic flux of the primary winding is coupled to the secondary winding with no loss whatsoever. Practical transformers are somewhat different.

In a transformer, the losses are often divided into two distinct groups: *iron losses*, so called because they are due to imperfections of the core, and *copper losses*, so called due to imperfections of the windings.

Iron Losses

The primary winding on the transformer has a finite winding inductance, so it presents a reactance across the supply that draws a current even when there is no secondary load. Rather than designing for a specific primary reactance or consequent current, older transformers simply used 'eight turns per volt' although many contemporary iron-cored transformers (particularly toroids) use only four turns per volt.

As the core is successively magnetised and demagnetised through opposite polarities, work has to be done to change the alignment of the magnetic dipoles. This loss is known as *hysteresis* loss, and it may be calculated by investigating the hysteresis curves for the particular core material used. Because it is the loss caused by changing the core magnetisation through one complete cycle of the applied AC waveform, there will be more loss in a given time if more cycles of magnetisation are traversed. Hysteresis loss is therefore directly proportional to frequency and can only be reduced by choosing a lower loss core material.

Magnetic cores are metal, and therefore conduct electricity. As far as the primary winding is concerned, there is no distinction between an intentional secondary winding connected to a load and a conductive path parallel to the primary winding through the core. Conductive paths through the core cause *eddy* currents to flow, which, because they are short circuits, cause losses. To reduce these losses, the core can be constructed from a stack of *laminations* that have had their surfaces chemically treated to make them insulators. The ultimate approach to this problem is to make the core of iron dust particles whose surface has been treated, and then bond these with a ceramic to form a solid core known as a *ferrite dust* core.

Eddy current loss is proportional to f^2 because not only is the loss proportional to the number of traverses of the magnetisation loop in a given time, but also higher frequencies have smaller wavelengths and allow more loops of current to form within the core. Although thin steel laminations are satisfactory for audio frequencies, ferrites are necessary for RF frequencies, and at VHF almost all core materials are excessively lossy, and air-cored transformers must be used.

Primary currents due to finite primary inductance, hysteresis loss and eddy current loss are often combined and known as magnetising current in power transformers and are responsible for core heating even when a load is not connected.

Not all of the flux from the primary flows through the secondary winding, and this loss, combined with hysteresis and eddy current loss, is known as *leakage inductance* in audio transformers. Theoretically, leakage inductance (referred to the primary) is found by measuring the primary inductance with the secondary short circuited. In practice, leakage inductance is awkward to measure because measurement at a single frequency is easily skewed by stray capacitances, necessitating a swept frequency measurement. Nevertheless, leakage inductance is an important theoretical concept, since it determines the high frequency operating limit of the transformer.

Leakage inductance is dependent on the size (q), the turns ratio N^2 and the geometry of the transformer (k), but is independent of μ_r:

$$L_{\text{leakage}} \propto qN^2k$$

For a given frequency, a higher power rating transformer will be larger than a lower power rating transformer and will consequently have higher leakage inductance.

Since leakage inductance is proportional to N^2, we should always: try to keep the turns ratio as low as possible, so paralleling output valves in a valve amplifier is beneficial because it reduces the turns ratio required.

Geometry can be improved in two fundamental ways: we can either improve the shape of the core, or improve our winding technique.

Standard transformers are made with E/I cores, where each lamination of the core is composed of an E shape and an I shape. A machine that looks (and sounds) rather like a card dealer inserts laminations alternately from either side of the coil so that, on alternate laminations, the orientation of the shapes is reversed to reduce the air gap at the joint (see Figure 4.28).

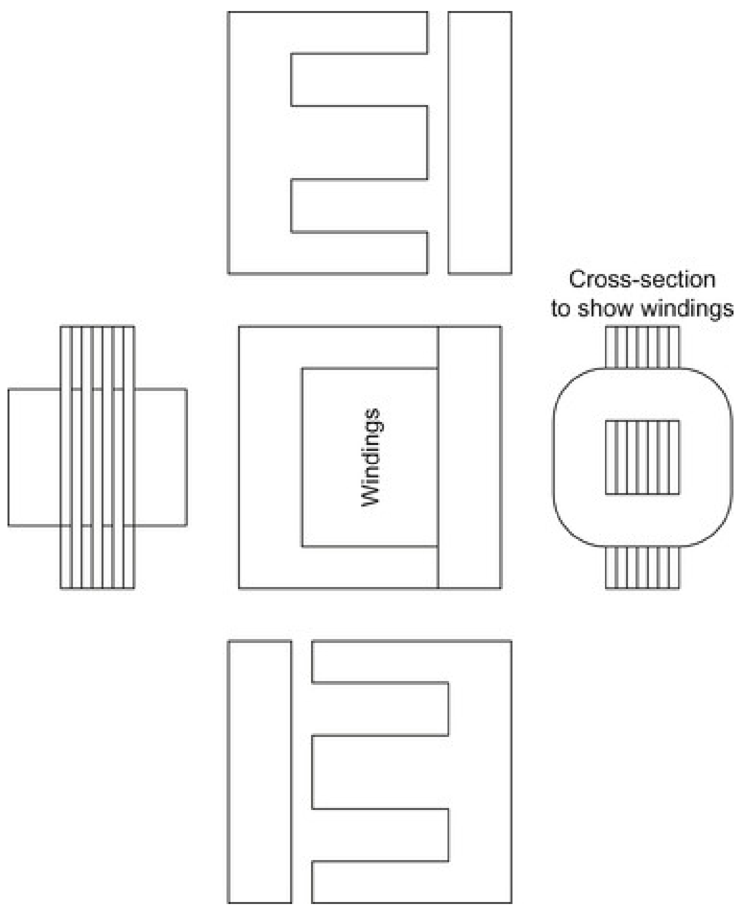


Figure 4.28 E/I core laminations' arrangement to reduce leakage flux.

Traditionally, superior cores were made as C cores. These were made by winding the core out of a continuous strip, which was then cut in half, and the resulting faces ground smooth. The coils were then wound, and the cores were inserted so that the ground faces were perfectly aligned with minimal air gap, and steel straps were used to hold the assembly firmly together (see Figure 4.29).

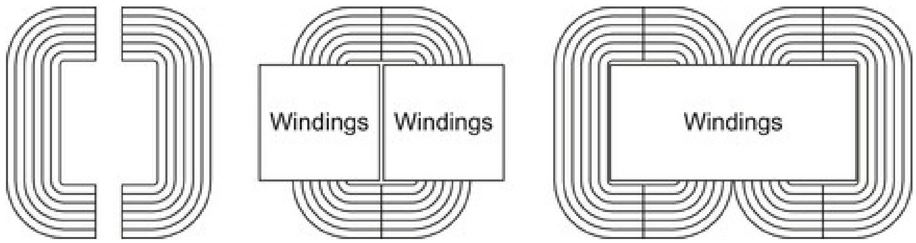


Figure 4.29 C-core arrangements.

The C core was an expensive process, and inaccurate assembly could create an air gap, thus creating the very imperfection that the design intended to avoid. The more modern approach is to wind the core as a toroid, but not cut it, and use a special coil winding machine to wind the coils directly onto the core, resulting in a very low leakage core (see Figure 4.30).

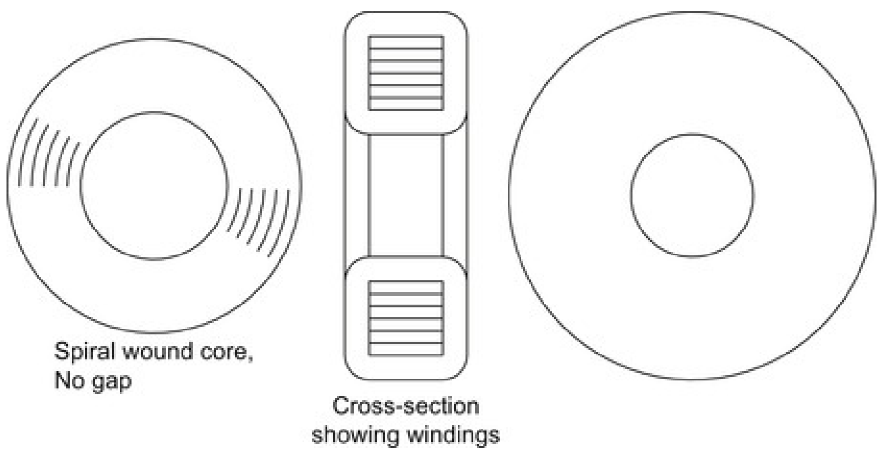


Figure 4.30 Toroidal core arrangement.

Incidentally, although toroids are thought of as being modern, the first transformer ever made was a toroid, using wire insulated with silk from his wife's wedding dress! (Michael Faraday, August 1831).

Both the C core and the toroid have the further advantage that the magnetic flux always flows in the same direction relative to the grain direction of the crystal structure of the core, whereas in the E/I core it has to flow across the grain in some parts of the core. This is significant because Grain Oriented Silicon Steel (GOSS) can tolerate a higher flux density before saturation in the direction of the grain than across the grain. E/I cores can therefore only operate at flux densities below saturation across the grain, whereas C cores and toroids can operate at significantly higher flux densities, allowing core size and turns per volt to be reduced.

The worst winding geometry for leakage inductance is split chamber (see Figure 4.31).

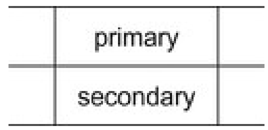


Figure 4.31 Split bobbin gives good primary/secondary isolation but high leakage inductance.

The geometry of a transformer can be improved by winding the primary and secondary out of many interleaving layers or sections, rather than winding one half of the bobbin with the primary and the other half with secondary. Increasing the number of sections improves the coupling between primary and secondary, thus reducing L_{leakage}, but usually increases stray capacitance.

Although sectioning the windings is relatively easy on an E/I or C core, it is very difficult on a toroid; moreover, winding geometry on a toroid is quite poor, and so it is easy to lose the benefits of the improved core by having a poor coil. Toroidal mains transformers are notorious for their leakage flux at the point where the windings exit for this very reason.

An alternative technique for improving winding geometry is to use *bifilar* winding, whereby two wires are simultaneously wound side by side. If one of these wires is part of the primary and the other is part of the secondary, this promotes excellent coupling between the windings, and leakage inductance is significantly reduced. The technique is cheaper than sectioning, and providing the coil winding machine can cope, there is no reason to stop with two wires – three or four could be used.

Unfortunately, there are two snags to multifilar winding. Firstly, the thin polyurethane insulation on the copper wire is easily damaged during winding and may break down if we have >100 V between the windings, making it difficult to make a transformer capable of isolating the HT supply. Nevertheless, the seminal 50 W McIntosh [5] amplifier used a multifilar output transformer and a 440 V HT supply! Secondly, the greatly increased capacitance between primary and secondary may resonate with the reduced leakage inductance to produce a lower resonant frequency than a sectioned transformer.

Multifilar winding is best used in small-signal transformers with a very low turns ratio (ideally 1:1), such as the balanced line output transformers used in studios.

DC Magnetisation

DC Magnetisation

If a net DC current is allowed to flow in a transformer, it shifts the AC operating point on the *B*/ *H* curve and causes significant distortion due to saturation on one half cycle. For this reason, output valve anode currents in push–pull amplifiers should be carefully balanced, and half-wave rectification should never be used on mains transformers. A traditional way of checking output valve DC balance in push–pull amplifiers is to measure the voltage between the anodes of the output valves, and adjust for zero volts. Zero voltage between anodes means equal voltage drops, and this implies equal currents with no out-of-balance current, *if the winding resistances are equal*. Checking these resistances before using this method is therefore essential. If the resistances are unequal, it will be a few tens of ohms at most, and it is perfectly permissible to add a permanent series resistor to balance the DC resistances, allowing a correct imbalance voltage to be measured between the anodes.

Because C-core and R-type transformers have a negligible gap and toroids are essentially gapless, they are far more susceptible to core saturation due to DC, particularly as all these transformers have the grain of their core material optimally aligned, allowing the designer to operate much closer to saturation.

Copper Losses

Copper wire has resistance, and in a well-designed transformer the losses due to resistance are equal in primary and secondary, and will therefore be related by:

$$R_s = \frac{R_p}{N^2}$$

where *N* is the primary to secondary turns ratio.

Having equalised copper losses between primary and secondaries, total copper losses may be traded against iron losses in a given transformer design so that two transformers may have different proportions of iron to copper to achieve equal power ratings.

Electrostatic Screens

Capacitance between primary and secondary sections is significant in audio transformers because it is multiplied by the turns ratio of the sections concerned in a manner similar to the Miller effect in the triode valve. The problem can be solved by interposing an earthed electrostatic screen, which should be made of foil, between the affected windings. We now have capacitance to earth, but the effect of this capacitance is minimal. It is most important that the two ends of the foil do not contact electrically, as this would form a shorted turn.

An electrostatic screen between primary and secondary is often fitted to mains transformers for a rather different reason. If the insulation was to break down between primary and secondary, mains voltage would be connected to the secondary, which would be a safety hazard. By interposing an electrostatic screen, the fault current flows directly to earth and blows the mains fuse, thus making the equipment safe. Unfortunately, such screens are usually a layer of wire, fulfilling the safety requirement but seriously degrading RF performance.

A foil electrostatic screen prevents RF interference on the mains from being capacitively coupled to the following circuitry. In audio, the significance of RF interference cannot be overemphasised, and this is sufficient incentive for using a foil electrostatic screen. Foil electrostatic screens are particularly beneficial for low-voltage secondaries because they prevent high-voltage noise from the mains being capacitively coupled directly into sensitive circuitry.

Magnetostriction

Valve amplifiers with output transformers may 'sing' audibly when operated at high power. Occasionally, this is due to a loose lamination, but it is more likely to be due to *magnetostriction*, which is an effect whereby a magnetic material changes its length according to the strength of the magnetic field passing through it. Output transformers support quite strong magnetic fields, so the effect can become noticeable. Since the magnetic field is varying, it causes vibration, but because magnetostriction is not polarity sensitive, in a push–pull amplifier the sound that is heard is pure second harmonic distortion.

Magnetostriction is inversely proportional to μ_r, so a higher-quality transformer is less likely to suffer from this (admittedly minor) problem [6].

Output Transformers, Feedback and Loudspeakers

It is far more convenient to derive feedback from a dedicated feedback winding, or from the end of a tapped winding, because it means that the user can change the matching of the amplifier to the loudspeaker without having to adjust feedback. The leak amplifiers were designed using this scheme, allowing a simple link to determine matching, but it means that the output transformer is not used optimally (see Figure 4.32).

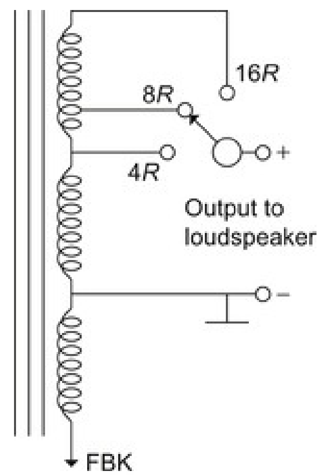


Figure 4.32 Secondary arrangement in leak output transformer.

As an example, when the 4 Ω setting is chosen, only half of the secondary winding is used, resulting in poorer leakage inductance. Worse, the feedback (which would ideally be applied at the output terminals) has to be coupled via the tapped secondary before being applied, and the coupling from one part of the secondary to another cannot be perfect. The optimum way to apply feedback is to derive it from the amplifier output terminals (or even better, the loudspeaker terminals). Ideally, transformer performance should be optimised by using as many of the secondary sections as possible in a carefully controlled way.

Old transformers tend to have a pair of secondary sections which are connected in series for 15 Ω loudspeakers and in parallel for 4 Ω loudspeakers. The sections are not necessarily from the same layers, so the sections have differing resistances and leakage inductances. Connected in series (15 Ω matching) this is not a problem, but when in parallel the mismatched Thévenin sources drive currents into one another, which is not ideal. Better-quality transformers have four secondary sections that cleanly give 1 Ω, 4 Ω, and 16 Ω, but still have the same problem as before if configured for 8 Ω.

Why not 16 Ω? 16 Ω loudspeakers would not need a particularly low-source resistance for optimum damping and would be less upset by loudspeaker cable resistance. In addition, transistor amplifiers could be designed more easily, valve amplifiers could have their secondary sections optimised and the reduced turns ratio would further improve the transformer. However, any manufacturer who introduced a 16 Ω loudspeaker would have that loudspeaker branded as inefficient because a given voltage would produce 3 dB less acoustic power than the 8 Ω design – read published comments about the BBC LS3/5a (12 Ω). So, we are stuck with 8 Ω, and the trend is firmly towards 4 Ω.

Although modern loudspeakers are nominally 8 Ω, two-way designs frequently combine a 4 Ω bass driver with an 8 Ω tweeter, so it is better to treat all loudspeakers as 4 Ω – the slight loss of 8 Ω measured power is insignificant, but the boost in quality is worthwhile.

Transformer Models

Because real transformers are such complex devices, it is usual to devise simplified models that attempt to represent operation at low, mid and high frequencies.

At low frequencies, the transformer may be represented as a perfect transformer in parallel with the primary inductance of the real transformer, driven by the non-zero resistance of the source (see Figure 4.33).

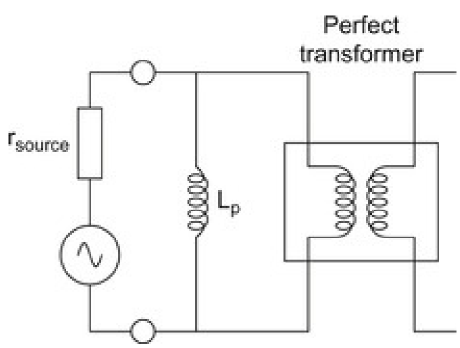


Figure 4.33 Transformer equivalent circuit at low frequency, showing effect of primary inductance.

The combination of source resistance and finite primary inductance creates a high-pass filter whose cut-off frequency is given by:

$$f_{-3\mathrm{dB}} = \frac{r_s}{2\pi L_p}$$

With a given transformer, we will obtain better low-frequency performance if we can reduce the source resistance. An EL34 pentode has r_a=15 kΩ, but the same EL34, used as a triode, has r_a=910 Ω and, used as a cathode follower, r_k<100 Ω.

Unfortunately, the previous model is only appropriate for small signals. In a power amplifier, output-valve operating conditions are invariably carefully matched to its load impedance, and the reduced reactance of L_p at low frequencies diverts signal current from the load into L_p. At high levels, the increased signal current flowing into L_p saturates the core, reducing L_p, further increasing signal current into L_p, leaving little current available for the loudspeaker load, so low frequency distortion increases catastrophically. Power amplifiers thus require the $f_{-3\ \mathrm{dB}}$ frequency L_p to be determined by R_L rather than r_a, and the advantage of an EL34 cathode follower would not be seen at full power.

Once we have decided on the relevant load resistance, we need a high primary inductance, which can be achieved either by increasing primary turns, or by using a core material with a higher μ_r. Although low frequency performance can obviously be improved by increasing μ_r, we can also use increased μ_r to improve high frequency performance. Primary inductance would be maintained by winding fewer turns, thus reducing stray capacitance, which would result in better high frequency performance.

A better core material is preferable because the bandwidth (in octaves) of a transformer with matched source and load resistances is:

$$\mathrm{BW}_{(\mathrm{octaves})} \approx \frac{\log(L_{\mathrm{primary}}/L_{\mathrm{leakage}})}{\log 2}$$

The bandwidth is dependent on the geometry of the transformer, and on μ_r, but not on size or number of turns. All other things being equal, a core with higher μ_r produces a transformer of greater bandwidth, which could be achieved either by a better core material, or by eliminating air gaps (toroid), or both.

Another factor that affects transformer bandwidth is the air gap:

$$\mathrm{BW}_{(\mathrm{octaves})} \propto \frac{1}{\text{air gap}}$$

We might have a good core material but need a large air gap to prevent it saturating. Alternatively, single-ended amplifiers force a magnetising current through their output transformers, necessitating a considerable air gap, reducing bandwidth.

At mid frequencies, we can consider the losses due to the resistance of the windings; it is usual to reflect the secondary circuit into the primary circuit (see Figure 4.34).

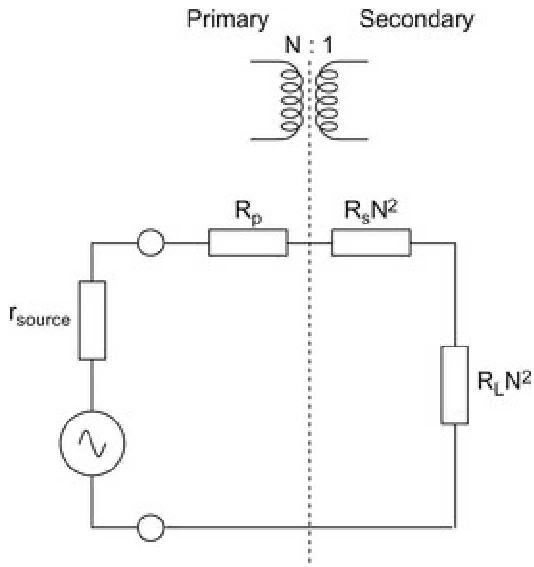


Figure 4.34 Transformer equivalent circuit at mid-frequency, showing effect of winding resistance.

The high frequency model is much more complex (see Figure 4.35).

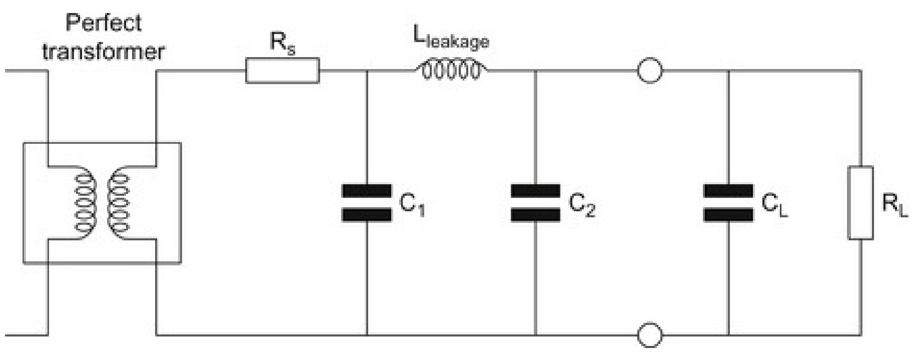


Figure 4.35 Transformer equivalent circuit at high frequency, and its similarity to classical third-order filter.

In this model, the primary circuit has been reflected into the secondary, and the source resistance, primary resistance and secondary resistance have been lumped together. Interwinding capacitance has been lumped and introduced in two positions, and leakage inductance is also included. The resulting circuit is a classic low-pass filter having an ultimate roll-off of 18 dB/octave, and with suitable choice of component values this model accurately simulates a real transformer at high frequencies.

Since the model is a classic filter, we can use the rules that apply to these filters. The most important of these rules is that performance is critically dependent on terminating resistances. For a normal filter, these terminating resistances are the source and load resistance, but a transformer is also sensitive to load capacitance, and since few manufacturers provide data for the effects of load capacitance, it is good practice to test the transformer with the expected source and load impedances and check for high frequency and low frequency resonances.

Input Transformer Loading

When using a moving coil cartridge step-up transformer, it is well worthwhile experimenting to find the optimum load resistance for the *transformer*, before adjusting the load on the cartridge.

Once the source resistance to the transformer is known, the optimum loading for the input transformer can be determined. Beware that better cartridges tend to have higher coil resistances (because thinner wire reduces moving mass), so upgrading a cartridge could require a change of transformer loading. Not only will the frequency response be affected, but a higher cartridge resistance could cause significant loss in the inevitable potential divider formed by the cartridge resistance and the reflected transformer loading resistance.

As an example, the Sowter 8055 was originally designed for a 3 Ω cartridge, and its optimum loading resistance was then a pure 2.7 kΩ resistance. Since the step-up ratio is 1:10 and impedances are transformed by a ratio of n^2, the 3 Ω cartridge saw a reflected resistance of 27 Ω, giving a tolerable loss of 0.9 dB. Replacing the 3 Ω cartridge with a 10 Ω model increases the loss to 2.7 dB, so a further 1.8 dB of sensitivity has been lost.

The significance of the additional 1.8 dB loss is that the noise at the input of the amplifier has remained constant, so the change of cartridge source resistance has degraded the S/N ratio by 1.8 dB. Coincidentally, increasing the g_m of the input valve by 50% gives an improvement of 1.8 dB in S/N ratio, but increasing g_m is always fearfully expensive, so we must avoid unnecessary losses before amplification. If we could increase the loading resistance on the transformer, the reflected resistance seen by the cartridge would rise, and the S/N ratio would improve.

Unfortunately, any transformer has an high frequency resonance caused by leakage inductance and interwinding capacitances, and increasing the loading resistance reduces damping, producing a peak in the frequency response and ringing on square waves. However, a carefully chosen Zobel network across the secondary can significantly tame the ringing. Values for the network are easily found by experiment (see Figure 4.36).

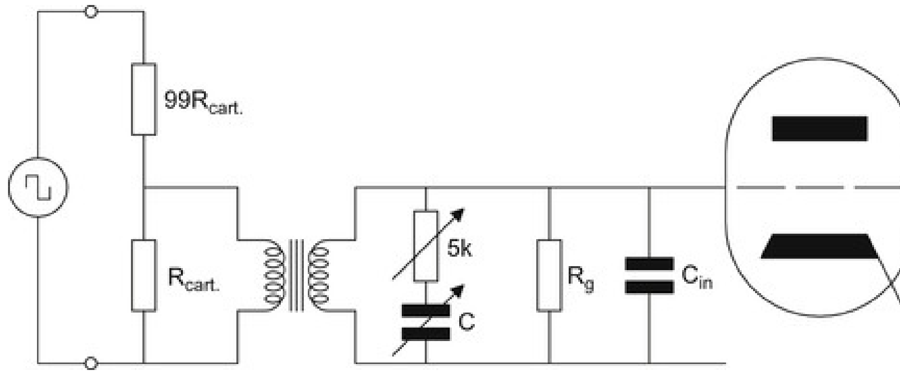


Figure 4.36 Determining Zobel network values for moving-coil transformer.

The purpose of the potential divider at the output of the square wave generator is two-fold:

- We want to drive the transformer from the same resistance as the cartridge plus arm wiring resistance. Typical generators do not have 10 Ω output resistance, so the potential divider is designed to have the correct output resistance.
- The output voltage of typical square wave generators is far too high for the transformer, so we can easily afford to attenuate by a factor of 100.

Calculating the potential divider rigorously is unnecessary because physical coil winding constraints mean that cartridge coil resistances cannot be equal (5% error is typical). Additionally, because the potential divider needs to attenuate by a factor of 100, $r_{out} \approx R_{lower}$, so we just set R_{lower} to be equal to the required resistance, and choose the nearest convenient value to make $R_{upper} \approx 100\ R_{lower}$.

We know that the transformer will drive a valve that has input capacitance, so this should be calculated, or measured, using the method given later in this chapter. Although ×10 oscilloscope probes reduce capacitance at the probe tip, they do not eliminate it, so this capacitance must also be considered during measurement. In this example, the transformer must drive an EC8010 triode whose input capacitance has been measured to be 190 pF. The author's Tektronix P6139A×10 probes present a tip capacitance of 8 pF, so a 180 pF loading capacitor was used (180 pF+8 pF≈190 pF).

The Zobel resistance is unlikely to be larger than that of the main loading resistor, so a 5 kΩ linear potentiometer was used.

The variable capacitance in the Zobel network was salvaged from an irredeemably faulty valve MW radio (these used to be plentiful, but you might now need to go to a radio fair). Air-spaced variable capacitors typically achieve 300–500 pF with the vanes fully closed, but sections can be paralleled, if

necessary.
Once the generator has been set to produce a 1 kHz square wave with an amplitude of ≈100 mV $_{pk-pk}$ at the output of the transformer, the Zobel resistance and capacitance can be simultaneously adjusted to give the best possible leading edge as viewed on the oscilloscope. In general, the resistor adjusts how much curve there is at the leading edge, whereas the capacitor adjusts the amplitude of the ringing superimposed on that curve. Finding the optimum point is surprisingly easy.
Once the optimum resistance and capacitance have been set, they can be carefully disconnected and measured. Your DVM may claim to be able to make both measurements, but a component bridge will almost certainly be better for the capacitance measurement.

Why Should I Use a Transformer?

With all that has been said about the imperfections of magnetic components, it might be thought that they should be avoided at all costs, particularly since they are invariably expensive.
An output transformer can be used to match a low impedance loudspeaker to the high resistance valve output stage, thereby greatly increasing efficiency. If multiple secondary windings are provided, it also allows user-selectable matching to various impedances without having to redesign the amplifier.
An input transformer, such as a moving-coil cartridge step-up transformer, can step up a small signal sufficiently that it can be amplified by the following amplifier with minimum noise due to the amplifier. As a bonus, the primary can be left floating so that any hum induced into the connecting cable from the cartridge to the transformer is rejected by the transformer. (See Chapter 7 for a fuller explanation of these benefits.)
With the possibility of multiple windings, a transformer may allow novel methods of feedback to be applied to a circuit, further improving its performance. This technique has frequently been exploited in power amplifiers [4].
Arguing the case for inter-stage transformers is rather harder. They invariably have to match high source and load impedances, requiring large inductances causing large stray capacitances that reduce bandwidth. Nevertheless, when expense is a secondary consideration, output valves such as the 845 may benefit from being driven by a robust driver valve coupled via a carefully specified and designed driver transformer.
A transformer isolates the DC on the primary from that on the secondary. This is often essential!

General Considerations in Choosing Transformers

These considerations only apply to audio transformers; power supply transformers will be considered in Chapter 5.

Unless you are building a standard circuit, for which a transformer has already been designed, you will almost certainly need a custom-designed transformer. It is therefore essential that you give the designer as many clues as possible so they may make the *right* compromises to suit your circuit.

- Is the transformer an output transformer, or is it a small-signal transformer?
- What is the maximum signal level (mV) that will be applied to the primary at the lowest frequency of interest? Is this level constant with frequency? How much distortion can you tolerate at this frequency/level?
- What is the source resistance?
- What primary to secondary turns ratio is required?
- What shunt resistance and shunt capacitance will load the secondary? Can either of these be varied, if necessary?
- What frequency range do you *need* the transformer to cover? Don't just say 5 Hz to 500 kHz ± 0.1 dB because it can't be done.
- Do you need an electrostatic screen?
- Do you need the transformer screened in a μ-metal can to reduce electromagnetic hum?
- Are there any special requirements that the designer ought to know about?

If the answer to the first question was 'power-output transformer', then these additional questions should be answered, and an annotated circuit diagram of the output stage is ideal.

- Is the output stage Class A or Class AB?
- What is the quiescent DC current? What is the maximum DC current?
- What is the maximum output power, and what is the lowest frequency at which this is required, for a given distortion level?
- Is the output stage push–pull or single-ended?
- Are the output valves triodes or pentodes? Will you need 'ultra-linear' taps? If so, at what ratio?
- What primaries and secondaries do you need? What DC is superimposed on

each?

• What form of physical mounting do you want? Open flanges, shrouds or drop-through?

All these questions may seem rather off-putting, but if you already have a clear idea of what you want, it is much more likely that the finished chapter will meet your expectations.

Uses and Abuses of Audio Transformers

Transformers are among the most reliable of electronic components, often lasting 40 years or more, but they *can* be damaged. Transformers are made of wire that can fail if excessive current is passed and insulation that can break down if it has to withstand too many volts.

The most common way of destroying an output transformer is to drive the amplifier well into overload so that one output valve switches off completely whilst the other is hard on. The leakage inductance of the half of the transformer associated with the switched-off valve tries to maintain its current, and in doing so, it produces a very large primary voltage causing the interwinding insulation to break down:

$$E = -L_{\mathrm{leakage}} \frac{\mathrm{d}i}{\mathrm{d}t}$$

Since d i/d $t \approx \infty$, the EMF developed is far higher than HT voltage, and it is easily capable of punching through the transformer interwinding insulation. If damp has been allowed to get at the transformer, then the (possibly paper) insulation will already be slightly conductive, and the possibility of breakdown is increased. In general, once a transformer's insulation has been damaged, the damage is permanent, so if there's a possibility that a transformer is damp, dry it out thoroughly before stressing it.

The author has still not managed to damage an output transformer, even when driving amplifiers to their full voltage output with only a very small electrostatic loudspeaker connected across the anodes, but the possibility should be considered.

Guitar Amplifiers and Arcs

Since the rate of change of current at overload is high, and output transformers for guitar amplifiers are deliberately poor, implying a large leakage inductance, a sufficiently high voltage can be developed to strike an arc externally, even though the transformer may be designed to survive the experience. The voltage needed to strike an arc depends partly on the cleanliness of the path, so a dirty (conductive) path lowers the voltage, and a carbonised trail from a previous arc certainly reduces the voltage needed.

Although a high voltage is needed to strike an arc, once struck it can be maintained by quite a low voltage. As an example, the xenon lamp used in a small cinema projector must be struck by a capacitive discharge of thousands of volts, yet it may be maintained by only 26 V at 75 A. If an amplifier strikes an arc from the anode, it can only maintain the arc to a place that has a low resistance to ground because a high resistance, such as a grid leak or cathode resistor, would limit the current and extinguish the arc. The heater pins are connected directly to ground via the Low Tension (LT) centre tap, so the most likely place for an external arc to strike is between anode and heater pins, because the only limiting resistance is the HT supply.

If we know that the amplifier will be thrashed, then a possible solution (depending on the amplifier) is to insert a resistor between LT and 0V HT, perhaps a 4.7 kΩ 6 W W/W, in order to extinguish the arc. However, floating the LT supply may now cause hum problems because of poor heater wiring (routing, dressing and connection to chassis).

Other Modes of Destruction

Excessive current through an output valve may cause thermal runaway from grid emission, melting the internal valve structure, thus dragging sufficient current through the output transformer that the primary winding fails. The simple cure is to keep your valve amplifier on display, and if a valve anode glows cherry red, switch it off immediately. (Output stages in valve amplifiers very rarely have fuses partly because the non-linear resistance of the fuse might cause distortion, but mostly because a fuse would not blow sufficiently quickly to protect the output valves.)

Small-signal transformers are usually damaged mechanically. They are fragile and have windings of very fine wire that is easily broken. Treat them with respect.

Magnetic Screening Cans

Transformers screened in μ-metal cans must be handled carefully and not be dropped as the impact work-hardens the μ-metal screen, greatly reducing its effectiveness (BBC 1:1 transformers designed to operate at −45 dBu had dire warnings on their screening cans about mechanical shock) (see Figure 4.37, photo of BBC toroid can with warning).

Figure 4.37 The μ-metal screening can of this transformer demands careful handling.

Magnetic Core Deterioration

Magnetic core materials can deteriorate with time (this was a stock fault for the mains transformer of a particular picture monitor), and the author has recently seen a number of chokes and transformers whose aberrant behaviour can only be explained by core material that has deteriorated. You might want to bear this in mind when choosing between an NOS part and a slightly more expensive new one.

Thermionic Valves

History

The thermionic valve was not so much invented as discovered and was a consequence of Thomas Edison's research into extending the longevity of incandescent filament light bulbs. It had been observed that as the light bulb neared the end of its useful life, the glass became discoloured and darkened. (This effect is not often clearly visible on domestic light bulbs, but it can be seen on non-quartz halogen torch bulbs and stage lamps.) The cause of the darkening was evaporation of the tungsten *filament* followed by deposition on the inner surface of the glass. In an attempt to counter tungsten evaporation, a *plate* was introduced into the (evacuated) *envelope*, and it was then noticed that if the plate was positively charged with respect to the filament, a current flowed across the vacuum. (Light bulbs need a vacuum because the incandescent tungsten filament would otherwise oxidise so rapidly that it would burn.) Glass darkening can clearly be seen in the author's Ediswan 'R' type valve (introduced in 1918, this valve had a 6-V 2.8-W heater, yet g_m was only 0.225 mA/V) (see Figure 4.38).

Figure 4.38 R-type valve; note darkening of envelope. (The etched legend declares this valve to be 'Type approved by Postmaster General BBC'.)

In 1904, John Ambrose Fleming [7] went rather further, and invented a new device that he termed an electrical *valve* (borrowing the term from the control of gases) because it only allowed current to flow in one direction. His device used two carbon filaments, one of which was heated 'to bright incandescence of greater intrinsic brilliancy than if used as an incandescent lamp', although this was qualified later in the patent as only being ≈2,000 K (the tungsten light bulbs now being phased out typically operate at ≈2,900 K). The other carbon filament, or *electrode*, was cold, and once a source of AC was connected between these two electrodes, Fleming found that current could only flow in one direction. The new device was termed the *thermionic* diode because of the *thermal* energy required to produce the *ion* flow. Strictly, only soft vacuum (or low pressure gas-filled) valves rely on the flow of gas ions – hard vacuum valves rely on the flow of electrons.

Although the diode had great curiosity value, and Fleming had suggested that it could be used for the detection of Hertzian (radio) waves, it was of limited practical use for two reasons. Firstly, carbon emits electrons reluctantly and, secondly, electron emission is strongly dependent on temperature, which had to be kept quite low to avoid premature failure of the carbon filament. These two factors meant that the carbon filament had only ≈0.003% of the emission of a tungsten filament at 2,900 K, necessitating a sensitive mirror galvanometer to

detect emission.
Lee de Forest's audion patent [8] of 1908 interposed a platinum wire in the shape of a grid iron between the heated filament and the plate, and showed experimentally that amplification could be achieved. Although the patent reveals that he did not understand how it worked at the time, the new device was useful and quickly led to the commercial birth of radio. The amplifying characteristics of the soft vacuum Audions were very variable, but the far more predictable hard vacuum triode soon evolved.

Emission

All metals have free electrons within their crystal structure, so some of the electrons must be at the surface of the metal, but because they are negatively charged, they are bound there by the electrostatic attraction between them and the adjacent positively charged nuclei. However, the atoms and electrons are constantly vibrating due to thermal energy, and if the metal is heated sufficiently, some electrons may gain sufficient kinetic energy to overcome the attractive forces of the atoms and escape.
The heated metal in the valve is the cathode, and when this is heated to a temperature determined by the work function of the metal, an electron cloud or space charge forms above the surface of the cathode. Because like charges repel, the cloud eventually accumulates sufficient charge to prevent other electrons escaping from the surface, and an equilibrium is reached.
If we connect the plate, or anode, to the positive terminal of a battery, electrons are attracted from the cloud and are accelerated through the vacuum to be captured by the anode. Because the electron cloud has been depleted, it no longer repels electrons as strongly, so more electrons escape the surface of the cathode to replenish the electron cloud.
Current cannot flow in the opposite direction because only the heated cathode can emit electrons, and only the positive anode can attract electrons.

Electron Velocity

We mentioned that electrons were accelerated towards the anode, and this is quite literally true. At the exact instant that an electron leaves the cloud, it has theoretically zero velocity, but it is constantly accelerated by the electric field of the anode, and acquires energy proportional to the accelerating voltage:

$$E = q_e V = \frac{1}{2} m_e v^2$$

where

E=energy

q_e=electronic charge ≈1.602×10^{-19} C

V=accelerating voltage

m_e=mass of electron ≈9.11×10^{-31} kg

v=electron velocity.

Rearranging, and solving for electron velocity:

$$\text{electron velocity} = \sqrt{2V \cdot \frac{q_e}{m_e}}$$

The ratio q_e/m_e is known as the electron charge/mass ratio and has an approximate value of 1.7588×10^{11} C/kg. If we apply 100 V between the anode and the cathode, the electrons will collide with the anode with a velocity of ≈6×10^{6} m/s or 13 million miles per hour.
Using the previous equation, it would appear that 512 kV (a common national distribution voltage) would be sufficient to accelerate the electrons to be faster than light speed – which is an impossibility. The flaw is that the simple equation assumes that the mass of the electron is constant, but at relativistic (approaching the speed of light) velocities, the mass of the electron increases in accordance with the Lorentz–Einstein equation [9], thus requiring an infinite voltage to accelerate it to light speed:

$$m = \frac{m_0}{\sqrt{(1-(v^2/c^2))}}$$

To account for this, the elegant equation given by Alley and Atwood [10] may be used:

$$\text{velocity} = c \cdot \sqrt{1 - \frac{1}{(1 + (q_e/m_e)\cdot(V/c^2))^2}}$$

where

c=velocity of light in a vacuum ≈2.998×10^{8} m/s

q_e/m_e=electron charge-to-mass ratio ≈1.759×10^{11} C/kg

V=anode voltage.

As an example of relativity at home, a good quality television using a cathode ray display tube needed a final anode voltage ≈25 kV, implying an electron collision velocity of 202 million miles per hour at the tube face, but the simple equation predicts a velocity 3.5% high.

Note that the distance between the anode and the cathode does not feature in either equation because an infinite distance would also allow an infinite time for acceleration, and even if the rate of acceleration was very low, the collision velocity would still be reached.

Many effects within valves can be understood by having an appreciation of the collision velocity of the electrons as they hit the anode.

Transit Time

It is occasionally suggested that the transit time between cathode and anode or even between g_2 and anode is significant in audio. Transit time can be calculated using [11]:

$$t_{\text{transit}} = \frac{2d}{\sqrt{2 \cdot (q_e/m_e) \cdot \sqrt{V}}}$$

where

d=distance between electrodes

q_e/m_e=electron charge-to-mass ratio ≈1.759×10^{11} C/kg

V=voltage between electrodes.

The KT66 has a particularly large cathode to anode spacing (estimated at 7 mm) and might be operated (sub-optimally) at 350 V, giving a transit time of 1.26 ns. Even for video, 1 ns is a very short time, and the example was specifically chosen to be slow, so most valves will be much faster, making transit time entirely irrelevant to audio.

Individual Elements of the Valve Structure

The Cathode

Early valves betrayed their light bulb origins and were *directly heated* using a tungsten filament that was also the cathode. Tungsten was used in incandescent light bulbs because it has the highest melting point of all electrical conductors (3,695 K or 3,422 °C) and could therefore withstand the ≈3,000 K temperature necessary to generate light that wasn't obviously yellow. Although producing bright light was not actually necessary for early valves, it was quickly found that reduced cathode temperature caused electron emission to fall drastically, so early valves became known as *bright emitters*. The emitted current per unit area is:

$$I \propto T^2 \mathrm{e}^{-(q_e \varphi / kT)}$$

where

T=absolute temperature of the cathode =°C+273.16

q_e=electronic charge ≈1.602×10^{-19} C

φ=work function of the cathode surface (≈4.55 for tungsten)

k=Boltzmann's constant ≈1.381×10^{-23} J/K

e=base of natural logarithms ≈2.718.

(See Appendix for full theoretical Richardson/Dushman equation.)

The emission efficiency of the cathode is important because not only does the filament dissipation increase the power requirement of the equipment, but the heat must also be lost without damaging any other components. We therefore want to maximise electron emission for a given filament power, so the history of the cathode is concerned with the developing chemistry of the cathode emissive surface.

The first improvement was to use a *thoriated* tungsten cathode which not only had improved emission, but also could operate at between 1,950 K and 2,000 K rather than 3,000 K. This reduced temperature was significant because valves primarily lose heat by radiation, and by Stefan's law:

$$E = \sigma T^4$$

where

E=power per unit area

σ=Stefan's constant ≈5.67×10^{-8} W/K ^{4}/m ^{2}

T=absolute temperature=°C+273.16.

Thus, 1,975 K only requires one-fifth of the heater power to overcome the losses due to radiation compared to 3,000 K, and these valves are sometimes known as *dull emitters*. Although the emission had only been doubled, the reduction in heater power by a factor of five meant that the total improvement in emission efficiency was a factor of ten.

The real improvement came with the oxide-coated cathode, which operated at only ≈1,100 K, and was 100 times as efficient as the pure tungsten cathode. As an example, even in 1929, the P215 battery valve had a directly heated barium azide coated cathode requiring only 2 V at 150 mA (see Figure 4.39).

Figure 4.39 P215 directly heated oxide-coated cathode valve.

Valve manufacturers often had proprietary cathode coating formulations, so the JT Baker Company manufactured the splendidly named ‘Radio Mixture 3’, which was composed of 57.3% barium carbonate, 42.2% strontium carbonate and 0.5% calcium carbonate [12]. Sadly, this concoction almost certainly bears no relation to the apocryphal ‘Love Potion 9’.

Unless the cathode is pure tungsten, the active emissive surface is only one molecule thick, and consequently fragile.

The vacuum in a valve is never perfect, and there will always be stray gas molecules between the anode and the cathode. A cold cathode prevents anode current, so zero voltage is dropped across the anode load resistor, causing V_a to rise to the full HT voltage. As the cathode warms from cold, a few electrons are attracted towards the anode, but some collide with stray gas molecules to produce positive ions which are repelled towards the cathode. If the only force on ions between the cathode and the anode was repulsion from the anode, space charge repulsion would prevent them from reaching the cathode. However, entropy motion ensures that some ions already have some momentum towards the cathode, allowing them to overcome the space charge, and because the ion is a molecule having a nucleus composed of (heavy) protons and neutrons, it has

considerable momentum when it strikes the cathode surface.
If this process of cathode bombardment occurs sufficiently often, the cathode emissive coating can be significantly impaired, and oxide-coated cathodes are even more vulnerable than thoriated tungsten cathodes, so if $V_{a(pk)}$>2 kV is required, thoriated tungsten cathodes are the norm. Because pure tungsten cathodes do not rely on a monomolecular layer for emission, they are almost immune to ion bombardment.
Given that the vacuum cannot be perfect, we must minimise bombardment by establishing the protective space charge above the cathode emissive surface before anode voltage is applied.
Another problem with oxide cathodes is cathode *poisoning*. If the cathode is kept at full operating temperature, but little or no current is drawn, a high resistance layer of barium orthosilicate forms at the interface between the barium oxide emissive surface and the nickel cathode structure. The interface resistance eventually reduces emission, but more significantly it increases the noise generated by the valve because the valve’s normal shot noise current develops a noise voltage across this resistance, which is in series with the input signal.
Poisoned cathodes can occasionally be gradually recovered by operating the valve at a high anode current. Another method, often used on the cathode ray display tube used in traditional televisions, is known as *rejuvenation*[13] and this works by temporarily increasing heater volts to heat the cathode to a higher temperature, and simultaneously drawing a large anode current. It should be realised that rejuvenation carries a risk of evaporating some of the cathode emissive surface, and contaminating the (nearby) control grid. Inadequate cathode activation is sometimes cited as a cause of noise, perhaps explaining the fact that rejuvenation occasionally improves a valve’s noise.
The final generation of colour television cameras using Plumbicon tubes used oxide cathodes and had a ‘standby’ mode whereby their heaters ran at half power (63% heater voltage) in order to extend the life of the tubes and reduce the wait for decent pictures when fully switched on. To be certain of avoiding cathode stripping *and* cathode poisoning, the author has previously adopted the same strategy in his RIAA stages, especially since one of them immediately sounded fine if the heaters had been left in standby, but took 2 h to recover from cold and meanwhile sounded terrible. In hindsight, it seems likely that leaving the RIAA stage heaters in standby kept the entire electronics sufficiently warm to prevent condensation and surface leakage currents causing increased noise and distortion. (Some recording engineers argue that valve condenser microphones sound better not because of the superior electronics, but simply because the heat

rising from the valve keeps the high impedance capsule and associated wiring dry despite the humid breath directed at it.)
‘Standby’ dissipation of 8 W in an RIAA stage heater equates to 25 W from the mains once regulator and transformer losses have been taken into account. Over a week, that amounts to 1.3 kWh – equivalent to running the hugely profligate ‘Crystal Palace’ amplifier for 2 h. With the current price of electricity, this sort of extravagance can no longer be tolerated, especially since surface leakage paths can be greatly reduced by PTFE insulators (PTFE repels water, causing condensation to form in isolated globules rather than wetting, joining, and making a conductive path). Thus, the author no longer recommends ‘standby’ mode on valve heaters, and the knock-on from this is that valve HT rectifiers become a necessity to avoid cold cathode bombardment.

Thoriated Tungsten Filament Fragility

Thoriated tungsten cathodes operate only slightly below the melting temperature of thorium (2,023 K), and to reduce the evaporation of thorium from the surface, the tungsten filament is partly converted to tungsten carbide [14]. Unfortunately, although hard tungsten carbide is brittle, so the degree of carbonisation is a delicate compromise between reducing thorium evaporation and fragility. Because thoriated tungsten filaments are very brittle, so valves such as the 211, 813 and 845 should be handled with extreme care and not be subjected to mechanical shock.
Unfortunately, thermal shock also kills thoriated tungsten filament valves. A 1994 study of transmitter valve longevity [15] found that each off/on cycle reduced filament life by 0.2% from its maximum life of 30,000 h. This doesn’t sound too bad, but it implies that 500 off/on cycles will destroy the filament, so if you switched the valve off and on every day, you could expect it to expire in less than 17 months. Understandably, the broadcasters took a dim view of this, and looked to see how life might be extended.
There are two reasons why the off/on cycle kills thoriated tungsten filaments:

- As the filament temperature passes through ≈900 K, the Miller–Larson effect causes the grains of the metal to reorient themselves, so that the wire becomes thinner and longer. Worse, if a given section of the filament is slightly thinner, the increased current density causes increased localised heating, which exacerbates the Miller–Larson effect and causes further necking of the filament. Eventually, this necking leads to such deep cracks that the remaining conductive material has sufficiently high current density and local heating to vaporise it, thus destroying the filament.

- The resistance of a cold filament is far less than that of a hot one, and assuming an operating temperature of 1,975 K, but an ambient temperature of 293 K (20 °C), the initial cold current is 8.6 times higher than the operating current. The inrush current through the filament interacts with the Earth's magnetic field to produce a small kick. Combined with the Miller–Larson effect, this gradually deepens the surface cracks in the brittle filament. The damage done to the filament is proportional to the cube of inrush current, so a 'soft start' can be worthwhile.

If you had bought a quartet of NOS 845s at considerable expense, you would have a vested interest in avoiding the Miller–Larson effect, might want to permanently operate the filaments in standby mode at 80% of full voltage, and only apply full voltage at full switch-on, but note that standby still expends the emissive life at a rate of 1% compared to full filament voltage (but no anode current).

Direct Versus Indirectly Heated Cathodes

Early valves used lead–acid cells to power their directly heated cathodes, so heater voltages insertion were multiples of 2 V. Having to take a heavy lead–acid battery to the local radio shop for periodic recharging was a nuisance, so later valves used the household AC mains supply. Unfortunately, AC heating caused audible hum owing to the three mechanisms described in the following subsections, ranked in order of significance.

The Thermal Problem

The filament has to be made of sufficiently fine wire to give a high resistance that can be heated by a low current and does not require excessively thick wires from the transformer. Because the wire is so fine, the thermal mass of the filament is low, and the temperature of the filament is partly able to track the applied power. This effect can be observed by noting that the thick filament of a car headlight dims slowly when switched off, yet the fine filament in a domestic AC mains light bulb of the same power dims quickly. Because emission is modulated by heating *power* (V^2 or I^2), the mechanism produces hum at twice the AC mains frequency.

The Electrostatic Problem

The voltage drop across the filament affects V_{gk} because if we consider any one point to be connected to HT 0 V, other points are at different (and changing) potentials. If we connect the centre of the filament to HT 0 V, then the ends will be at equal and opposite voltages, so although one end of the filament emits more electrons, the other emits fewer. However:

$$I_a \propto V_{gk}^{-(3/2)}$$

The significance of the three halves power law is that the excess of electrons drawn from the more positive end of the filament is *not* exactly nulled by the deficiency from the opposite end. The imbalance is most apparent when the (AC) filament supply is at a positive or negative peak, so this mechanism also produces hum at twice AC mains frequency.

The thermal and electrostatic problems arise because the filament supply is a sine wave. A square wave filament supply would eliminate these problems, although preventing breakthrough of the higher harmonics into the audio circuitry via C_{hg} would be a major headache.

The Electromagnetic Problem

The magnetic field created by the filament heating current curves the flight of the electrons so that some miss the anode. When AC is applied to the filament, the direction of curvature alternates with the polarity of the filament current, so this mechanism produces hum at mains frequency.

The Indirectly Heated Cathode Solution

Although various alternatives were tried, the best solution to the three previous problems was the *indirectly heated* cathode [16], whereby the emissive material was applied to a coated metal sleeve surrounding the filament, which was then termed the *heater*. Although not explicitly stated in the patent, the key idea was that if the sleeve cathode had sufficient thermal mass, it would be unable to track the changing temperature of an arbitrarily thin heater wire. Because the metal sleeve did not pass heater current, its entire surface was at the same potential, so it was named a *unipotential* cathode, and this solved the electrostatic problem. Making the sleeve from a magnetic material such as nickel tends to screen the filament's magnetic field, reducing the electromagnetic problem.

Since the sole purpose of the indirectly heated cathode is to reduce hum, the AC powering the heater must be electrically insulated from the signals on the cathode. Unfortunately, good electrical insulators also tend to be good thermal insulators, and the thermal resistance of the aluminium oxide electrical insulator means that the heater must be at 1,650 K to raise the cathode to 1,100 K, so indirectly heated cathodes need more heater power than directly heated cathodes. They also take longer to reach operating temperature, but for small-signal valves, the reduction in hum is invaluable, so the slow warm-up and loss of efficiency can be tolerated.

The cathode emissive surface is sprayed onto the outside surface of the nickel sleeve, and the heater reverts to pure tungsten. However, the cathode sleeve looks like an anode to the heater filament, and if a current was allowed to flow from the heater to the cathode, then this would add to the intended cathode to anode current, and hum would result. It has previously been supposed that such a current was directly due to electron emission from the hot tungsten heater, and this idea was bolstered by the fact that superimposing a small (+10 V seems sufficient) DC voltage on the heater supply prevents the hum current. However, it seems more likely that electrolysis deposits a permanent conductive path through the porous aluminium oxide insulator and that biassing the heater positive prevents formation of this path. Either way, RCA [17] frequently recommended +40 V as the optimum voltage between the heater and the cathode, and until more evidence arrives it seems churlish to ignore their practical recommendation.

Despite all these efforts to eliminate hum, the heater filament could still induce hum into the signal circuitry either by leakage currents, or because of the imperfect magnetic shielding of the nickel cathode sleeve. In a further effort to reduce hum, the heater filament of the EF86 was wound as a helix in order to cancel the magnetic field caused by the heater current.

The only way to eliminate heater-induced hum is to use a DC heater supply with no AC content whatsoever, and this implies a *stabilised* supply, which has other benefits. Because cathode emission is so strongly temperature dependent, it is essential that the heater voltage is correct, and Mullard quoted a maximum permissible heater voltage variation of ±5%, which is exceeded by the current UK legal limit for mains voltage variation (+10%, −6%). A stabilised heater supply stabilises the characteristics of the valve, and the elimination of thermal cycling of the cathode surface reduces LF noise.

As an aside, when the author installed an Automatic Voltage Regulator (AVR) to the test bench supplying his AVO VCM163 valve tester to combat mains voltage fluctuations, the AVR worked hardest between 4 pm and 11 pm. Is it a coincidence that the sound of the Hi-Fi seems to improve after midnight?

All the final generation of small-signal valves (except battery and electrometer valves) use indirectly heated cathodes, and directly heated pure tungsten cathodes are now only used for high power transmitters.

Heater/Cathode Insulation

An indirectly heated cathode consists of a heater filament insulated by aluminium oxide folded and slid into a tightly enclosing tubular cathode (see Figure 4.40).

Figure 4.40 6K7 heater filament removed from cathode. The aluminium oxide insulation has been removed at the apex of the hoop to expose the much thinner tungsten filament.

No insulator is perfect, and they all deteriorate rapidly as temperature rises, which is unfortunate, since the heater/cathode insulator is red-hot. Typically, the resistivity of aluminium oxide at the cathode's operating temperature is less than one-millionth of its room temperature value. All heater/cathode insulation must therefore be electrically leaky, allowing leakage currents to flow between heater and cathode. Even worse, if the insulation is contaminated, this imperfection produces $1/f$ noise. Irritatingly, one of the worst offenders for poor heater/cathode insulation is the otherwise excellent 12B4-A, so this valve must be screened to exclude those samples with poor (hot) insulation if noise is critical.

Heater power could be reduced, and the valve could be made more efficient, by reducing the thickness of the heater/cathode insulation, and this is exactly what was done in the transition from the International Octal based generation to the later B9A generation, but this compromises electrical heater/cathode insulation.

Increasing the voltage across the heater/cathode insulation increases leakage currents. Although V_{kh}(max) is specified on datasheets as being anywhere from 90 V to 150 V (except for some ruggedised and 'P' series TV valves), this is a very 'soft' limit, since it is usually given at an arbitrary leakage current. Nevertheless, a sufficiently high voltage *will* punch through the insulation to rupture the heater. Heater failure due to heater/cathode insulation breakdown is uncommon, but it is most likely in cathode followers with high-signal voltages or output stages with distributed loads (such as the McIntosh design).

Cathode Temperature Considerations

Because of the Richardson/Dushmann equation, electron emission, and therefore anode characteristics, are critically dependent on cathode temperature. Provided that anode dissipation is sufficiently low that it does not further heat the cathode, cathode temperature is related to heater power (P) by:

$$T \propto \sqrt[4.5]{P}$$

Szepesi [18] also found that the oxide-coated cathode of the Tungsram HL4G produced minimum noise when operated at ≈1,200 K and that a 60 K drop in cathode temperature caused by a 25% drop in heater voltage doubled the noise power.

Operating oxide-coated cathodes at higher heater voltages dramatically shortens life because it increases evaporation of the emissive material, so $V_h \ngtr 105\%$. Thus, long life, low noise and stable anode characteristics demand heater supplies stabilised at the correct voltage.

Heaters and their Supplies

It is usual to supply power to parallel heaters in from a constant voltage source (typically 6.3 V), or series heaters from a constant current source (typically 300 mA). If either type of supply drifts from its nominal value, undesirable changes in anode characteristics occur. Although 6.3 V regulators are easily made, linear regulators become increasingly inefficient as load current rises, and unless carefully designed and constructed, switched mode regulators can be electrically noisy. By comparison, a 300 mA constant current supply feeding a pure series heater chain is easily and efficiently implemented.

Valve manufacturers often specify series or parallel heaters, but is there actually any fundamental difference between the filaments of the two types, and could we use 6.3 V heaters (of equal current requirements) in a constant current chain?

Some 6.3 V valves were tested to see if there was any significant difference between the behaviour of their heaters. The valves were deliberately chosen to be as different as possible to magnify any difference between heaters. The 12AT7 was selected to be one whose heater flashed white at switch-on (Table 4.3).

Table 4.3 Percentage of Normalised Heater Current Against Heater Voltage for Parallel (6.3 V) Heaters

Heater voltage (V)	CV4024 12AT7 (0.30 A)	Mullard ECC83 (0.29 A)	Raytheon 5842 (0.30 A)	GE 6BX7 (1.45 A)	Mullard EL84 (0.79 A)	Mullard EL34 (1.475 A)	Mean	1 σ
6.30	100	100	100	100	100	100	100	0

6.00	97	97	100	97	96	97	97.3	1.37
5.50	92	93	93	93	92	93	92.7	0.52
5.00	87	90	90	88	87	87	88.2	1.47
4.50	83	83	83	83	82	85	83.2	0.98
4.00	77	78	80	78	77	78	78.0	1.10
3.50	70	66	73	72	72	73	71.0	2.68
3.00	67	60	67	66	67	67	65.7	2.80
2.50	58	62	60	61	60	60	60.2	1.33
2.00	52	55	53	54	51	53	53.0	1.41
1.50	42	45	47	45	43	44	44.3	1.75
1.00	30	34	37	37	32	34	34.0	2.76
0.50	20	21	20	26	19	20	21.0	2.53

1 σ=Standard deviation.

Within the limits of experimental error (which worsened significantly towards 1 V), the heater currents are in very close agreement, suggesting that parallel heater valves have essentially similar filaments. This is broadly to be expected, since tungsten is the only practical filament material.

A PL508 was then tested set to its correct current (300 mA), the voltage at that current was measured, and a series of measurements was taken, which were normalised to 6.3 V and compared with the mean currents from Table 4.3 (Table 4.4).

Table 4.4 Percentage Comparison of Current Heater Valve (PL508) with Mean Current of Voltage Heater Valves

Voltage	6.3	6.0	5.5	5.0	4.5	4.0	3.5	3.0	2.5	2.0	1.5	1.0	0.5
Mean	100	97	93	88	83	78	71	66	60	53	44	34	21
PL508	100	97	93	88	83	78	71	66	60	54	45	35	23

A somewhat improved measurement technique was available when the PL508 was tested; nevertheless, the correlation between this and the parallel heaters is remarkable. There appears to be no significant difference between the filaments in valves specified for series or parallel heaters, and provided that individual heaters consume their correct power, there seems to be no reason why we should not mix the two types at will. This technique will be used in the EC8010 RIAA stage described in Chapter 7.

As a further test, the heater current of an EL34 was investigated at 0.5% intervals within a ±5% nominal voltage range, and the results were plotted as a graph. An extremely close fit to a straight line was observed, indicating that the heater behaves as a constant resistance over this very limited range. Since $P = I^2R$ and $P = V^2/R$, two conclusions emerge.

Firstly, parallel chains should be constant voltage (Thévenin) regulated, and series chains should be constant current (Norton) regulated.

Secondly, we should not mix topologies – series/parallel heater chains cause errors – because each heater no longer sees a perfect Thévenin or Norton source.

The author investigated his stock and found that all 24 of his 6SN7 and 12SN7 double triodes have their heaters internally wired in parallel (except for a single RCA 12SN7), whereas his 14 (12.6 V) 14N7 uses two 6.3-V heaters wired in series. The significance of this observation is that 14N7 should be more stable with constant current heating, whereas 6SN7 and 12SN7 should be more stable with constant voltage.

As a more insidious example, a double triode initially tested on a valve tester with 6.3 V parallel heaters, and found to have perfectly matched anode characteristics between sections, would be mismatched by configuring the heaters in series unless the heaters were also perfectly matched. Matching should closely replicate the proposed conditions of use.

Current Hogging and Heater Power

Having established that valve heaters are essentially the same, is the series or parallel chain best in terms of heater power regulation? Fortunately, although parallel connection with a Thévenin source is rather better in terms of power regulation and thermal runaway, the filament/cathode loses most of its heat by radiation, which is proportional to the fourth power of absolute temperature, so this is a very effective regulating element.

It has been suggested that valves intended by the manufacturer for constant voltage heating could exhibit voltage hogging (analogous to current hogging in parallel power transistors) when connected in a series string and powered by constant current. The hypothesis is that if one valve starts with a larger voltage than another, it must be absorbing more power ($P = IV$, and I has been held constant), and if it is absorbing more power, its filament temperature must rise, causing its resistance to rise, making it absorb even more power and causing it to hog heater voltage. The author tested an EC8010 (with 315 Ω parallel resistor to draw 20 mA and correct total current to 300 mA) in series with a 6J5GT to see if there was any evidence of current hogging occuring (see Figure 4.41).

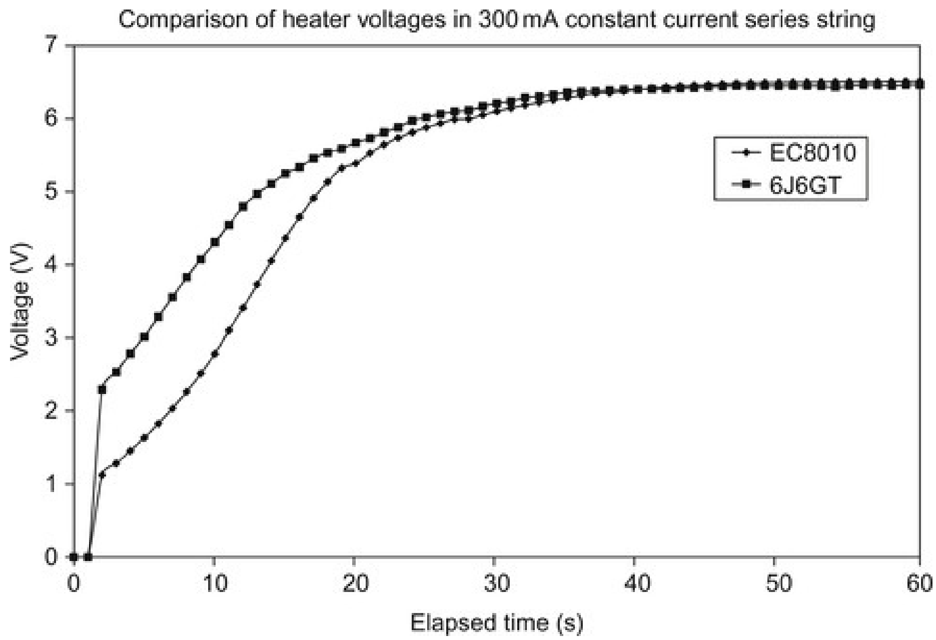


Figure 4.41 Heater voltages of two dissimilar valves fed in series from a constant current source.

As might be expected from two entirely different valves, the voltage across each valve is completely different at switch-on, but they gently converge and after 45 s settle to very nearly equal values, with no evidence of current hogging.

However, in another series of tests, the author monitored heater power for a pair of valves of the same type but selected for a large difference in their heater powers. The valves were first connected as a series string and driven constant current at the manufacturer's rated current and their heater voltages logged at 1 s intervals for 120 s from switch-on, and then the pairs were driven individually constant voltage at the rated 6.3 V (measured at the pins) and both voltage and current logged at 1 s intervals for 120 s. For each valve, the disparity in power dissipation between the pair was calculated for both constant current heating and constant voltage heating (Table 4.5).

Table 4.5 Comparison of Constant Voltage and Constant Current Heating

Valve type	Constant voltage (%)	Constant current (%)	Ratio
KT88 (1.6 A)	3.8	11	2.9
6J5GT (0.3 A)	2.6	4.8	1.9

The tabulated results tell us two things. Firstly, constant current heating exacerbates differences between valves, almost doubling the 6J5GT disparity and almost trebling the KT88 disparity. Secondly, it seems that the problem becomes worse with higher heater powers – it would not be a good idea to use series KT88 heaters.

Constant current 300 mA series heater chains have a number of advantages compared to conventional 6.3 V $_{DC}$ regulated supplies for parallel heaters:

- The assumed linear regulator is more efficient.
- They are inherently proof against accidental short or open circuits.
- The thermal shock to the valve heaters at cold switch-on is eliminated.
- Individual heater resistances can be used as part of a staged Radio Frequency Interference (RFI) filter.
- Heater wiring resistance becomes irrelevant (a complex pre-amplifier using Octal valves, perhaps consuming 6.3 V at >5 A, would require thick heater wiring).

The disadvantages of series heater chains are:

- Any failure is catastrophic and affects all valves in the chain. Having said that, the author has had a total of two heater failures in 30 years (one self-inflicted by flouting $V_{hk(max)}$). Sadly, the second failure *was* in a series heater chain, and the consequential damage was horrendous.
- Constant current heating exacerbates differences between valves – which might be a problem in a differential pair.

Heater Voltage and Current

Typical indirectly heated valves require approximately 1 min to reach 99% heater temperature from cold, or 40 s when preheated at 80% current (63% voltage). When driven from a constant current source, heater terminal voltage is a very sensitive measure of heater (and by implication) cathode temperature. An International Servicemaster 14N7 (I_h=300 mA) was tested by a 4½ digit DVM set to log heater voltage at 5 s intervals. The results were normalised to 100% of the final heater voltage and are presented in the accompanying table (Table 4.6).

Table 4.6 Heater Voltage Against Time When Driven by a Constant Current

Time (s)	V_h (%)
0	22.62
5	36.28
10	46.60
15	58.93
20	78.75
25	89.29
30	93.14
35	95.24
40	96.87
45	98.06

50	98.80
55	99.19
60	99.38
90	99.54
120	99.70
150	99.80
180	99.87
210	99.93
240	99.98
270	100

As can be seen from Table 4.6, although the valve operates correctly within 60 s, a considerably longer time is required before the heater/cathode reaches thermal equillibrium. Since emission, and therefore valve operation, is temperature dependent, we cannot expect stable operation until 5 min after switch-on.

The most marked changes occur during the first minute of operation, so the heater voltage over the first 50 s of a selection of different types was tested using an HP54600B oscilloscope set to 20% V_h/div vertically and 5 s/div horizontally (see Figure 4.42).

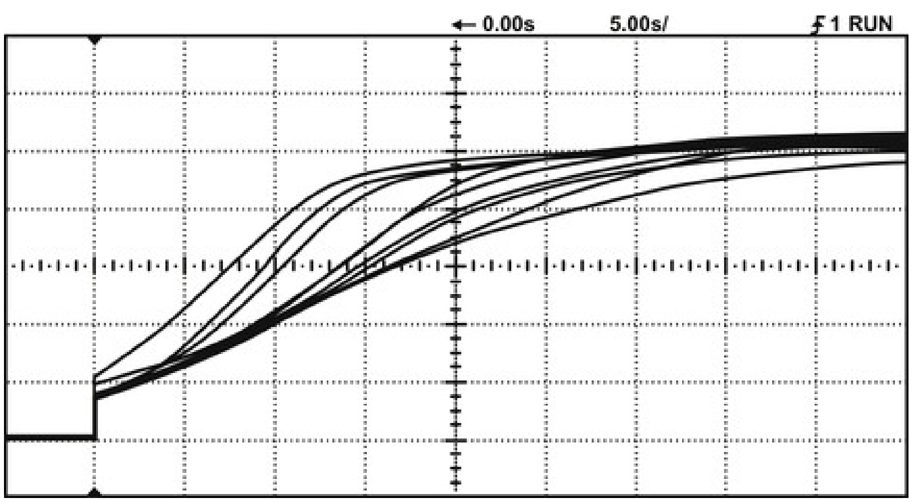


Figure 4.42 Heater voltage against time for various valves fed from a constant current source.

As can be seen, all the valves warm at different rates, hence the manufacturer's caveats about series heater chains.

Without changing oscilloscope settings, a Brimar ECC88 was preheated with 80% heater current, and the effect of this was compared with heating the same valve from cold, and this showed that preheating reduced warm-up time from 40 s to <10 s (see Figure 4.43).

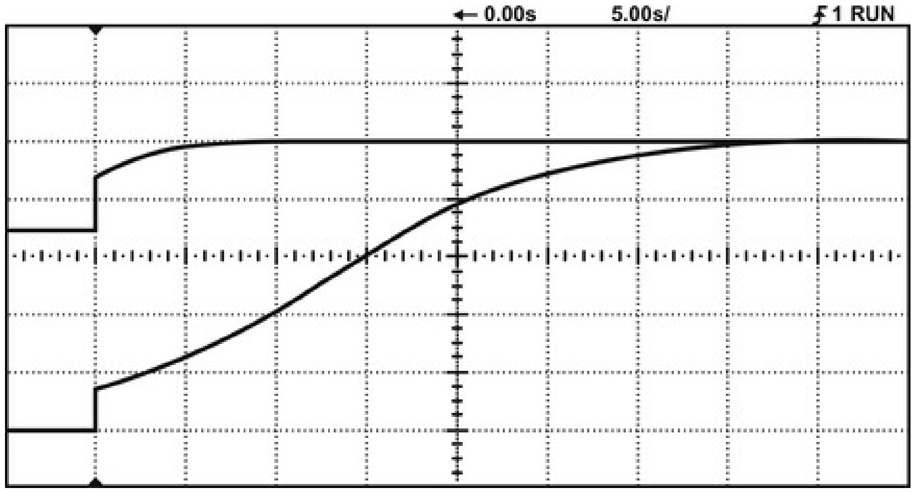


Figure 4.43 Heater voltage against time from constant current source. Upper trace: Preheated. Lower trace: From cold.

The Control Grid

The control grid is wound from stiff, fine wire (often tungsten) as a helix around the cathode, and it is most effective close to the cathode surface, where the velocity of the electrons is less, than near the anode, by which time the electrons have acquired considerable momentum and are not so easily repelled. Therefore, even in a valve having a succession of grids, the control grid is *always* the grid nearest to the cathode. The pitch of the grid winding and its positioning relative to the anode and cathode influence g_m and μ (see Figure 4.44).

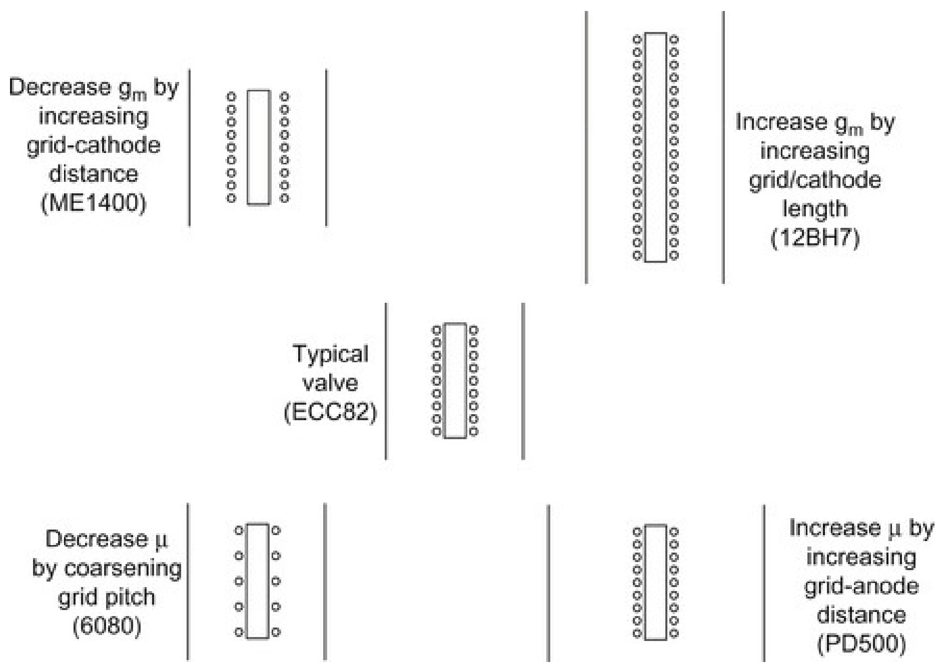


Figure 4.44 The effect of valve geometry on μ and g_m.

As a practical example of the concepts shown in the diagram, two dissected valves having comparable mutual conductance were inspected under a travelling microscope. The 6080 (μ=2) had a grid pitch of ≈1.64 mm per turn, whereas the ECC81 (μ=65) had a pitch of ≈0.16 mm per turn.

As an extreme example of the effect of anode–cathode spacing on μ, good quality valve televisions regulated their extra high tension (EHT) supply because this avoided changes in picture size with brightness. Because their EHT was typically 15 kV, series regulators could not be used (heater/cathode insulation and efficiency problems), so shunt regulators were necessary. No valve has a perfect vacuum, so an increased anode–cathode spacing was needed to avoid arcing, resulting in humungous μ (1050 for the PD500).

Increasing g_m requires that the grid be moved closer to the cathode, but if high μ is also required, then the grid winding pitch must be very fine, necessitating extremely fine wire for a uniform field. Thus, the WE416C (g_m≈65 mA/V and μ≈250) was specified to have a grid with a pitch of 1,000 turns per inch (0.0254 mm), using 0.0003″ (12 µm) diameter wire, spaced 0.0005″ (20 µm) from the cathode [19].

Because the control grid is so close to the cathode, a very small movement of the grid has a significant effect on the flow of electrons, and this is the cause of valve microphony.

Grid Current

Although the control grid is normally a high-resistance point, it can pass positive or negative grid current.

If we charge the grid positively with respect to the cathode, the grid reduces the repulsive effect of the space charge on electron emission at the surface of the cathode, and assists in pulling electrons away from the surface of the cathode. A much higher anode current flows, but some electrons are captured by the grid and flow out into the grid circuit, resulting in positive grid current, which drastically reduces input resistance. This is why Class AB2 output stages, which operate with positive grid current, are invariably preceded by a power driver.

Thermal Runaway due to Grid Current

If the grid is allowed to emit electrons, negative grid current results, and depending on the value of the grid-leak resistor and biassing, the potential of the grid may rise (*lowering* V_{gk}), causing an increase in anode current, and further heating the valve. The emissive material of the cathode then begins to evaporate, contaminating the grid and increasing grid emission. At worst, the grid may become so hot that it slumps and touches the cathode, completing the destruction of the valve, but permanently increased valve noise is inevitable even if the valve is not actually destroyed.

Grid Emission

Cathode stripping has been mentioned earlier as a problem for the cathode, but the stripped cathode material must go somewhere. The control grid is nearest to the cathode, so this sputtering process contaminates it with emissive material, greatly increasing the likelihood of grid emission.

The Richardson/Dushmann equation (see Appendix) shows that there are two ways in which the emission of an uncontaminated grid may be minimised:

- *Reduce grid temperature*: Power valves cool their control grid by winding it on thick copper axial supporting wires that conduct the heat to radiant heatsinks at their ends. Note that a hot anode inevitably increases grid temperature.

- *Increase grid work function*: Electron emission is proportional to the inverse power of work function, so selecting a grid material with increased work function reduces grid emission (Table 4.7).

Table 4.7 Comparison of Emission for Different Grid Materials

Metal	φ	Relative emission at 1,100 K (%)
Tungsten	4.55	100
Gold	5.28	0.05
Platinum	5.63	0.001

As can be seen, a slight change in work function makes a huge difference to emission. Platinum gives by far the best performance, but a pure platinum grid would be expensive and too weak to be rigid. One practical solution was to use platinum-clad molybdenum wire (25% platinum by weight [20]), but even this was expensive. Conversely, thinly gold-plating a grid is cheap, so this technique is common (6080, 6545P).

Frame-Grid Valves

If a horizontal brace is welded from one vertical control grid support rod to the other at both top and bottom, a frame is formed, so these are known as frame-grid valves. The advantage of this construction is that the grid wire can now be *tensioned* across the frame, greatly reducing sag, which enables far more precise grid-cathode positioning of each wire, allowing closer spacing to the cathode, which increases g_m. As an example, the frame-grid E88CC comfortably achieves g_m=10 mA/V, whereas the traditional construction of the otherwise comparable ECC82 struggles to better 2 mA/V.

Variable- μ Grids and Distortion

Variable- μ valves are also known as remote cut-off valves, which refers to the gentle curve of their mutual characteristics, requiring an unusually large negative grid voltage to reduce I_a to 0. Radio receivers have to cope with a very large dynamic range of RF signals because they may be tuned from a strong local signal to a weak distant signal. To allow sufficient gain for the weak signal, but avoid overload on the strong signal, the RF gain of the receiver is made variable by an Automatic Gain Control (AGC) system. Typically, the AGC affects the gain of two or three valves simultaneously, and because the total gain is the product of individual gains, each valve only needs to change its gain by a small amount for the total gain to change significantly. The control grid is deliberately wound with an uneven pitch (which causes μ and g_m to change with V_a), so these valves are known as variable- μ triodes or pentodes (see Figure 4.45).

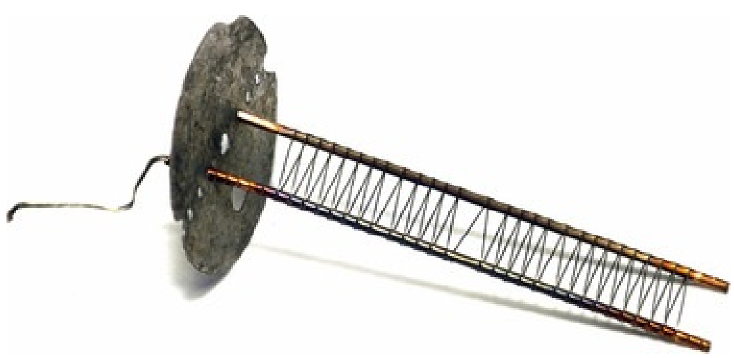


Figure 4.45 Control grid of 6K7 variable- μ pentode. Note the deliberate deviations from constant pitch winding in the centre and a quarter of the way in from each end.

Although variable- μ valves are *designed* to change their gain with V_a, this is not a problem at RF because the signals are so small compared to audio, and distortion is proportional to amplitude.

In an audio triode, distortion is dominated by the variation of r_a with I_a, but once the loadline is flattened (R_L>> r_a) to lower distortion, we rely on constant μ, so the evenness of the grid pitch becomes important. Mechanically, the grid wire is wound and swaged into guide slots cut into the grid support rods whose position was determined by a lead screw.

As the lead screw wears, backlash develops and the position of the guide slots becomes less precise, resulting in an uneven grid winding. Backlash is more of a problem with finer grid pitches (high- μ valves), so it is more difficult to make a low-distortion high- μ valve than a low- μ valve.

Reducing backlash and tightening production tolerances to reduce variation of μ are not insurmountable problems – they just cost money, so this specification would be driven by the application. Fortuitously, it appears that the CV1988 (military 6SN7) required low variation of μ, so it has lower distortion than the commercial 6SN7.

Other Grids

Tetrodes and pentodes have a helical screen grid (g_2) wound concentrically between the control grid and the anode, so a proportion of the cathode current is accelerated into the screen grid, causing its temperature to rise. If the screen grid is coated with zirconium, this not only helps cooling by radiation, but also absorbs residual gas, so the screen grid assists the getter in maintaining a good vacuum.

Beam tetrodes have beam-forming plates rather than a helical wire grid (see Figure 4.46).

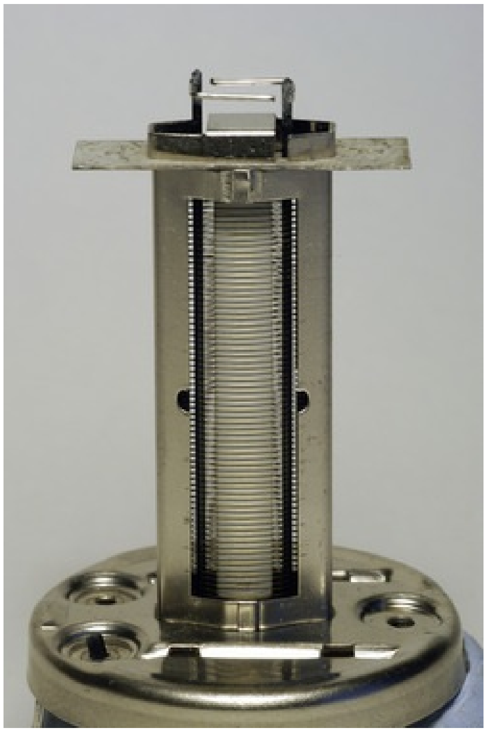


Figure 4.46 Beam-forming plates and cathode/grid structure of QQV07-50 VHF dual beam tetrode.

Electrons are only weakly attracted to the suppressor grid (g_3) in a pentode, so this grid does not self-heat.

The Anode

The anode is constantly bombarded by high velocity electrons. Although these electrons have very little mass, their high velocity means that they possess considerable kinetic energy, which is converted into heat when they are stopped by the anode. An important specification for a valve is therefore the *anode* dissipation, because a hot anode heats the grid, causing grid emission. A secondary effect is that a hot anode releases gas, a phenomenon known as *outgassing*, which contaminates the vacuum.

Occasionally, some early valves used coarsely woven wire mesh anodes, which was claimed to 'prevent the grid becoming overheated by reflected radiation'[21] (see Figure 4.47).

Figure 4.47 True mesh anode of a Philco 37.

The mesh anode idea has been recently resurrected, although the 'mesh anode' 300B is actually pressed from metal sheet having rectangular holes punched through it (see Figure 4.48).

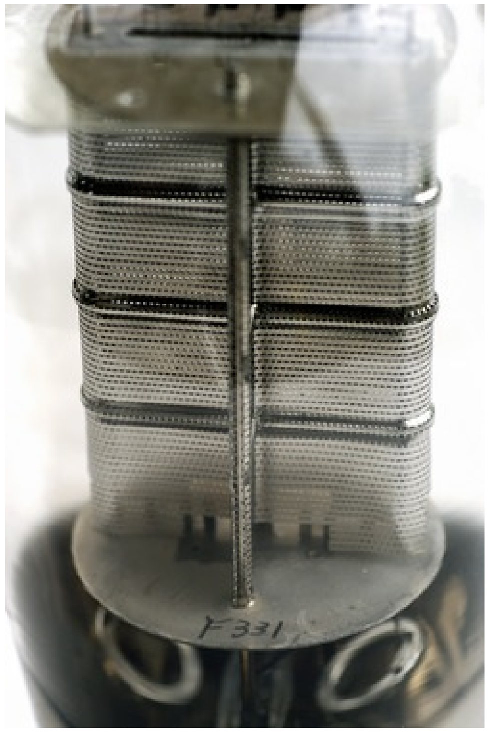


Figure 4.48 Punched plate ‘mesh’ anode 300B.

The anode dissipates heat by radiation, and to maximise radiating area, anodes often have fins. Another way to improve the loss of heat by radiation is to colour the anode black by coating the nickel anode with graphite. Even better, the 813 transmitting tetrode and 6528 series regulator dual triode use solid graphite anodes because they do not warp at high anode temperatures. The manufacturer of the 6528 made a virtue out of the very hot graphite anode by coating it with zirconium, which has a great affinity for hydrogen, nitrogen and oxygen once heated above 800 K (incipient red heat).

The anode’s surroundings must be cool and capable of absorbing radiant heat. Otherwise, they will emit or reflect heat back to the anode, which is perfectly coloured to absorb radiant heat, thus raising anode temperature. Chrome-plated output transformers, etc., may look nice, but they could raise anode temperature. The worst possible surroundings for a valve would be a concentric chrome-plated cylinder, since this would focus the radiant heat back to the anode.

The EF86 low-noise pentode has an electrostatic screen surrounding its anode to reduce hum, and in some examples is formed from a shiny metal sheet, but this should not be confused with the anode. This screen severely restricts anode cooling, but the g_m of the EF86 is quite low anyway, so operating it at a high current (which would increase g_m, but also increases P_a) would be pointless. The EF86 typically operates with a low P_a, so the electrostatic screen does not a cause a cooling problem.

Because electrons collide with the anode at high velocity, there is a possibility of dislodging more than one electron from the anode surface for each electron strike – an effect known as *secondary emission*. If the secondaries barely left the surface of the anode before returning, this would not be a problem, but if they stray any distance, they affect the electric field between the cathode and the anode, causing distortion. The relative level of this emission is determined by the Secondary Emission Ratio (SER) of the material concerned. Nickel has a fairly low SER (≈1.3) and, together with its malleability, this is why it is commonly used for anodes and other pressed sheet valve electrodes. The SER of an electrode can be dramatically reduced by zirconium plating or by graphitising the surface (coating it with colloidal graphite).

Graphite has a very low SER, but is rather fragile, so it can only be used in quite thick (>1 mm) structures. Moulded graphite anodes have increased thermal inertia, allowing $P_{a(peak)} >> P_{a(continuous)}$, so they are popular in transmitter valves such as the 813, 211 and 845. The Mullard QV08/100 has a massive anode structure formed out of two graphite slabs ≈8 mm thick, making an anode that is very tolerant of momentary overloads (see Figure 4.49).

Figure 4.49 Graphite slab anode of QV08/100 tetrode.

The Vacuum and Ionisation Noise

The quality of the vacuum within the valve is critical because initially uncharged gas molecules in the valve are likely to be struck by high velocity electrons on their way to the anode, possibly dislodging electrons to create positive gas ions. Positively charged ions are repelled from the anode, but are attracted to the grid/cathode structure, whereupon they are immediately discharged by a balancing number of electrons flowing up from the external paths to ground. Since the formation of ions and their subsequent discharge by the grid/cathode structure is random, it creates random noise currents leading to *ionisation noise*.

Ionisation noise currents become a problem only when they flow through an external resistance such as a grid-leak resistor. They then develop a voltage across that resistance in accordance with Ohm’s law, and because valves are voltage-operated devices (V_a or $I_a \propto V_{gk}$), the ionisation noise voltage is amplified. If the grid resistance was zero, the ionisation current would be unable to develop a noise voltage.

Low-noise input stages use a high- μ valve, so that subsequent stages do not degrade the noise performance, but a high- μ valve implies fine grid pitch, which greatly increases the probability of ions striking the grid rather than the cathode. Since R_g, the grid-leak resistor, is invariably quite a high value, a significant noise voltage can be developed and amplified. Because the grid effectively screens the cathode, very few ions strike the cathode, so cathode ionisation current is greatly reduced, and because even an undecoupled R_k has a low resistance to ground compared with the grid-leak resistor, this further reduces any noise voltage developed in the cathode circuit. Ionisation noise in high- μ stages is thus dominated by grid ionisation noise current and can be minimised by a low source impedance to ground at audio frequencies, so transformer input coupling reduces the effects of ionisation noise currents at low frequencies compared with capacitor coupling. (At very low frequencies, $Z_{sec} \approx R_{DC(sec)}$, which is quite low, whereas $X_C \approx \infty$, so capacitor coupling produces more 1/ f noise.)

Output stages tend to use low- μ valves to minimise Miller capacitance and preserve bandwidth, but low- μ valves have a coarse grid pitch, biassing the probability of ion strike in favour of the cathode. Low- μ valves operate with high bias voltages (V_{gk}), requiring high values of R_k. The combination of an increased proportion of ionisation current and high R_k means that low- μ valves should not leave their cathodes undecoupled if the effects of ionisation noise currents are to be minimised.

A good vacuum is referred to as being *hard*, whereas a poor vacuum is *soft*.

Therefore, valves are sometimes described as having 'gone soft'. During manufacturing, the air in the valve is pumped out, but some air will remain that cannot be removed by pumps, and the remaining gas is removed by the *getter*.

The Getter

Gas molecules are trapped in the interstices of the metal parts – not dissolved, so heating to red heat can force them out, just before the getter is flashed.

The getter is a metal structure often fitted near the top of the valve, coated with a highly volatile powder (usually a barium compound similar to the cathode emissive surface). Once the valve has been sealed and as much gas as possible has been pumped out, the getter is heated and the powder explodes, consuming the remaining gas. The force of the explosion throws molten barium onto the inside surface of the envelope to give the familiar mirrored coating at the top of the valve. The explosion is initiated electrically, either by directly passing a heating current through the getter's metallic supporting structure (metal envelope valves), or by shaping the getter as a short-circuited turn and using transformer action to induce the heating current from an external RF source at ≈450 kHz (glass envelope valves).

Although some of the getter material is deactivated by the explosion, the getter must continue to consume gas molecules throughout the life of the valve because gas continuously permeates the vacuum, either via leaks at the seals where the leads leave the envelope, or by outgassing from hot structures. The rate of a chemical reaction doubles with each 10 °C rise in temperature, so most valves thermally bond the getter to the hot anode with a thick wire. Because the anode is not at a constant temperature over its entire surface, and the wire has thermal resistance, some valves mount getters via short stubs onto the hottest parts of the anode. If bonded with only one stub, the mass of the getter combined with the spring of the stub can form a mechanically resonant system with a very high Q, so some special quality valves support their getters with two stubs to reduce the Q, and hence microphony.

To be consumed by the getter, the gas molecules must touch it, but this is ensured by entropy, and provided the heater reaches operating temperature before HT is applied to the anode, it should be effective.

Soft valves can often be spotted by the gentle blue glow near the glass envelope, which is due to the collision of ionised gas molecules with the glass. This should not be confused with the blue fluorescence that can be seen on the inner surface of the anode in valves such as the EL84, and which is perfectly normal.

Valves stored for decades in a cold warehouse may have an imperfect vacuum because the getter was too cold to be fully effective. Fortunately, 24 h in a domestic oven at 100 °C warms the getter and will often clear residual gas [22], but beware that the phenolic bases of Octal valves can easily be damaged by a higher temperature. When first used, even new valves should have their heaters powered for a least half an hour before applying HT. Although a few hours of electrical use also clears the gas, the previous methods avoid the damaging ion bombardment suffered by the cathode until the getter has cleaned the vacuum.

Occasionally, rather than air gently seeping in through the Kovar alloy glass to metal seals at the valve's pins, a severely overheated anode may release sufficient gas (known as outgassing) to poison the cathode and turn the getter sputtering from silver/brown to an almost transparent brown. Long-term overheating results in stains on the anode, so the combination of stains and nearly transparent getter sputtering may reveal the damage even before the 'gas' test on a valve tester (see Figure 4.50).

Figure 4.50 The right-hand valve of this pair has almost lost its gettering due to outgassing caused by catastrophic anode overheating.

The Mica Wafers and Envelope Temperature

The electrode structure, heatsinks and getter are supported and held rigidly in position by insulating mica wafers at the top and bottom of the anode structure.

If this mica is not a perfect insulator, then leakage current paths will form, perhaps from the anode to the control grid, which would cause noise in a small-signal valve, but could cause destruction in a power valve.
When the getter is exploded, some molten metal may strike the mica wafer and make it slightly conductive. To lengthen the leakage current paths and increase their resistance, slots are cut in the wafer between the control grid and the anode.
Alternatively, the getter may be positioned such that it is less likely to spray onto the mica wafers, or the electrode supporting wafers may be shielded by a sacrificial mica wafer or metal plate. The designers of the KT88 not only used all of the previous techniques to reduce leakage, but also made the electrode supporting wafers undersize, so that they did not touch the (assumed contaminated) envelope (see Figure 4.51).

Figure 4.51 View of GEC KT88; note the measures taken to reduce leakage currents.

Osram made a big fuss about quality control, perhaps justifiably – their KT88 tolerates a higher anode voltage than a modern one.
Even if the mica wafers have not been contaminated with conductive getter material, mica is not a perfect insulator, and like all insulators its resistance falls with increasing temperature. The Sony C-800G studio condenser microphone used a Peltier effect heat pump to cool the envelope of its pre-amplifier valve. Since the mica wafers are in contact with the envelope, they are also cooled, and it seems probable that the reduction in leakage currents through the wafers, together with reduced anode temperature and consequent outgassing is

responsible for the reported improvement in noise [23].

Not only does reduced envelope temperature reduce noise, but it also improves valve life. To directly quote the ‘Brimar Valves Components Group Mobile Exhibition’ (November 1959) manual, ‘The use of close-fitting screening cans of high thermal conductivity in intimate thermal contact with a large area of the bulb, in conjunction with an adequate heat sink, can materially reduce the operating bulb temperature and very considerably improve the life of the valve.’

Contrast this with a further comment from the same source, ‘The use of screening cans which are not in thermal contact with the valve may seriously interfere with the cooling of the valve.’

A hot envelope implies a hot anode, and because the mica wafers support the anode, the anode conducts heat directly to the wafers. The wafers can be heated by an excessively hot anode to the point that they outgas water vapour (mica unavoidably contains water), which is particularly poisonous to oxide-coated cathodes [24].

Micas can also poison cathodes because of vibration. A vibrating electrode chafes its support holes in the micas, producing fine mica dust [25]. Weight for weight, dust has a far greater surface area than a solid block, so it readily releases water vapour.

Ceramic support wafers are popular in transmitting valves such as the 845, and in ruggedised valves such as the 6384, because it avoids the water vapour problem.

Valve Sockets – Losses and Noise

The author has a good variety of B9A valve sockets in both chassis mounting and PCB types ranging from phenolic through ceramic to PTFE, so their frequency-dependent losses were investigated. After considerable testing, it transpired that the ceramic and PTFE types were essentially indistinguishable, but that other types had measurable defects (see Figure 4.52).

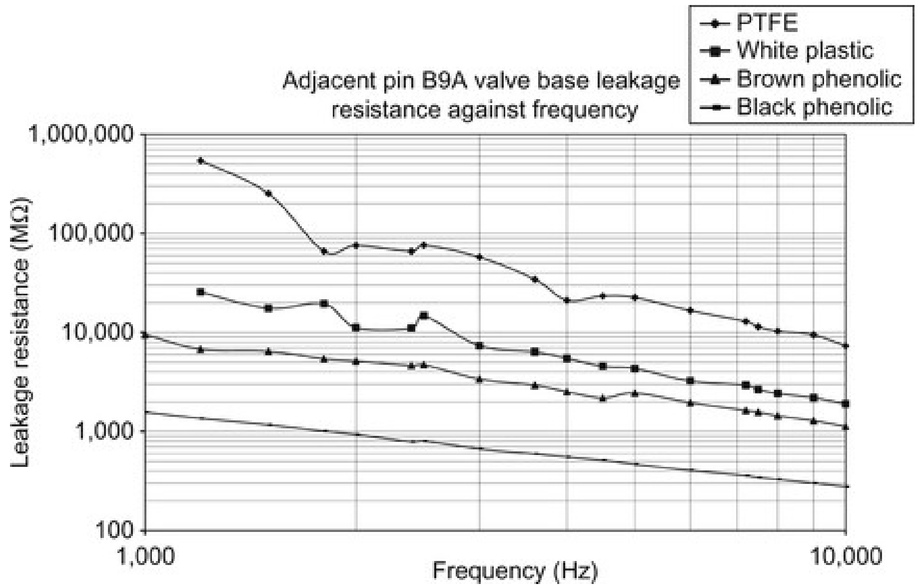


Figure 4.52 Leakage resistance against frequency for different valve socket insulating materials.

A clear trend is apparent – if you can't use a PTFE or ceramic socket, get as far away from black as possible (black is carbon, which is conductive) because DC leakage currents (especially from anode to grid) imply noise and/or distortion in high impedance circuits.

Valve Bases and the Loktal™ Base

Octal valves require an insulating base into which hollow brass pins are riveted, so the author investigated the anode to grid-leakage resistance of 44 triodes (see Figure 4.53).

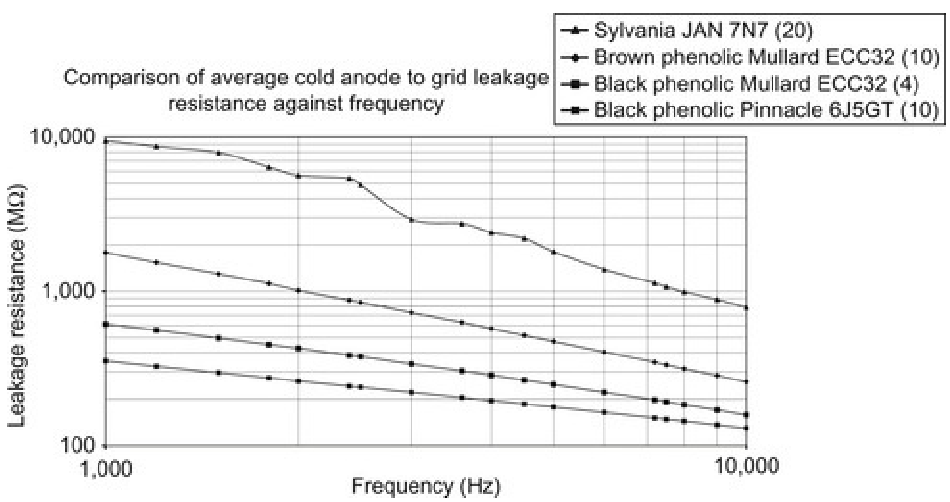


Figure 4.53 Average anode to grid leakage resistance against frequency for different valve base insulating materials (Number in brackets refers to the number of samples tested.)

The results are similar to the valve socket comparison, with the worst base being the black phenolic Pinnacle 6J5GT – which is an interesting result since this triode performed very well in the author’s distortion tests. Clearly, the Pinnacle 6J5GT is more suited to high signal levels where distortion is important than to small signal levels from high impedances where anode to grid leakage and resulting noise is important. However, the clear winner in this comparison is the Sylvania JAN 7N7, which has a Loktal™ base.

Despite its technical superiority over the established octal base, the Loktal™ base introduced by Sylvania in 1938 was an unpopular diversion and an evolutionary dead end, since the resulting valve was little smaller than an Octal valve. To produce a valve that could operate up to 225 MHz (7A4 versus the otherwise equivalent 6J5), capacitances, inductances and leakage paths had to be reduced. Eliminating the glass *pinch* within the valve and bringing electrode support wires directly to the pins shortened the valve, which reduced internal stray capacitances and lead inductance. Eliminating the phenolic base reduced leakage between pins, and the addition of an earthed metal base through which the pins protruded both screened and electrically guarded individual pins. The central metal spigot not only provided keying to ensure correct orientation, but also had a ring that locked the valve into the base once inserted. To avoid trademark infringement, competing companies changed the nomenclature to ‘Loctal’, or referred to them as ‘Lock-in’ valves and designated the base B8G.

For high quality audio, Loctal valves are excellent, since their electrode construction (and consequent distortion) is the same as the Octal generation, but they have the obvious technical superiority of the Loctal base. Since Loctal valves were designed for use above 100 MHz (when socket losses become very significant), PTFE-insulated Loctal bases were made, and are still available.

The 1939 Sylvania manual misleadingly specifies Loktal™ heater voltage as being 7 V or 14 V for 130 V line, allowing them to use the new ‘7’ and ‘14’ prefixes for their type designation, but this is directly equivalent to 6.3 V or 12.6 V for 117 V – their nominal line voltage, so these valves actually have standard heaters. The contrived prefixes were simply a marketing ruse to distinguish the new base.

Brimar [26] noted that leakage currents in the base are a cause of noise, and once poor insulators have been eliminated, the only way to reduce leakage currents is to lengthen the leakage path. Thus, triodes requiring low grid current lengthen the leakage path by bringing their grid out through the top of the valve (ME1400, EC1000), or include a central nipple inside the glass button base for the grid connection (ME1401).

The Glass Envelope and the Pins

The Glass Envelope and the Pins

The envelope maintains the vacuum within the valve; careless handling will crack the valve, and air will enter. The easiest way to crack the envelope is to bend the pins whilst attempting to insert the valve into a new socket. It is therefore a good practice to plug old valves into the new sockets of a new amplifier, and test the amplifier before inserting the expensive new valves. This then spreads the fingers of the socket slightly, and the new valves can be inserted without fear of damage. A hidden problem is that repeated plugging and unplugging can create micro-fractures in the glass near the pins, allowing just sufficient gas to leak into the valve to subtly degrade the vacuum, and increase ionisation noise (for this reason, the British Standards Institution [27] warned against repeated testing of special quality valves). A terminally damaged envelope is easily spotted because the mirror coating due to the getter turns white.

If the envelope is allowed to accumulate fluff, it will be thermally insulated, and the valve will run hotter, with all the dire consequences that entails. Valves should be clean and shiny to promote long life.

Not all electrons accelerated towards the anode are captured, and so a charge can build up on the inside surface of the (insulating) glass envelope. Coating the glass with graphite renders it conductive, and connecting it to the cathode allows the charge to be drained away. Conversely, the internal glass nipple around each pin on a button base lengthens paths and reduces leakage currents between adjacent pins.

The pins are made of Kovar, which is an iron–nickel–cobalt alloy having the same coefficient of thermal expansion as the glass envelope in order that leaks do not occur at the seals as the valve warms. If stored under damp conditions Kovar can rust, so valves should ideally be stored in evacuated plastic bags. Loctal valves were the earliest valves to be introduced with Kovar pins, and their unpopularity at the time means that a 50-year-old NOS valve is not uncommon, so beware of this problem. If you have a good stock of valves, store them in a warm, dry place – not in a shed or garage.

On some valves, the pins may be gold-plated, but this plating will be quickly removed by repeated plugging and unplugging. Gold-plated pins used to be a sign of quality (although Brimar did not always bother to gold-plate their excellent E88CC), but some modern valve manufacturers cheerfully gold-plate selected valves sourced by their standard production line, whereas traditional special quality (Mullard) or trustworthy (Brimar) valves were consciously designed/produced to be better, rather than selected from a standard production

line.
Although gold is corrosion resistant and assists in achieving a good contact with the socket, silver has better conductivity (ρ=1.47×10 $^{-8}$ Ω/m compared with 2.05×10 $^{-8}$ Ω/m for gold and 1.54×10 $^{-8}$ Ω/m for copper). As frequency rises, skin effect causes conduction to occur principally at the surface of the conductor, so the MUSA video connector originated at Post Office telecommunications exchange London *MUS*eum *A* was silver-plated to improve conductivity. Similarly, transmitting valves intended for use at VHF sometimes have silver-plated pins. Sadly, unattended silver corrodes quite badly.

PCB Materials

Glass Reinforced Plastic (GRP) boards are slightly less than ideal for valve audio because of leakage resistance. The leakage occurs because the epoxy resin does not always seal perfectly to the glass fibres and surface tension draws water vapour into the resulting gaps, never to be released. Many years ago, the author built circuits on Synthetic Resin Bonded Paper (SRBP) and felt that they sounded better than the same circuit built on GRP, but at the time he could not see any engineering reason behind it, and put it down to imagination. The crucial difference between GRP and SRBP is that SRBP is porous over its entire surface and not just at its edges. SRBP can therefore lose water vapour over its entire surface as it heats, whereas a GRP board can lose water vapour only at its edges. A warm SRBP board could therefore actually be less leaky (even though nominally a poorer material) than a warm GRP board. Because water is polar, the problem of dielectric leakage becomes even more acute at high frequencies, so microwave and >200 MHz oscilloscope designers have avoided the material for decades, preferring to use PTFE.

References

[1] Scroggie, MG, *Radio and electronic laboratory handbook*. 8th ed. (1971)Iliffe; P475.

[2] Grover, FW, *Bull Bur Stands* **VII** (1911) 495.

[3] In: (Editors: Hughes, LEC; Holland, FW) *Electronic engineer's reference book*3rd ed. (1967)Iliffe.

[4] Thiele, AN, Air-cored inductors for audio, *J Audio Eng Soc* **24** (June (5)) (1976) 374.

[5] McIntosh, FH; Gow, GJ, Description and analysis of a new 50-watt amplifier circuit, *Audio Eng* **December** (1949).

[6] Sowter, GAV, Soft magnetic materials for audio transformers: history, production, and applications, *J Audio Eng Soc* **35** (October (10)) (1987) 760.

[7] Improvements in instruments for detecting and measuring alternating electric currents. British Patent 24,850; 1904.

[8] US Patent 879,532; 18 February 1908.

[9] von Ardenne, M, *Cathode-ray tubes*. (1939)Pitman, New York; P4.

[10] Alley, CL; Atwood, KW, *Electronic engineering*. (1962)Wiley, New York.

[11] van der Ziel, A, *Noise, sources, characterization, measurement*. (1970)Prentice-Hall; P101.

[12] Rosebury, F, *Handbook of electron tube and vacuum techniques*. (1965)Addison-Wesley.

[13] Upton, LS, The rejuvenation of vacuum tubes, *Old Timer's Bulletin* **December** (1973).

[14] Spangenberg, K, *Vacuum tubes*. 1st ed. (1948)McGraw-Hill; P41.

[15] Johnson W. Techniques to Extend the Service Life of High-Power Vacuum Tubes. In: NAB Engineering Conference Proceedings. National Association of Broadcasters; 1994. pp. 285–93.

[16] Improvements in cathode structures for vacuum thermionic tubes. British Patent 209,415; 1924.

[17] RCA. Receiving tube manual, Technical Series RC-19; 1959.

[18] Szepesi, Z, The temperature response of the shot effect of valves with oxide-coated cathode, *Wireless Engineer* **February** (1939) 67–70.

[19] 416C valve data sheet. Western Electric; 1971.

[20] Knoll, M, *Materials and processes of electron devices*. (1959)Springer-Verlag; (assisted by B Kazan.

[21] Henderson FE. Introduction to valves. 2nd ed. Wireless World; 1943.

[22] Jones, M, New life for old valves, *Electron World* **November** (2000) 863–867.

[23] Spencer-Allen, K, Chill factor, *Studio Sound* **November** (1992) 56–59.

[24] Rowe, EG; Welch, P, Developments in trustworthy-valve techniques, *Electr Commun* **September** (1954) 172–188.

[25] 6384 valve data sheet. Bendix Corporation.

[26] Brimar application report VAD/508.3 6BR7; October 1951.

[27] The use of electronic valves. British Standard Code of Practice CP1005.

BSI; 1962.

Recommended Further Reading

- Eastman, AV, *Fundamentals of vacuum tubes*. 3rd ed. (1949)McGraw-Hill; Very little on audio but good for the fundamental physics behind device operation.
- Chaffee, EL, *Theory of thermionic vacuum tubes*. (1933)McGraw-Hill; There are two superb books that major on the electron physics needed to design valves, and this is the older (but more easily obtainable) one. Brush up your mathematics and physics.
- Spangenberg, KA, *Vacuum tubes*. (1948)McGraw-Hill; This is the other electron physics book. Very wide-ranging, it includes special tubes such as klystrons, magnetrons, photomultipliers, and display tubes. Nevertheless, there is still plenty of physics pertaining to audio valves. Superb, but rare.
- Reimann, AL, *Thermionic emission*. (1934)Chapman & Hall; Reimann worked for Marconi-Osram and it shows. Lighter on mathematics than the previous two books, this book is primarily about the practical problems of manufacturing valves and particularly the chemistry.
- Vyse, B; Jessop, G, *The saga of Marconi Osram valve*. (2000)Vyse; Written by two ex-Marconi-Osram employees, but from a more modern perspective and more history than physics, this is a fascinating read because it details valve development and puts it into context.
- Knoll, M, *Materials and processes of electron devices*. (1959)Springer-Verlag; (assisted by B Kazan). This is one of the last books written about valve manufacture and therefore contains all the significant innovations. If you ever wanted to know which materials were used where, and why, this book has the answer, and will probably answer any supplementary questions. Superb.
- Tomer, RB, *Getting the most out of vacuum tubes*. (1960)Sams; Reprinted by Audio Amateur Press (2000). One of the very few books written on the reliability of valves and should be read by all designers. Its slimness belies its worth.
- The use of electronic valves. British Standard Code of Practice CP1005 (1962). Not the world’s most exciting pamphlet, but useful for cautions on valve use. Read in conjunction with Tomer.
- Thrower, KR, *History of the British radio valve to 1940*. (1992)MMA International; Does what it says on the tin. No physics, but a great deal of

history and an exposé of the shenanigans employed by the valve manufacturing cartels to manipulate the market.

Chapter 5. Power Supplies

Valve amplifiers need a DC High Tension (HT) supply and one or more heater, or Low Tension (LT) supplies, which may be AC or DC. Often, the supplies for the pre-amplifier and power amplifier will be derived from the same power supply, which is frequently integral to the power amplifier, but this need not be so. The advent of Power Supply Unit Designer Version 2 (PSUD2) freeware in 2003 has transformed the design of linear supplies, but an understanding of the underlying principles enables much faster convergence to an optimum design.

In this chapter, we will identify the major blocks of a power supply, see how to design them, and then design a complete supply.

The Major Blocks

There are two fundamental types of power supply: *linear* and *switchers* (see Figure 5.1).

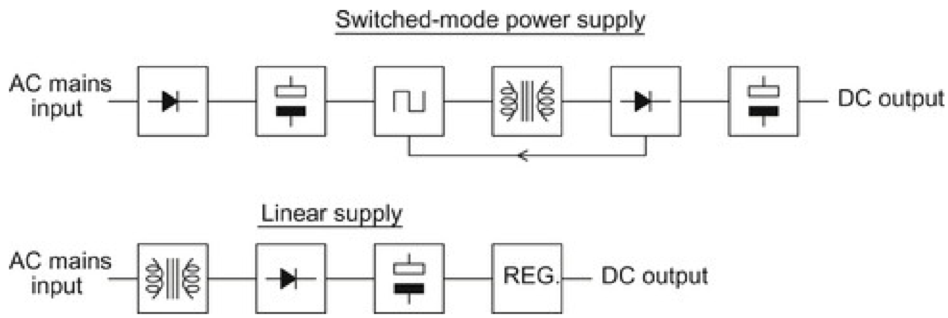


Figure 5.1 Comparison of linear and switcher power supplies.

In a switcher, the AC mains input is rectified, switched at a high frequency, typically >50 kHz, transformed, rectified, and smoothed; regulation is part of the switching element. Switchers are small, light and efficient. Their design is highly specialised, and early designs generated copious Radio Frequency (RF) noise, but designs conforming to modern Electro-Magnetic Compatibility (EMC) standards are surprisingly quiet and can be useful for heater supplies.

By contrast, a linear supply transforms the 50 Hz or 60 Hz AC mains directly, requiring a bulky mains transformer. This is then rectified by valves or semiconductors, smoothed by large capacitors, and possibly even larger inductors, and then regulated if necessary. Linear supplies are heavy and inefficient, but easily designed and low noise. Valve amplifiers use lots of them, so we had better know how to design them.

Power supplies are designed from the output back to the input. Since they are designed after the amplification stages, it is tempting to think of them as an afterthought; indeed, some commercial products reflect this attitude. It is most important to realise that an amplifier is merely a modulator and controls the flow of energy from the power supply to the load. If the power supply is poor and has insufficient energy to meet the amplifier's peak demands, then the most beautifully designed amplifier will be junk.

Rectification and Smoothing

Although we may not have a regulator on the output of the supply, we always have rectification and smoothing. The two functions are inextricably linked and determine the specification of the mains transformer, so this is usually the starting point for design. Since we need to rectify the sinusoidal AC leaving the transformer with maximum efficiency, we will only consider full-wave rectification. Half-wave rectification is not only inefficient (because it only uses alternate half-cycles), but also causes DC to flow through the transformer, and even small DC currents can cause core saturation. A saturated core is lossy and produces leakage flux, which can induce hum currents into nearby circuitry.

Choice of Rectifiers/Diodes

There are two forms of full-wave rectification: the centre-tapped rectifier and the bridge rectifier (see Figure 5.2).

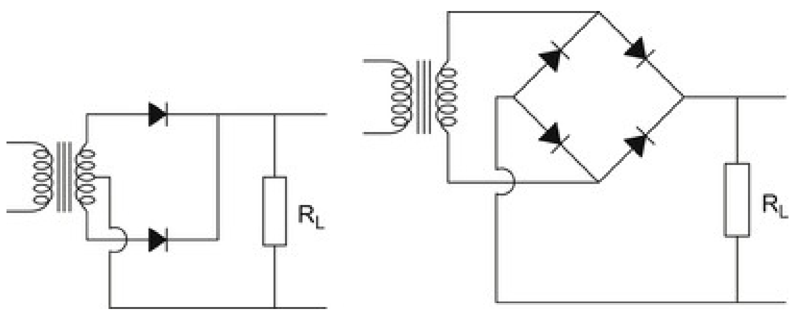


Figure 5.2 Full-wave rectification.

The bridge rectifier is the usual modern topology because it economises on transformer windings. The centre-tapped rectifier was traditional in valve circuits because it economised on rectifiers (which were expensive).

When we consider HT supplies producing V_{DC}<1 kV, we have a choice between silicon and hard vacuum thermionic rectifiers. The GZ34 was the ubiquitous dual diode rectifier, but if we're designing from scratch, the single diodes intended for use in television line scan circuits are cheaper and better, but beware that although the Novar (North American) and Magnoval (European) bases are superficially identical, plugging the 1 mm pins of the 6CL3 into the 1.3 mm receptacles of a Magnoval socket guarantees arcing and unreliability.

Cathode ray tubes for television needed large deflections (120°, or more) to minimise the depth of the tube and used magnetic deflection, so line scan was current drive, rather than voltage (as in electrostatic deflection analogue oscilloscopes). Applying a constant current to an inductor (the line scan coil) causes a voltage ramp, producing the required linear horizontal sweep. Because flyback had to occur in <12 μs (rather than 52 μs for active picture), a correspondingly larger negative current was required. However, this larger negative current was effectively a high amplitude pulse, and it excited all sorts of resonances. One early solution was to *damp* the resonances using a resistor, but that could waste 20 W of power. The more elegant approach was to deliberately excite one half-cycle of resonance and use it as the flyback pulse (who cares about linearity during flyback?), and then use a diode to clamp the following half-cycle. This eliminated the power wasted in the damping resistor, so the clamping diode became known in Europe as an efficiency diode, but a damper diode in the USA. Examples suitable for audio use are 6CL3/6CK3 (USA) and PY500A (Europe).

Sadly, valve rectifiers are inefficient. Not only do they need a heater supply, but they also drop tens of HT volts across themselves and increase the supply's output resistance. They are fragile in terms of ripple current (which we will investigate in a moment), so there is a maximum value of capacitance they can tolerate, and even they need a minimum total resistance in series with each anode to limit this ripple current:

$$R_{series} \geq R_s + n^2 R_p$$

where

R_s=secondary resistance

R_p=primary resistance

n=secondary-to-primary turns ratio.

Although Table 5.1 allows for quick comparison, for detailed design we should refer to a manufacturer's datasheet or full analysis.

Table 5.1 Comparison of Common Hard-Vacuum Rectifier Valves

	$I_{DC\ (max)}$ (mA)	R_{series} (V_{out}=300 V) (Ω)	$C_{(max)}$ (μF)	I_{heater}	V_{heater} (V)
EZ90/6×4	70	520	16	600 mA	6.3
EZ80/6V4	90	215	50	600 mA	6.3
EZ81/6CA4	150	190	50	1 A	6.3
GZ34/5AR4	250	75	60	1.9 A	5
GZ37	250	75	60 [a]	2.8 A	5
6CL3/6CK3	250 [b]	–	–	1.2 A [b]	6.3
PY500A	440 [b]	–	–	300 mA	42 [b]

[a]Mullard did not specify $C_{(max)}$ for the GZ37, but $i_{a(pk)}$=750 mA for both the GZ34 and GZ37, so $C_{(max)}$=60 μF may be assumed.

[b]Per diode, a full-wave rectifier would double this value.

The unpopular EZ90 is only really suitable for pre-amplifiers but is very cheap. EZ80 and EZ81 are cheap and are ideal for pre-amplifiers or small mono power amplifiers. New Old Stock (NOS) Mullard GZ34 and even the current-hungry GZ37 are now scarce, inflating prices. A pair of 6CL3/6CK3 is far cheaper than a GZ34 – and don't forget the even cheaper 12CL3/12CK3, but the price paid for their higher ratings is increased heater power consumption. The PY500A has to be used carefully because its peak current rating is lower than might be expected.

Indirectly heated rectifiers (EZ8* and EZ9* series) are designed to operate from the same 6.3 V heater supply as the signal valves, but this means that $V_{hk} \approx 300$ V, which applies severe stress to the heater/cathode insulation, implying noise currents from the rectifier's cathode into the grounded common heater supply. If low heater noise is paramount, we could transfer the stress from the fragile heater/cathode insulation to the more robust mains transformer by providing a dedicated rectifier heater winding tied to the rectifier cathode, so the GZ3* series forces this design decision by internally tying the indirectly heated cathode to the heater.

Hard vacuum valve rectifiers have only one clear advantage over silicon valve rectifiers, but this advantage may be sufficient to make us tolerate their foibles. The warm-up *rise time* (time taken to change from 10% to 90%) of their output voltage when fully loaded ≈5 s, which greatly reduces the inrush current to electrolytic capacitors in comparison with semiconductor rectifiers (see Figure 5.3).

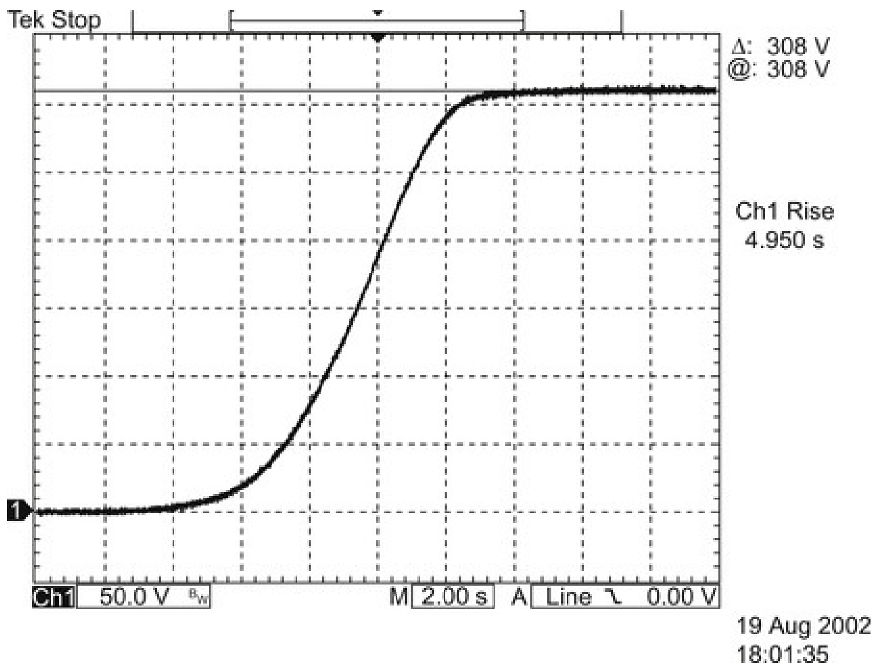


Figure 5.3 Gentle rise of HT supplied by EZ81 rectifier with 120-mA load.

Devotees of hard vacuum valve rectifiers point out that the valve switches on and off more cleanly than silicon valve rectifiers, thus exciting fewer resonances in the power supply, but the author's experience is that both types of rectifier produce switching spikes, and it is the smoothing/snubbing arrangements that are important. If there is an improvement due to valve rectifiers, it is more likely because of their enforced low ripple current.

Whichever rectifier topology we choose, we must ensure that it is capable of withstanding the stresses imposed upon it by the surrounding circuit. When considering low-frequency rectification derived from the mains, we need only specify the voltage and current ratings. However, neither of these ratings is quite as straightforward as it might seem (see Figure 5.4).

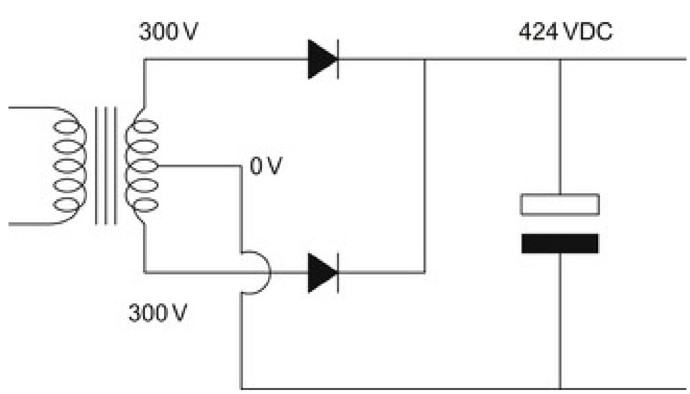


Figure 5.4 Effect of capacitor on rectifier ratings.

The diagram shows a centre-tapped rectifier using silicon diodes to rectify the 300–0–300 V $_{RMS}$ secondary. The off-load voltage at the reservoir capacitor will be the peak AC voltage less diode drop:

$$V_{DC(off-load)} = V_{AC(pk)} - \text{diode drop} = 300\ \text{V} \times \sqrt{2} - 0.7\ \text{V} = 424\ \text{V}$$

Note that this is much higher than if a valve rectifier had been used – silicon diodes and valve rectifiers are *not* directly interchangeable, although for a short period solid state plug-in replacements were made (see Figure 5.5).

Figure 5.5 Solid-state plug-in replacement for valve rectifier.

The diode voltage rating that concerns us is the reverse voltage rating, known as V_{RRM} (Reverse Repetitive Maximum), also known historically as PIV (Peak Inverse Volts). A good safety factor to allow for mains spikes is twice the peak incoming voltage, which is why the 5 A 1,200 V STTA512F silicon diode is so useful for valve electronics (Table 5.2).

Table 5.2 Comparison of Diode V_{RRM} Ratings Needed for Full-Wave Rectification

	Diode rating (V_{RRM}/ V_{RMS})	V_f
Centre-tapped	2√2	1
Bridge	√2	2

When rectifying high voltages, the centre-tapped rectifier has the disadvantage that it needs diodes having double the V_{RRM} rating. As an example, a 300–0–300-V transformer with centre-tapped rectifier would require diodes having

V_{RRM}>849 V, but a single 300 V winding plus bridge rectifier could produce the same output voltage, and would only require diodes having V_{RRM}>424 V. It is comparatively easy to manufacture a hard vacuum rectifier with a high V_{RRM} rating, so centre-tapped rectifiers tend to be valve, whereas bridge rectifiers tend to use silicon.

We can stack ' *n*' silicon diodes in series to multiply V_{RRM} by ' *n*' if necessary, but unmatched diode 'off' capacitances could cause one diode's V_{RRM} to be exceeded, causing failure. This problem can be avoided by adding a parallel capacitance of about 10 times the diode's low-voltage 'off' capacitance across each diode so that individual reverse voltages are defined by the (perhaps 5% tolerance) external capacitor rather than the (rather looser) internal diode capacitance. Thus, the 1,200 V STTA512F would require a 4.7 nF 1,200 V film capacitor to swamp its ≈600 pF low-voltage 'off' capacitance.

Although most popular in valve electronics, the centre-tapped rectifier is occasionally useful in low-voltage/high-current silicon circuits because of the lower forward diode drop; only 1 V_f, compared to 2 V_f for the bridge.

Valves such as GZ34, EZ81, EZ80, *etc.* that are intended for use in centre-tapped rectifiers naturally require a centre-tapped transformer, but a hybrid valve/semiconductor rectifier circumvents this problem [1] (see Figure 5.6).

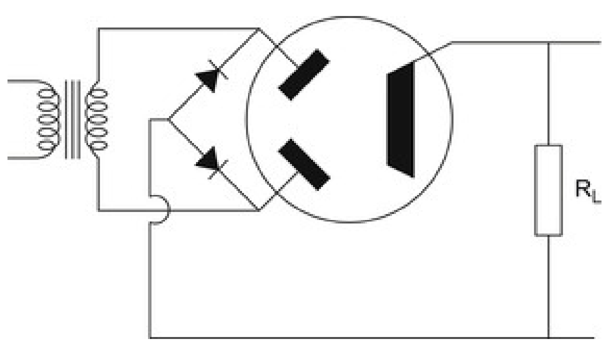


Figure 5.6 Hybrid valve/semiconductor rectifier.

When the diode feeds a reservoir capacitor, current pulses many times greater than the DC load current flow. Fortunately, modern silicon diodes are designed with these peaks in mind, and it is usually sufficient to choose a current rating for each diode equal to the DC load current divided by the number of polarities used. Thus full-wave rectification allows a halving of diode rating because the load current is shared between the diodes of the two polarities, whereas half-wave rectification forces the load current to be supplied from a single diode. If you are lucky enough to have three-phase mains (Germany), you can further divide by the number of phases, allowing six 6CL3/6CK3 to deliver 1.5 A.

Rectifiers To Be Avoided (Gas)

The first gas in common usage was (poisonous) mercury vapour [2] , but it was supplanted by xenon (a noble gas, and therefore non-poisonous). Despite this, the soft blue glow of mercury vapour rectifiers has recently become mildly fashionable.

Because the valve is filled with a low pressure gas, as an electron accelerates away from the heated oxide-coated cathode, it is likely to strike a gas molecule, displacing a second electron. The two electrons accelerate towards the anode, and the process repeats. It is easy to see that this process could occur many times before the anode captures all the electrons generated by a single electron leaving the cathode. This mechanism of *gas amplification* reduces slope resistance and forward drop compared to hard-vacuum rectifiers, making them useful at higher currents.

Rectification relies on vapour, yet mercury vapour quickly condenses as metallic droplets, so it must be evaporated by the heater and the correct vapour pressure reached before HT can be safely applied, as shown in Table 5.3 .

Table 5.3 Delays Needed by Mercury Vapour Rectifiers

Pre-Heat Time	Ediswan [3]	Mullard [4]
After storage/mechanical disturbance	>15 min	>30 min
Day-to-day	>60 s	>60 s

To avoid flashback, mercury vapour rectifiers are typically only rated to operate between 20 °C and 60 °C, although some can only tolerate ≤50 °C, so a fan could be needed to disperse hot air from other components.

Unfortunately, gas amplification is not entirely a bonus. Each gas molecule that releases an electron becomes a positively charged ion and is accelerated towards the cathode, whereupon it is discharged by an electron. But even the lightest gas ion consists of a proton (1,836 times the mass of an electron), so gas ions are slow, and more complex ions are heavier and slower.

Imagine an electron striking a gas molecule close to the surface of the anode and displacing an electron so that the two electrons almost immediately reach the anode, but the ion still has to reach the cathode to be discharged by an electron. Thus, any change in V_a that would switch a vacuum valve off is delayed by the transit time of gas ions from anode to cathode because these ions determine when the last electron is emitted from the cathode, so gas valves switch off far slower than they switch on due to the charge storage caused by the lower mobility of the heavy ions.

One practical consequence of the slow switch-off or current overshoot of gas valves is that they are likely to oscillate if the anode lead is not snubbed with lossy ferrite beads or RF chokes, and they may need to be enclosed by a metal screening can. The best way to detect oscillation is not with an oscilloscope, but by moving an AM radio nearby and listening for a buzz as the radio approaches the rectifier.

As Table 5.4 shows, mercury vapour rectifiers exhibit most of the nastier disadvantages of silicon and valves then add some uniquely their own, but they do produce a pretty blue glow. If absolutely necessary, perhaps an equally pretty effect could be obtained by fitting a blue LED in the base of a 6CK3/6CL3?

Table 5.4 Summary of Mercury Vapour Rectifier Features

Disadvantages	Advantages
Mercury vapour is poisonous	Pretty blue glow
That (poisonous) mercury vapour is contained by a fragile glass envelope	Slope resistance comparable to silicon
Needs heater supply	
Needs delay before HT can be safely applied	
Charge storage due to slow ions may cause oscillation	
Limited temperature range	
That pretty blue glow includes ultraviolet light, which can damage eyesight	

Rectifiers To Be Avoided (Selenium)

The author recently had the dubious privilege of testing an International Rectifier E150L 130 V 150 mA selenium rectifier. (Note that this is an American device, so the similarity to European Pro-Electron numbering is purely coincidental.) The device begins to turn on at about 4.5 V but has a forward drop of 15.8 V at its rated current of 150 mA; this is slightly better than an EZ81 (valve) at the same current (19.8 V), but the EZ81 has 10 times the reverse voltage rating! The reverse current is best not talked about in polite company; 7.92 mA at 100 V implies ≈10 mA at the rated voltage of 130 V (see Figure 5.7).

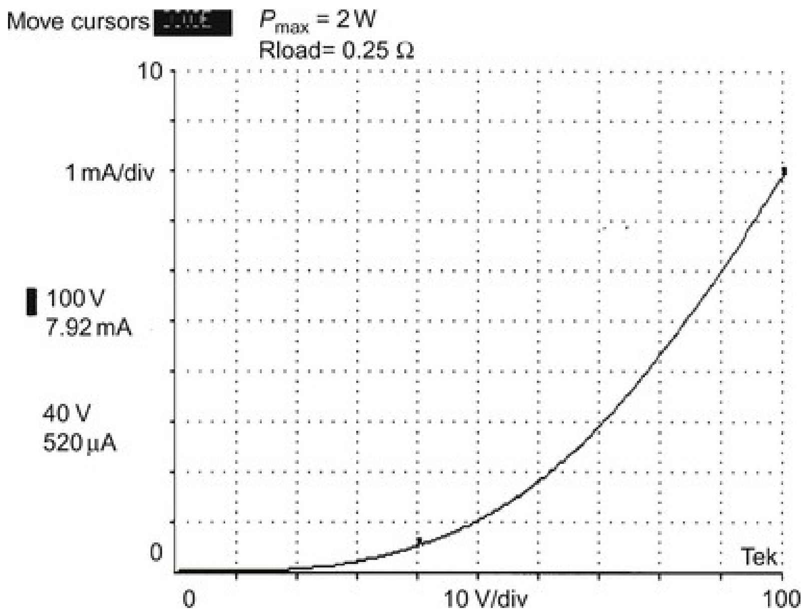


Figure 5.7 E130L selenium rectifier forward current for applied forward voltage.

These devices are recognisable by the square fins needed not just to dissipate the power wasted in forward mode – the power dissipated when the device should be *off* is also significant (10 mA×130 V=1.3 W!).

Selenium rectifiers fail short-circuit and cause wholesale smelly destruction when they do so. Take them out before they take out the mains transformer, but beware that their replacement will almost certainly need some added series resistance to maintain the original design voltage.

Rectifiers To Be Avoided (Copper Oxide)

The forward characteristics of copper oxide are only a little worse than those of silicon, which is why they were used for low voltages. When tested, reverse current of a WB 1/6 A rectifier was 10 µA at 0.5 V but rose to 1.14 mA at 60 V, not quite as alarming as selenium, but a long way from silicon. Copper oxide rectifiers are not often seen but look like a collection of stacked washers on a central screw.

RF Interference/Spikes

Rectifiers are switches. Although the following argument implies a resistive load on the rectifier, the results are also valid for the load presented by a reservoir capacitor.

As the input AC waveform rises through 0 V, one or more diodes will switch on, and stay switched on until the waveform falls through 0 V, when the other diode, or diodes, switch on. All diodes need a minimum forward bias before they can conduct, even if it is only the 0.7 V required by silicon. This means that there is a dead zone symmetrically about 0 V where no diodes conduct. The transformer, which is inductive, has been switched off, and it tries to maintain the current flow, but in doing so, it generates a spike of Electro Motive Force (EMF):

$$E = -L\frac{dI}{dt}$$

Fortunately, diodes don't switch quite as abruptly as postulated and there are usually plenty of stray capacitances within the transformer to prevent this EMF from rising very far, but if we are unlucky, the shock applied to the system can excite a resonance resulting in a damped train of oscillations. Using a search coil, the author once observed 100 Hz repetition rate bursts of 200 kHz leaking from a mains transformer for this very reason. Happily, the problem can usually be cured by bypassing each individual rectifier diode with a 4.7 nF film capacitor having a voltage rating equal to the diode V_{RRM} rating. Very occasionally it may be necessary to add a series resistor to each snubbing capacitor to prevent ringing.

Whether we use a bridge rectifier or a centre-tapped rectifier, we still apply the same waveform to the succeeding circuit. The waveform, although it is of only one polarity, is not a smooth DC. The function of the smoothing element is to reduce the ripple, either to a satisfactory level, or to a level such that a regulator can cope with it.

The Single Reservoir Capacitor Approach

The simplest way of smoothing the output of the rectifier is to connect a reservoir capacitor across it, and feed the load from this reservoir (see Figure 5.8).

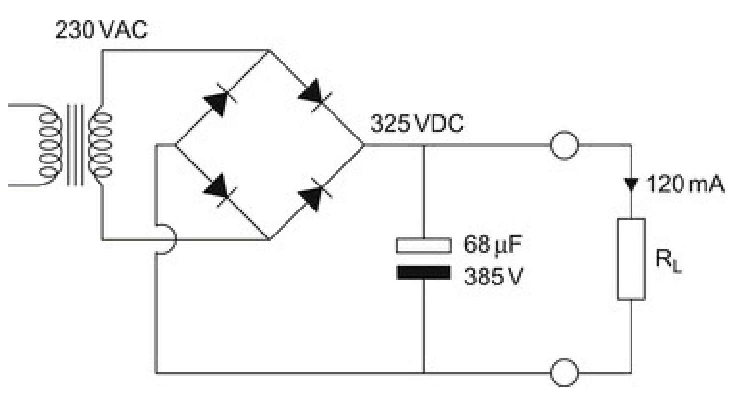


Figure 5.8 Power supply using reservoir capacitor.

Assuming no load current, the capacitor must charge to the full peak value of the AC leaving the transformer ($V_{sec}\times\sqrt{2}$).

Ripple Voltage

The output of the rectifier tops up the charge in the capacitor every cycle, so that at the peak of the waveform, the capacitor is fully charged. The voltage from the transformer then falls away sharply, so the rectifier diodes switch off. Load current is now supplied purely from the capacitor, which discharges exponentially into the (assumed resistive) load until the transformer output voltage rises sufficiently to recharge the capacitor and restart the cycle (see Figure 5.9).

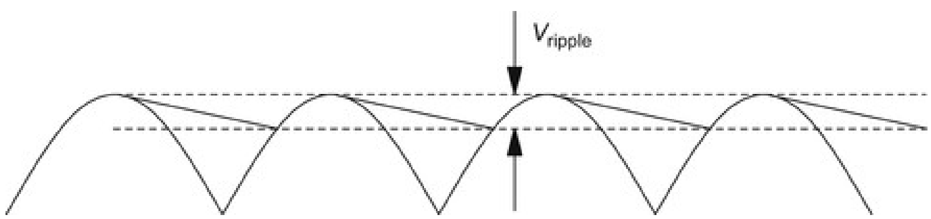


Figure 5.9 Ripple voltage across reservoir capacitor caused by charge/discharge cycle.

Although the reservoir capacitor theoretically discharges exponentially, for any practical value the discharge curve may be taken to be a straight line. (If the load is a series regulator, the discharge curve truly *is* a straight line because any circuit that supplies a constant load current with a regulated constant voltage with negligible wasted current must be a constant current sink.) Given this approximation, it is easy to calculate what the output ripple voltage will be.
The charge stored in a capacitor is:

$$Q = CV$$

The total charge, due to a current *I*, flowing for time *t* is:

$$Q = It$$

We can combine these equations:

$$CV = It$$

Rearranging:

$$V = \frac{It}{C}$$

This equation gives the voltage *change* on the capacitor due to the capacitor supplying current I, for time t. If mains frequency is 50 Hz, then each half-cycle is 0.01 s. If we now make another approximation, and say that the capacitor supplies current *all* of the time, then t=0.01 s. We now have a useful equation:

$$V_{\text{ripple(pk-pk)}} = \frac{0.01I}{C}$$

It might be thought that this equation is of little use, since two sweeping approximations were used to derive it, but as the reservoir capacitor is nearly always an electrolytic capacitor, whose tolerance could be ±20%, we would need a very inaccurate equation before it approached the error introduced by component tolerances!

We can now calculate the ripple voltage at the output of our example circuit in Figure 5.8 , which had a 68 μF capacitor and a load current of 120 mA:

$$V_{\text{ripple}} = \frac{0.01 \times 0.12}{68 \times 10^{-6}} = 18\,\text{V}_{\text{pk-pk}}$$

which is about 5% of full voltage, a good design choice.

The previous method produces sensible results provided that the ripple voltage is between 5% and 20% of the total voltage (it is unusual to allow ripple voltage to rise above this limit).

The Effect of Ripple Voltage on Output Voltage

The reservoir capacitor charges to the voltage peaks leaving the rectifier, so the ripple voltage is subtracted from this and reduces the output voltage. The output voltage V_{out} can be considered to be made up of two components: V_{DC} which is pure DC and V_{ripple} which is the superimposed AC ripple voltage. The significance of making this distinction is that subsequent filtering blocks the AC component to leave only the DC component.

$$V_{\text{out}} = V_{\text{DC}} + V_{\text{ripple}}$$

The AC ripple voltage swings symmetrically about V_{DC} and, at its positive peak, reaches V_{peak}, therefore:

$$V_{\text{DC}} = V_{\text{peak}} - \frac{V_{\text{ripple}}}{2}$$

Considering our previous example, which had V_{ripple}=18 V and V_{peak}=325 V, the DC voltage that would be seen after subsequent perfect AC filtering would be:

$$V_{\text{DC}} = 325\,\text{V} - \frac{18\,\text{V}}{2} = 316\,\text{V}$$

Summarising, V_{DC} is always reduced by a factor of half the ripple voltage.

Ripple Current and Conduction Angle

Now that we have looked at ripple voltage, we need to look at ripple current. This is the current required by the capacitor to fully recharge it every half-cycle. To do this, we need to find the conduction angle, which is the half portion of the cycle for which the diodes are switched on and the capacitor is charging (see Figure 5.10).

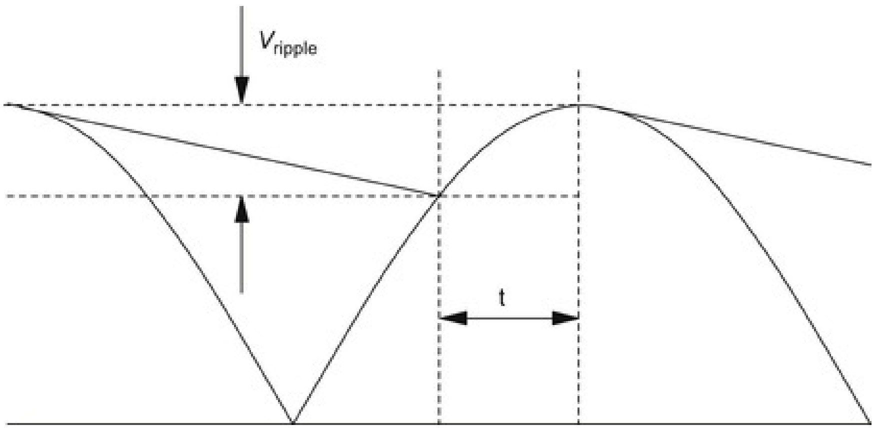


Figure 5.10 Determination of conduction angle from ripple voltage.

To do this, we work backwards from the time that the capacitor is fully charged. We know the ripple voltage, so we can find the absolute voltage on the capacitor at the instant that the diodes switch on. The instantaneous voltage at the output of the rectifier (ignoring the polarity) is:

$$v = V_{\text{peak}} \cos(\omega t)$$

At the instant that the diodes switch on, the capacitor voltage must be:

$$V_{\text{peak}} - V_{\text{ripple}} = V_{\text{peak}} \cos(\omega t)$$

Rearranging:

$$\frac{V_{\text{peak}} - V_{\text{ripple}}}{V_{\text{peak}}} = \cos(\omega t)$$

$$\omega t = \cos^{-1}\left(\frac{V_{\text{peak}} - V_{\text{ripple}}}{V_{\text{peak}}}\right)$$

$$t = \frac{1}{2\pi f} \cdot \cos^{-1}\left(\frac{V_{\text{peak}} - V_{\text{ripple}}}{V_{\text{peak}}}\right)$$

If we now put some figures into this equation from our earlier example (Figure 5.8), *remembering to work in radians, and not degrees*:

$$t = \frac{1}{2 \times \pi \times 50} \cdot \cos^{-1}\left(\frac{325 - 18}{325}\right)$$
$$t \geq 1 \text{ ms}$$

Note that 50 Hz was used in the above equation despite the fact that full-wave rectification of 50 Hz produces 100 Hz ripple. The reason is that full-wave rectification simply inverts the polarity of alternate half-cycles, so the shape and timing of each half-cycle pertains to the original AC frequency.

The capacitor draws current from the mains transformer only for 1 ms in every 10 ms, or 10% of the time. We should therefore expect this ripple current to consist of short, high current pulses (see Figure 5.11).

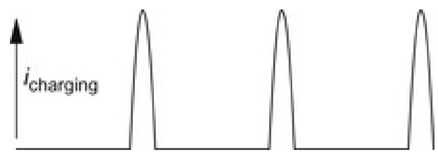


Figure 5.11 Ripple current waveform.

We can now find the ripple current using the relationship:

$$i = C\left(\frac{dV}{dt}\right)$$

But we need an expression for d *V*/d *t*, so we start with our original expression:

$$v = V_{peak}\cos(\omega t)$$

Differentiating:

$$\frac{dV}{dt} = -\omega V_{peak}\sin(\omega t)$$

And substituting:

$$I_{ripple} = -\omega C V_{peak}\sin(\omega t)$$

If we now put some values into this equation:

$$I_{ripple} = 2 \times \pi \times 50 \times 68 \times 10^{-6} \times 340 \times \sin(2 \times \pi \times 50 \times 1 \times 10^{-3})$$
$$I_{ripple(pk)} = 2.2 \text{ A}$$

which is a factor of 18 greater than the 120 mA load current!

Quick check: Charge is equal to current multiplied by time, which would be *area* on a graph of current against time. If the capacitor has to charge in a tenth of the time that it takes to discharge, then it is reasonable to suppose that it will require

10 times the current ($Q = It$). This brings us to 1.2 A. However, we observed earlier that the shape of the charging pulse is not rectangular, and because the area under this pulse is smaller than that of a rectangle of equivalent height and width, this accounts for the final difference in the two answers.
Summarising, the answer is unexpectedly large, but believable.
In practice, peak ripple current is reduced by:

- Transformer core saturation
- Series resistance made up of diode slope resistance, capacitor Effective Series Resistance (ESR), wiring resistance and transformer winding resistance (secondary and reflected primary).

As a result of these factors, peak ripple current is typically between four and six times the DC load current. As a measured example, a transformer with silicon bridge rectifier and capacitor input filter producing 108 V_{DC} loaded by a resistor drawing 35 mA_{DC} drew $I_{ripple(pk)}$=160 mA, a ratio of 4.6:1.
Valve rectifiers have far higher internal resistance than silicon rectifiers and usually need additional series resistance in deference to their limited ripple current ratings, so their ratio of I_{ripple}/I_{DC} is even lower. A DC 50 MHz Tektronix TCP202 current probe was used to investigate a 300 V HT power supply using a GZ34 to feed its 47 μF polypropylene reservoir capacitor and the load current of 88 mA_{DC} caused $I_{ripple(pk)}$=340.3 mA, a ratio of 3.9:1 (see Figure 5.12).

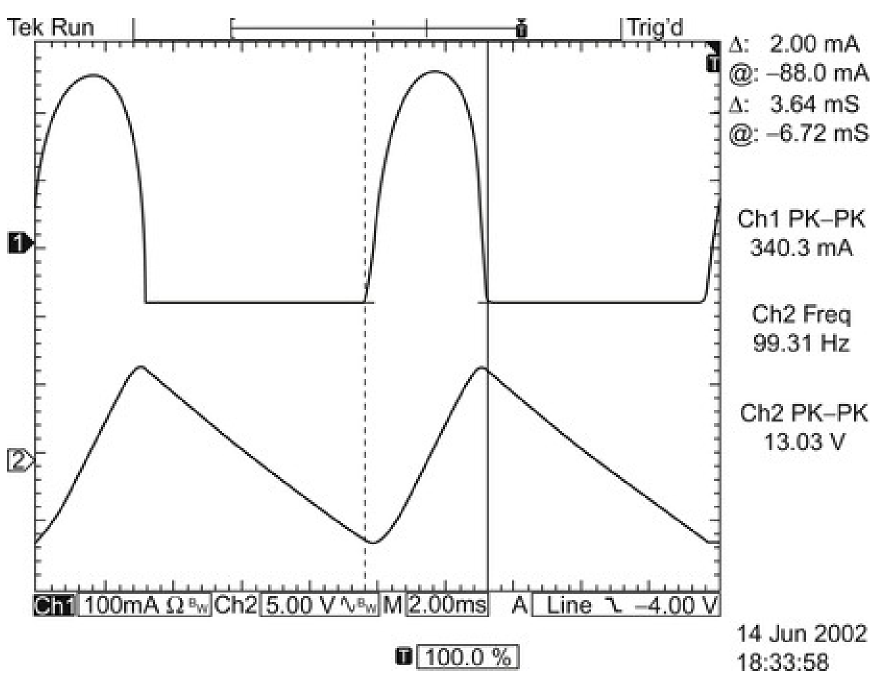


Figure 5.12 Captured reservoir capacitor waveforms due to 88 mA load current. Upper trace (Ch1): current waveform (I_{pk}=340 mA). Lower trace (Ch2): ripple voltage (13 $V_{pk–pk}$).

The ripple current pulses contain harmonics of 100 Hz that theoretically extend into low RF. Fast Fourier Transform (FFT) mode was selected on the oscilloscope, allowing spectrum analysis of the reservoir capacitor ripple current (see Figure 5.13).

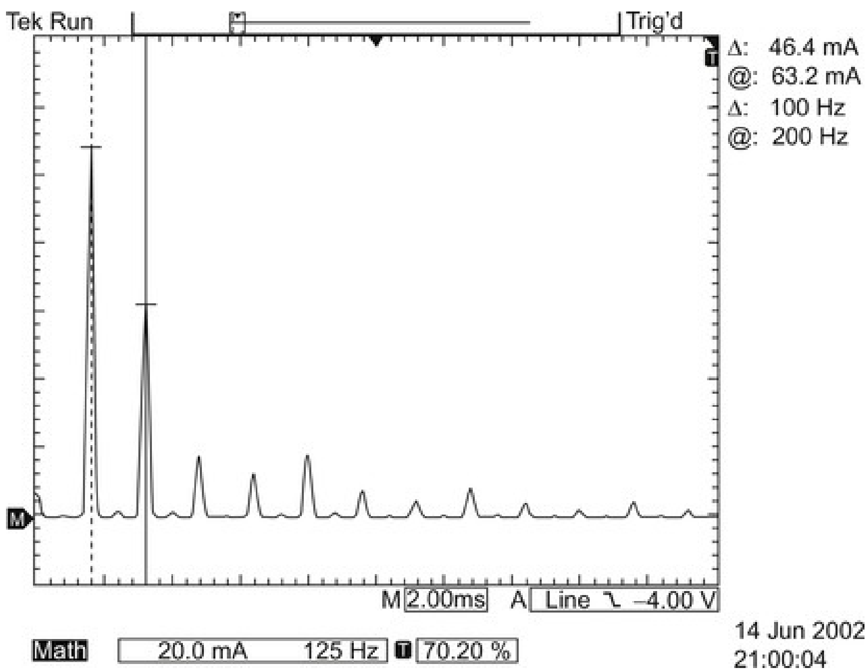


Figure 5.13 Spectrum of reservoir capacitor ripple current.

The spectrum sweeps linearly from DC (left) to 1.25 kHz (right), so the dominant 100 Hz fundamental can be clearly seen, followed by a train of harmonics. Although the 20 μA/div linear vertical scale of this FFT implies that the harmonics die away rapidly, logarithmic scaling revealed that harmonics at 2.5 kHz were only 45 dB below those at 100 Hz. Audio engineers might complain about mains noise, but their power supplies are not innocent.

Transformer Core Saturation

Toroidal transformers are more susceptible to core saturation as a direct result of their more nearly perfect magnetic design. Whether mains or audio, power transformer cores are normally made of Grain Oriented Silicon Steel (GOSS), which has the advantage of allowing a higher flux density in the direction of the grain. Traditional stacked EI cores are unable to take full advantage of this, since there is always a region where the flux is at right angles to the grain, but C cores and toroids have all their flux aligned with the grain and can operate closer to saturation (permitting a smaller core), so this is why toroids are smaller and cheaper. Consequently, toroids saturate sharply, whereas EI cores have a much gentler limit.

Transformer core saturation is undesirable because it releases a leakage field of magnetic flux to induce currents in nearby circuitry. Even worse, this saturation happens cyclically (100 Hz or 120 Hz) and so produces *bursts* of interference with harmonics extending to radio frequencies. Sharper saturation produces a greater proportion of higher harmonics in the leakage field.

This is not merely an apocryphal tale of woe. Many years ago, the author tore his hair out searching for the source of (video) hum in a picture monitor, only to find that the cause was a saturating mains toroid inducing hum directly into the neck of the picture tube.

Choosing the Reservoir Capacitor and Transformer

If we have designed our supply to have a ripple voltage of 5% of supply voltage, then for 90% of the time the transformer is disconnected, and the output resistance of the power supply is determined purely by the capacitor ESR and associated output wiring resistance. This is why changing reservoir capacitors from general purpose types to high ripple current types produces a noticeable effect on the sound of an amplifier; they have a lower ESR (but a higher price).

The transformer/rectifier/capacitor combination is a non-linear system. This makes its behaviour considerably more complex than the ideal Thévenin source, so we need to investigate it over different periods of time.

In the short term (less than one charging cycle), the output resistance of the supply is equal to capacitor ESR plus wiring resistances. This will be true even for very high current transient demands, which may appear in each and every charging cycle, provided that they do not significantly deplete the charge on the capacitor. All that is required is that the capacitor should be able to source these transient currents. To be able to do this, the capacitor needs a low ESR, not just at mains frequencies, but also up to at least 40 kHz, because a Class B power output stage causes a rectified (and therefore frequency doubled) version of the audio signal to appear on the power supply rails. (See Chapter 6 for explanation of Class B.) We can cope with this requirement by using an electrolytic capacitor designed for use in switched mode power supplies as the main reservoir.

A power amplifier may significantly deplete the charge in the reservoir capacitor, causing output voltage to fall either by drawing a sustained high current, due to a continuous full power sine-wave test, or by reproducing a short, but loud, sound – such as a bass drum.

Supplying a constant load is relatively easy, because we know exactly how much current will be drawn, and we simply design for that current. If the ripple voltage for a sensible ripple current is higher than we would like, then we simply add a regulator to remove it.

The difficulties start when we want to supply a changing load. It might seem that if the power amplifier is rated at 50 W continuous into 8 Ω, then all we have to do is to calculate what load current that implies, and design for that current. The drawbacks of this approach are more easily demonstrated using a transistor amplifier, where the load is directly coupled to the output stage and the power supply is very simple (see Figure 5.14).

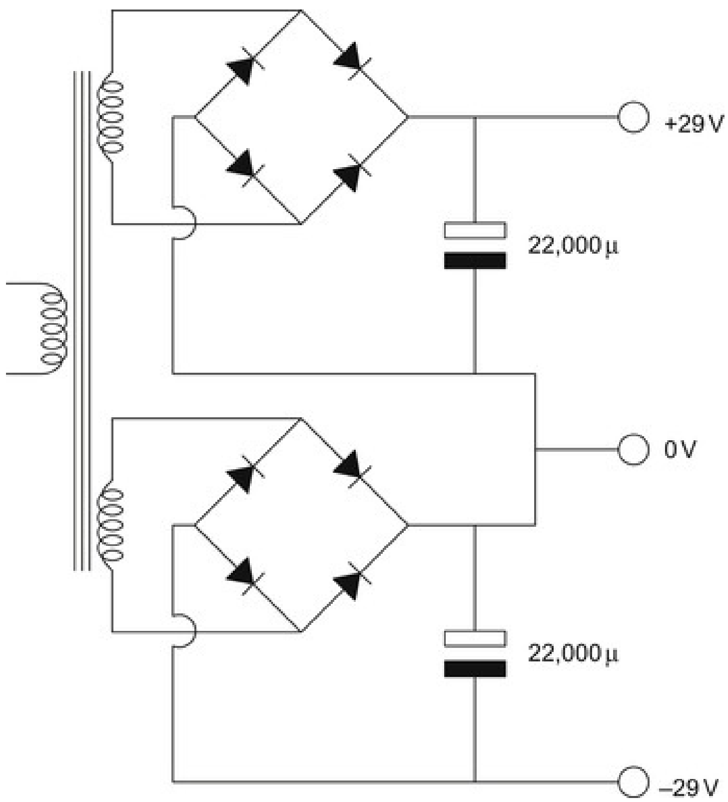


Figure 5.14 Typical power supply for transistor amplifier.

Considering our 50 W 8 Ω example:

$$P = I^2 R$$

Therefore, for a sine wave:

$$I_{\mathrm{RMS}} = \sqrt{\frac{P}{R}} = \sqrt{\frac{50}{8}} = 2.5\ \mathrm{A}$$

But we have to supply the *peak* current, which is √2 greater, at 3.5 A:

$$V_{\mathrm{RMS}} = \sqrt{PR} = \sqrt{50 \times 8} = 20\,\mathrm{V}$$

But we have to supply the *peak* voltage, which is √2 greater, at 28.3 V. Transistor amplifiers can typically swing to within about a volt of rail, so we might just tolerate ±29 V rails, and a power supply capable of delivering ±29 V at 3.5 A is implied. We therefore need 203 W per channel and 406 W for a 50 W stereo amplifier! This is a very large and expensive power supply, and we would need some astonishingly good reasons for using it.

The key to the problem lies in the class of the output stage. If the output stage operates in Class A, then the quiescent current equals the peak current required at maximum power output, in this case, 3.5 A. If each channel genuinely draws a constant 3.5 A from the ±29 V power supply, then we really do need the 406 W power supply. The classic Krell KSA50 50 W stereo amplifier drew 300 W from the mains at idle [5] , suggesting that it wasn't quite true Class A, but it was certainly far closer than most Class A pretenders.

The reservoir capacitor value was easy to determine using our earlier formula and 5% ripple voltage criterion, but the transformer is quite a different matter. It is possible to determine the requirements of the transformer exactly, using the graphs originally devised by Schade [6] . In practice, the required transformer information may not be available, so a practical rule of thumb is to make the VA rating of the transformer at least equal to the required output power.

If our example stereo 50 W amplifier output stage becomes Class B, then each channel still supplies 3.5 A to the load on the crests of the sine wave, but at other points in the cycle the required current from the power supply is much lower. The effect of the reservoir capacitor is to average the fluctuating current demand, and for a full-wave rectified sine wave:

$$I_{\text{average}} = \frac{2}{\pi} \cdot I_{\text{peak}} = 0.637 I_{\text{peak}}$$

The average supply current is 2.2 A, so a 250 VA transformer would be chosen. We could further argue that the amplifier does not operate at full power all the time and that the short term musical peaks requiring maximum output power do not last long. A smaller transformer could therefore be used, since the reservoir capacitor could supply the peak currents. This is a very seductive argument, and many commercial amplifier manufacturers have been persuaded by it, since £1 extra on component cost generally adds £4–5 to the retail price.

We do not have to work to such tight commercial considerations and, within reason, the bigger the mains transformer, the better it is.

Back-to-Back Mains Transformers for HT Supplies

This idea pops up from time to time as a cheap way of obtaining the high voltage needed for the HT supply. As an example, a 240 V:6 V mains transformer provides 6 V to the valve heaters (5% low, but perhaps tolerable) and an identical 240 V:6 V transformer is connected with its 6 V winding across the first transformer's 6 V winding, with the intention of producing an isolated 240 V to be rectified for the HT. If transformers were ideal, then using one transformer to step down followed by another to step up would be fine. However…

The first problem is that a practical 240 V:6 V transformer does not have the 40:1 turns ratio expected of an ideal transformer. The transformer manufacturer knows that there will be resistive losses in the primary and secondary, so the turns ratio is reduced to ensure that the rated secondary voltage appears at full load current. Transformers having a low VA rating invariably have poor regulation. Regulation is defined as:

$$\text{Regulation}(\%) = 100 \times \left(\frac{\text{Off-load voltage}}{\text{Full-load voltage}} - 1\right)$$

Very small transformers may have regulation as bad as 20%. In other words, their turns ratio has had to be reduced by 20% to ensure that they deliver the rated secondary voltage at full load current. Thus, our example 240 V:6 V transformer might actually have a turns ratio of 32:1, so when we try to step 6 V back up to 240 V, we only get 192 V. Worse, that 192 V is the off-load voltage and falls by 20% at full load current, so instead of seeing 192 V, we would only see 154 V at full load current.

The second problem is to do with magnetising current. Although the ratio by which a transformer steps up (or steps down) a voltage is dictated by the turns ratio between primary and secondary, this implies that so long as you had the right ratio, you could use as many or as few turns as you liked. Imagine a transformer with a secondary not connected to anything. If the secondary isn't connected to anything, we can simply throw it away. We are now left with a choke across the mains having a reactance (due to the transformer's primary inductance):

$$X_L = 2\pi f L$$

We have a voltage across this reactance, so a current must flow that is inversely proportional to the reactance ($V/X_L = I$), and this is generally known as the magnetising current. Inductance is proportional to the square of turns, so the more turns we have on the primary, the higher the reactance, and the lower the magnetising current. The key issue is that the higher the magnetising current, the more power is wasted as heat in the wire and the core.

It might seem that the ideal would be to use as many turns as possible to maximise primary inductance and thus minimise magnetising current, but thin wire has a higher resistance and would cause higher resistive losses when a load current was drawn. Thus, the primary inductance of a transformer must be a careful balance of a number of factors to maximise efficiency.

Large transformers (anything bigger than a domestic washing machine) are carefully designed to be >99% efficient, but small transformers are designed to be cheap. The upshot is that a small transformer might be so inefficient and draw such a large magnetising current that connecting two transformers back-to-back could overload the first simply due to the magnetising current of the second.

The author used a Tektronix P6302 50 MHz current probe and associated AM503 amplifier connected to a TDS3032 oscilloscope to measure transformer magnetising current from 240 V_{RMS} mains. A 30 VA toroid drew a negligible magnetising current of 1.5 mA_{RMS} (albeit very distorted), whereas a 20 VA EI drew a magnetising current of 24 mA_{RMS} – equivalent to 5.8 VA. Thus, although the toroid would be eminently suitable for back-to-back connection, a pair of the EI transformers would reduce the first transformer's VA rating to 14 VA. The situation became farcical with a very small 6 V 250 mA (1.5 VA) EI transformer that drew a very distorted magnetising current of 19.6 mA_{RMS} – equivalent to 4.7 VA, so back-to-back connection of a pair of these transformers would grossly overload the first (see Figure 5.15).

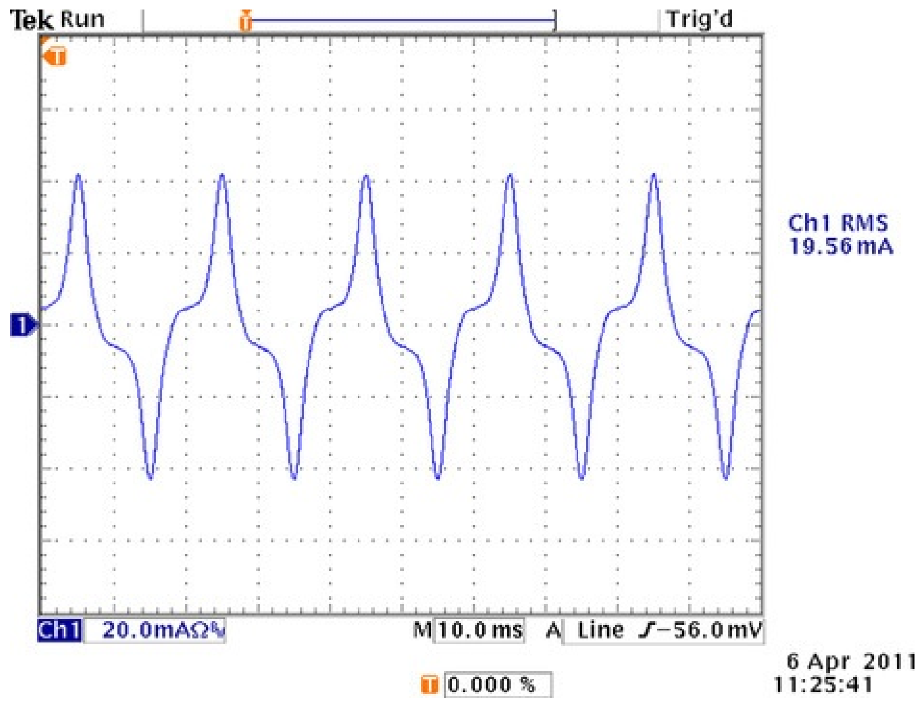


Figure 5.15 Highly distorted magnetising current drawn by 1.5-VA transformer.

Summarising, connecting transformers back-to-back might appear to be convenient, but it relies on the second transformer having a low magnetising current (suggesting a toroid) to avoid overloading the first transformer, and the output voltage will always be significantly lower than mains voltage.

If you only need a small current, better quality bathroom strip lights include a ≈20 VA split bobbin isolating transformer for their shaver socket, and this is a much better solution. When measured, the turns ratio of such a 20 VA transformer stepped up by 1:1.09 in order to overcome its full load losses, and the ratio of the winding resistances was identical, implying that the same gauge of wire was used for primary as for secondary. This implies that if you wanted a voltage lower than 240 V, this split bobbin transformer could safely have its primary and secondary interchanged so that instead of the turns ratio compensating for losses, it would step down to give ≈200 V at full load current.

Voltage Multipliers

Sometimes the mains transformer we are forced to use can't produce a high enough voltage, and one of the cheapest and most popular ways of generating high voltages at relatively low currents is the classic diode/capacitor voltage multiplier. Multipliers are most common where a constant and negligible current is required, such as the polarising bias required by electrostatic loudspeakers (≈5 kV).

The voltage multiplier was invented [7] in 1920 by the Swiss physicist H. Greinacher for polarising ionisation chambers and a Swiss patent [8] granted in 1922, but nobody seemed to take much notice. However, 10 years later in 1932, J.D. Cockcroft and E.T.S. Walton published a paper [9] describing the use of a 125 kV voltage multiplier identical to Greinacher's for accelerating hydrogen ions to split the lithium atom's nucleus. Perhaps the glamour of splitting the atom meant that the patent examiners were not as diligent as they should have been, for Cockcroft and Walton were granted UK [10] and US [11] patents for the multiplier. Certainly, splitting the atom later won them the 1951 Nobel physics prize, and the attendant voltage multiplier has become incorrectly associated with them.

The multiplier, often known as a ladder or cascode, can be extended indefinitely, with each step theoretically adding $\sqrt{2}\ V_{in(RMS)}$ to the output, but its regulation is very poor. Each diode needs a voltage rating $>\sqrt{2}\ V_{in(RMS)}$. Unfortunately, all but the lowest capacitor must be rated at $>2\sqrt{2}\ V_{in(RMS)}$. Additionally, because succeeding capacitors are charged by rectifier switching that partly discharges the lowest capacitor, this capacitor must be of a higher value to reduce voltage drop (see Figure 5.16).

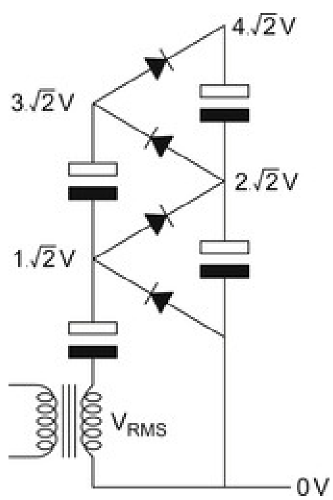


Figure 5.16 Greinacher voltage multiplier.

Although voltage multipliers were initially used for generating EHT, they can be useful for providing negative grid bias, and the diminutive Rogers Cadet stereo power amplifier even used a voltage doubler for its main HT. There are two forms (see Figure 5.17).

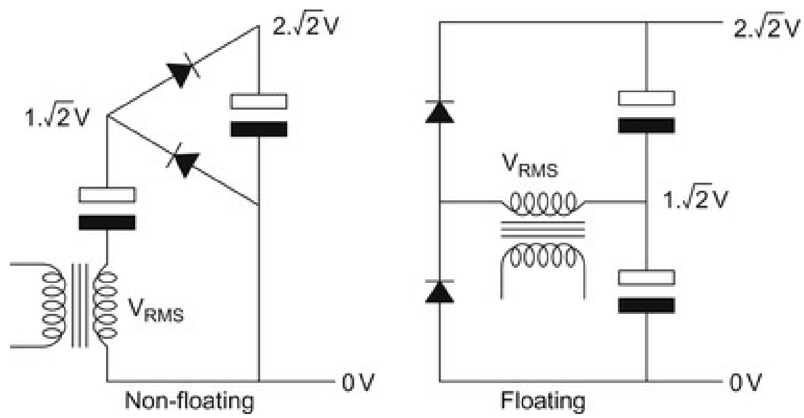


Figure 5.17 Non-floating and floating voltage doublers.

The non-floating doubler is a truncated Greinacher multiplier. It can be connected in parallel with a conventional centre-tapped rectifier/transformer combination allowing a subsidiary (higher voltage) HT to be developed – perhaps for polarising a dedicated high-frequency electrostatic loudspeaker.

The advantage of the floating doubler is that it uses two *identical* capacitors, each rated at half the output voltage, but the diodes must be rated at >2√2 $V_{in(RMS)}$. Because each capacitor is only charged on alternate half-cycles, ripple voltage is doubled compared to a conventional full-wave rectifier. Because the ripple voltages of the two capacitors are in series, a further doubling of ripple voltage occurs. Thus, for a given ripple voltage, the floating doubler requires each capacitor to have four times the capacitance needed by a conventional full-wave rectifier.

The Choke Input Power Supply

Choke input power supplies were very popular in the heyday of valve amplifiers because even if large value capacitors had been available, the ripple current they would have drawn would have destroyed the rectifiers of the time, so chokes had to be used for smoothing (see Figure 5.18).

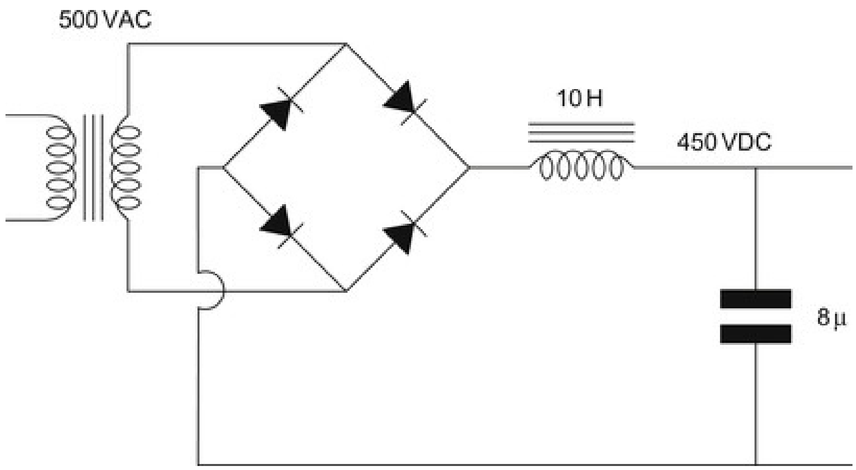


Figure 5.18 Choke input power supply.

If we could make a choke input supply having a choke of infinite inductance, the mains transformer current would be identical to the DC load current. Practical supplies do not quite achieve this ideal, so the transformer current is a combination of DC load current and a somewhat smaller, nearly sinusoidal current drawn by the choke. Nevertheless, the choke input power supply has the great advantage that it draws a very nearly continuous current from the mains transformer rather than a series of high current pulses. To understand why this is so, we need to consider the output waveform of the rectifier in detail (see Figure 5.19).

Figure 5.19 Full-wave rectified AC sine wave.

This waveform is a full-wave rectified sine wave, but because it has undergone a non-linear process (rectification), the frequencies present in this waveform are not the same as went into the rectifier. Fourier analysis reveals that the result of full-wave rectification of a sine wave is:

$$v = \frac{2\sqrt{2}\, V_{\mathrm{in(RMS)}}}{\pi}\left(1 + \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\cos n\omega t}{4n^2 - 1}\right)$$

Note that $V_{\mathrm{in(RMS)}}$ is the voltage *before* rectification.

The previous equation is a mathematical way of expressing an infinite series, but for our purposes it is simpler to present the information as follows:

$$V = V_{\mathrm{in(RMS)}}[0.90 + 0.6(2f) - 0.12(4f) + 0.05(6f) - 0.03(8f)\ldots]$$

This shows us that a full-wave rectified sine wave is made up of a DC component corresponding to 0.90 $V_{in(RMS)}$, plus a series of decaying even harmonics of the input frequency (f) *before* rectification. The choke has such a high reactance to these AC terms that only the DC component reaches the load. The output voltage of a choke input power supply is therefore 0.90 $V_{in(RMS)}$, rather than $\sqrt{2}$ $V_{in(RMS)}$ for the capacitor input supply.

Minimum Load Current for a Choke Input Supply

Choke input power supplies require a minimum load current to be drawn before they operate correctly. If less than this current is drawn, the circuit reverts to pulse charging of the capacitor, and the output voltage rises to a maximum of $\sqrt{2}$ $V_{in(RMS)}$. The absolute minimum current that should be drawn is:

$$I_{min} = \frac{\sqrt{2}\, V_{in(RMS)}}{3\pi^2 fL}$$

In practice, the inductance of any power supply choke varies with the current through it, so it is wise to draw rather more current than this, and a handy approximation (appropriate for 50 Hz or 60 Hz mains) is:

$$I_{min(mA)} \geq \frac{V_{in(RMS)}}{L_{(H)}}$$

Choke input supplies invariably feed a capacitor, and the minimum current requirement is therefore important, since insufficient current could cause the voltage across the capacitor to rise to ≈157% of nominal voltage, which might destroy it. The traditional way of dealing with this problem was to use a *swinging* choke, whose small air gap caused high inductance at low currents that fell as current increased, and although these became unfashionable after the 1960s, they have returned – just like flares and platform shoes.

Once the minimum current has been exceeded, the output ripple becomes constant with load current and because for any practical power supply filter, $X_L >> X_C$, the simple potential divider's attenuation can be approximated to a ratio:

$$\frac{V_{out}}{V_{in}} \geq \frac{X_C}{X_L} \geq \frac{1}{\omega^2 LC}$$

where $\omega = 2\pi f$.

If we consider that only the amplitude of the second harmonic is significant, we can incorporate its 0.6 factor from the Fourier series we saw earlier and determine the peak ripple voltage:

$$V_{\text{ripple(pk)}} \geq \frac{1.6\ V_{\text{in(RMS)}}}{\omega^2 LC}$$

It is more usual to work with peak-to-peak ripple voltages (because they're much easier to measure on an oscilloscope), and the waveform is sinusoidal so $V_{\text{pk–pk}}=2\ V_{\text{pk}}$:

$$V_{\text{ripple(pk-pk)}} \geq \frac{1.2\ V_{\text{in(RMS)}}}{\omega^2 LC}$$

If we now change the equation to accept capacitance directly in μF and include 100 Hz or 120 Hz ripple frequencies:

$$V_{\text{ripple(pk-pk)}} \geq \frac{3\ V_{\text{in(RMS)}}}{L_{\text{(H)}} C_{(\mu F)}} (50\ \text{Hz}),\quad \geq \frac{2\ V_{\text{in(RMS)}}}{L_{\text{(H)}} C_{(\mu F)}} (60\ \text{Hz})$$

where L is in H, C in μF and $V_{\text{in(RMS)}}$ is the mains transformer secondary voltage. Strictly, the factor of 2 in the 60 Hz approximation should be 2.1, but it is not unusual for an iron-cored choke's inductance to vary by 50% or more depending on DC and AC current, and the tolerance of an electrolytic capacitor is likely to be ±20%, so the 5% error that allows an easily remembered approximation is insignificant. More importantly, the approximations agree well with measurement and PSUD2 predictions.

Current Rating of the Choke

Although a choke of infinite inductance would allow both choke and mains transformer to have a current rating equal to the maximum DC load current, they actually have to support a somewhat higher current, and it is particularly important that the choke is correctly rated. Remember that the choke generates magnetic flux in its core proportional to the current passing through the coil, but if too much magnetising force is applied, the core saturates, causing inductance to fall to zero.

Since the output of the rectifier comprises a DC component and an AC component, it is the summation of these components that determines the current rating of the choke. The DC component is simply the load current, but the AC component requires a little more thought.

Because the choke is followed by a capacitor, which is a short-circuit to AC, the entire AC component leaving the rectifier is developed across the reactance of the choke, causing an AC current to flow. Once we know the AC voltage across the choke, we can easily calculate the current.

As previously mentioned, the AC component is dominated by the second harmonic, so we can simplify the calculation to deal exclusively with this

component.
The instantaneous AC voltage across the choke is therefore:

$$v = 0.6\ V_{\text{in(RMS)}}\cos(2\pi ft)$$

where *f* is the second harmonic of mains frequency.
The reactance of the choke is:

$$X_{\text{L}} = 2\pi fL$$

Using Ohm's law to combine the two equations, the instantaneous current through the choke is:

$$I_{\text{AC}} = \frac{0.6\ V_{\text{in(RMS)}}\cos(2\pi ft)}{2\pi fL}$$

We are only concerned with the maximum current, which occurs when cos(2*πft*)=1, so this factor can be removed, leaving:

$$I_{\text{AC(peak)}} = \frac{0.6\ V_{\text{in(RMS)}}}{2\pi fL}\ (\text{second harmonic only})$$

It was stated that only the second harmonic was significant, but this assumption should now be examined. Referring to the Fourier series, the fourth harmonic is 20% (0.12/0.6) of the voltage of the second harmonic. The doubled reactance of the choke at the fourth harmonic halves the choke current, resulting in a fourth harmonic current that is only 10% of the second harmonic, so the approximation is fair, but there is room for improvement.
The sum of the AC currents drawn by each of the Fourier terms, up to and including the eighth harmonic, was investigated to find the maximum positive peak. (The negative peak is irrelevant since when added to the DC load current, it reduces the total peak current.) The result of this exercise modified the equation to:

$$I_{\text{AC(positive peak)}} = \frac{0.544\ V_{\text{in(RMS)}}}{2\pi fL} = \frac{V_{\text{in(RMS)}}}{1{,}155L}\ (50\ \text{Hz}) = \frac{V_{\text{in(RMS)}}}{1{,}386L}\ (60\ \text{Hz})$$

But the total peak current flowing through the choke is the sum of the AC peak current and the DC load current:

$$I_{\text{pk}} = I_{\text{DC}} + I_{\text{AC(positive peak)}}$$

As an example, a Class A power amplifier using a pair of push–pull 845 valves requires a raw HT of 1,100 V at 218 mA, and a 10 H 350 mA choke is available, but is this adequate? The transformer supplying the choke input filter has an output voltage of 1,224 V_{RMS}. Using the previously derived equation and assuming 50 Hz mains:

$$I_{\mathrm{AC(positive\ \ peak)}} = \frac{V_{\mathrm{in(RMS)}}}{1{,}155L} = \frac{1{,}224}{1{,}155 \times 10} = 0.106\ \mathrm{A} = 106\ \mathrm{mA}$$

$$I_{\mathrm{pk}} = I_{\mathrm{DC}} + I_{\mathrm{AC(positive\ \ peak)}} = 218\ \mathrm{mA} + 106\ \mathrm{mA} = 324\ \mathrm{mA}$$

The total peak current is 324 mA, so the 350 mA rated choke is just adequate, but the example shows that choke AC current can be surprisingly high, particularly when high HT voltages are contemplated.

As a sweeping generalisation, chokes for HT choke input supplies generally need to be ≥15 H, otherwise the AC current becomes crippling, and the quickest way of determining a choke's suitability for a choke input supply is either to model it in PSUD2, or to put all the choke equations into a spreadsheet.

Mains Transformer Current Rating for a Choke Input Supply

The peak choke current must be supplied by the transformer, so the transformer should be rated appropriately. However, since transformer ratings assume resistive loads and sine waves, their current ratings are RMS of sine wave, and they can deliver a peak current of √2 this value, so the previous example would require a transformer with an RMS sine wave current rating of 229 mA (324 mA $_{\mathrm{pk}}$). This is sufficiently close (5% error) to the DC load current of 218 mA that a common approximation is to assume that the transformer should have an AC $_{\mathrm{RMS}}$ current rating equal to the DC load current.

Current Spikes and Snubbers

Choke input power supplies are not perfect and have two main problems, electrical switching spikes and mechanical vibration.

Although we said earlier that the choke input power supply drew a continuous current from the mains transformer, this cannot be exactly true. Since the rectifier diodes require a certain voltage across them before they switch on (irrespective of whether they are thermionic or semiconductor), there must be a time, as the input waveform crosses through zero volts, when neither diode is switched on. The current drawn from the transformer is therefore not quite continuous and must momentarily fall to zero. The choke will try to maintain current and, in doing so, will develop an EMF:

$$E = -L\frac{\mathrm{d}i}{\mathrm{d}t}$$

In any full-wave rectifier, the diodes switch off at twice mains frequency, and at that instant, d i/d t=∞, so theoretically infinite voltage spikes are produced with a repetition rate of twice mains frequency (see Figure 5.20).

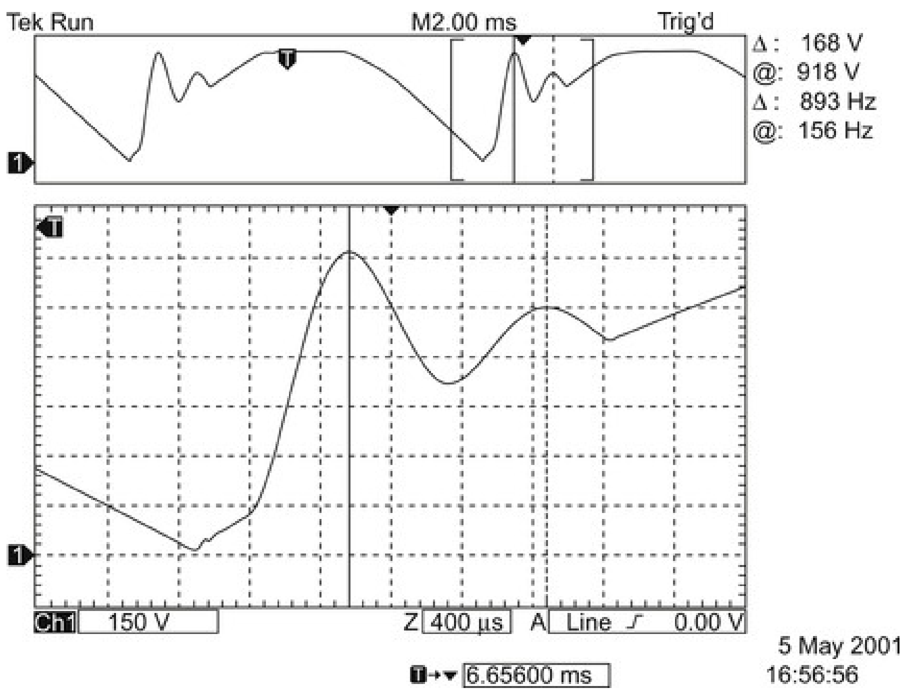


Figure 5.20 Extreme choke ringing caused by rectifier switching without load current.

Although drawing a significant load current greatly damps the ringing of the choke, the current waveform still has a glitch (see Figure 5.21).

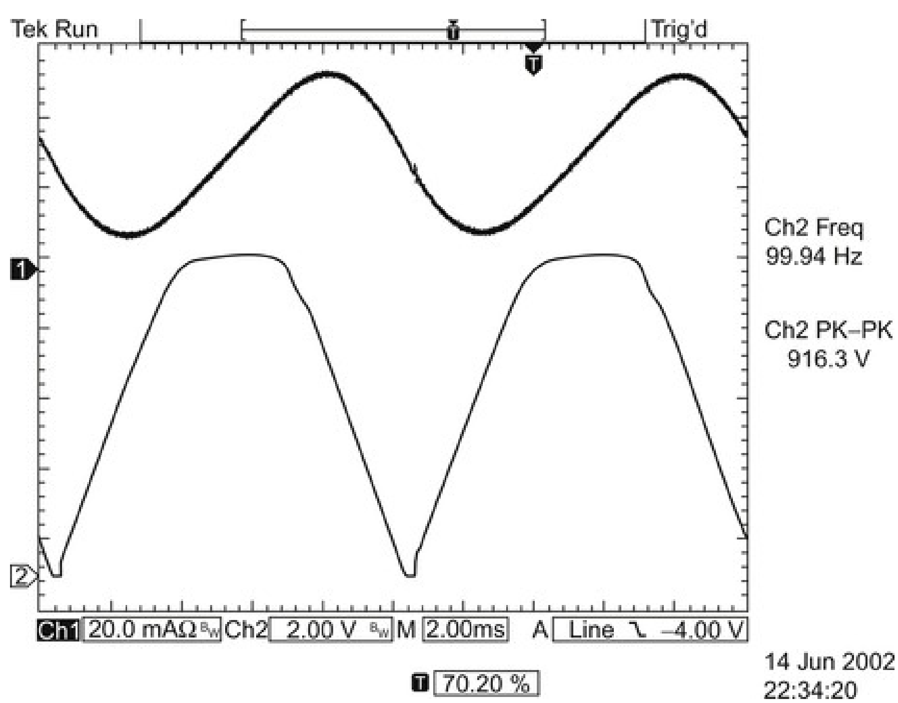


Figure 5.21 Without snubber (but with load current). Upper trace (Ch1): transformer load current. Lower trace (Ch2): input voltage to rectifier.

Traditionally, a resistor/capacitor snubber network was connected across the choke to protect the inter-winding insulation of the mains transformer from the spikes (see Figure 5.22 a).

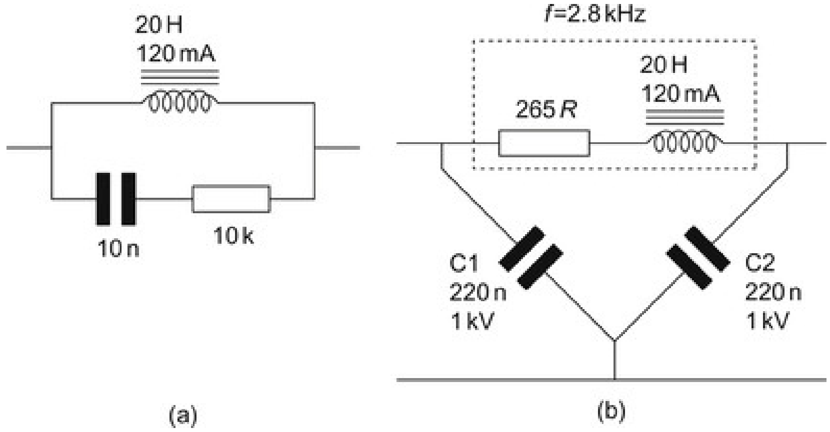


Figure 5.22 Traditional and improved choke snubber networks.

Although fitting the traditional 10 nF+10 kΩ snubber across the choke tames the voltage spikes, it degrades high-frequency filtering and worsens the glitch in the current waveform (see Figure 5.23).

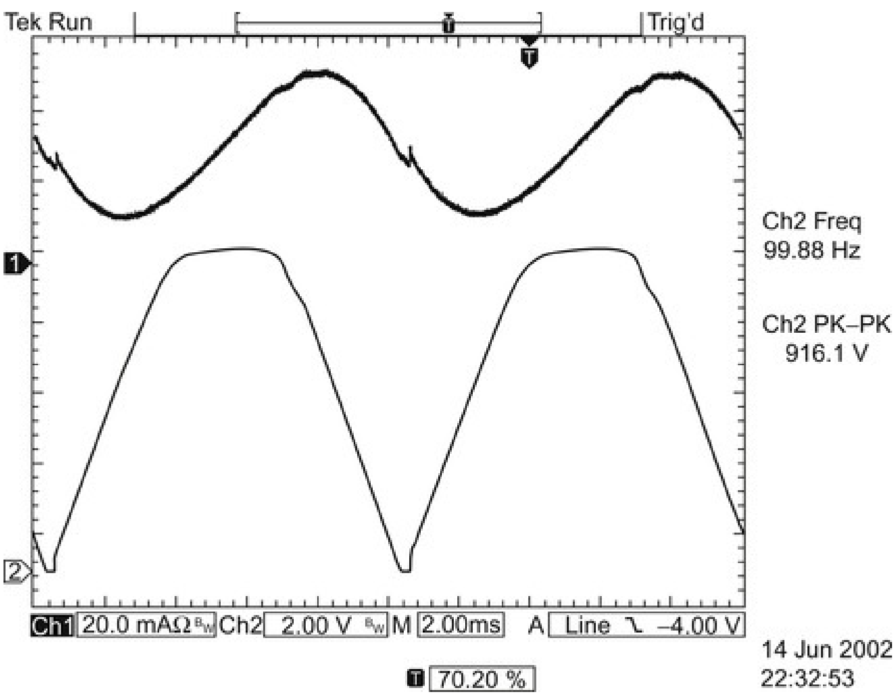


Figure 5.23 With 10 nF+10 kΩ snubber. Upper trace (Ch1): transformer load current. Lower trace (Ch2): input voltage to rectifier. Note the worsened current waveform.

A snubbing method that significantly improves high-frequency filtering is to fit

back-to-back capacitors across the choke, with their centre tap connected to 0 V, and use the internal resistance of the choke as the snubbing resistance. Optimum high frequency filtering is obtained by choosing C1 so that it resonates with the leakage inductance of the mains transformer at the same frequency as the self-resonance of the choke, but this seems not to be critical, and curiously 220 nF is often a practical value for both HT and LT supplies (see Figure 5.22 b).
The modified snubber network removes the voltage spikes without compromising high frequency filtering or adding glitches to the current waveform (see Figure 5.24).

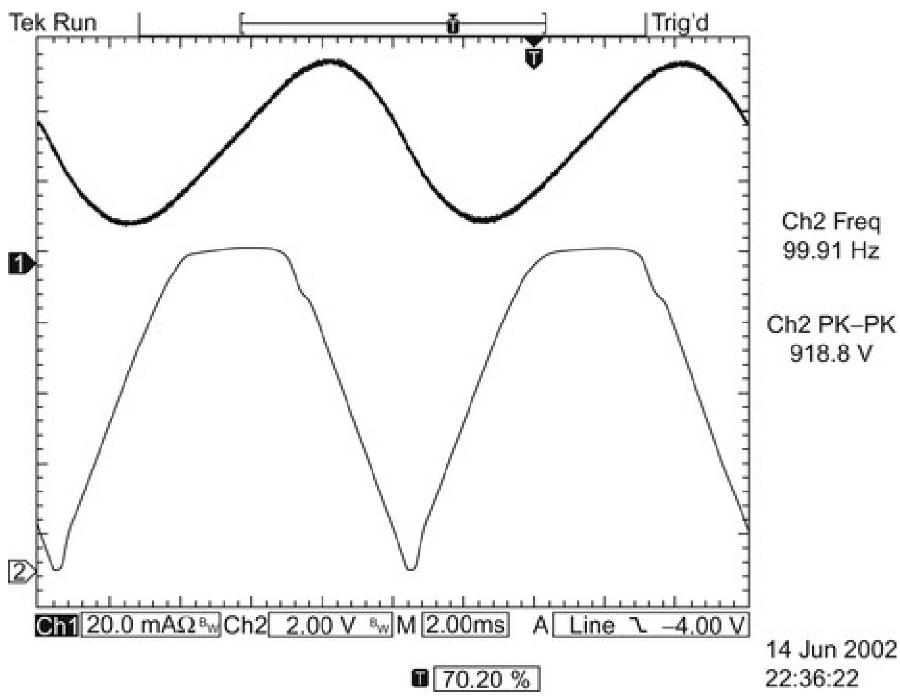


Figure 5.24 With 220 nF back-to-back snubber. Upper trace (Ch1): transformer load current. Lower trace (Ch2): input voltage to rectifier. Note the complete absence of glitches.

As mentioned previously, the entire AC component at the output of the rectifier is across the choke. In Chapter 4 , we observed that output transformers could 'sing' due to loose laminations or magnetostriction, and the same is true here. The choke could buzz at twice mains frequency, and if it has any loose parts, such as a loose screening can, it will rattle loudly. Even worse, the choke is bolted to a resonant sounding board (the chassis), which will amplify the buzz.
The author has recently investigated a number of choke input power supplies. A buzzing choke implies core saturation. Unfortunately, it seems that iron cores can deteriorate over the decades, reducing inductance, which increases the AC current, perhaps to the point of saturation, thus causing buzz. If you must use old chokes (and transformers), check them for buzz under load *before* drilling holes

in the chassis.

Intermediate Mode: The Region Between Choke Input and Capacitor Input

For a given input voltage, a choke input supply produces the lowest output voltage (0.90 $V_{in(RMS)}$) because only the DC component from the rectifier reaches the load, whereas a capacitor input supply with $C_{reservoir}=\infty$ achieves the maximum ($\sqrt{2}$ $V_{in(RMS)}$) because it can use the AC component. Another way of looking at a choke input supply is to consider it to be a capacitor input supply where $C_{reservoir}=0$. We now see that changing reservoir capacitor value could be a useful way of adjusting output voltage between 0.90 $V_{in(RMS)}$ and $\sqrt{2}$ $V_{in(RMS)}$, thus allowing a previously unsuitable transformer secondary to provide the required output voltage without wasting power in a resistor (see Figure 5.25).

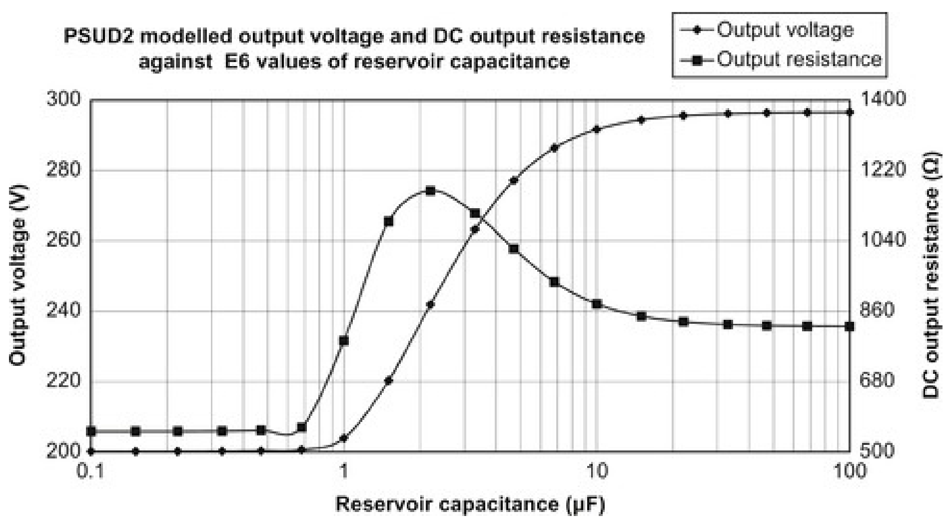


Figure 5.25 Intermediate mode: the effect of reservoir capacitance on output voltage and DC output resistance.

This modelled example shows that a reservoir capacitor of between 1 μF and 10 μF is needed to operate in intermediate mode, and this is fairly typical. Noting the critical dependence of output voltage on reservoir value, a plastic capacitor (rather than an electrolytic) is necessary to ensure that the practical capacitor matches the modelled value. Fortunately, the high peak voltage appearing across the reservoir capacitor tends to force a polypropylene type, guaranteeing a sufficiently close tolerance.

DC output resistance of a supply is important because it shows how output voltage changes with load current (AC output impedance will be investigated shortly). The graph shows the lowest DC output resistance when the supply operates in choke input mode ($C_{reservoir}$=0), and a somewhat higher DC output resistance in capacitor input mode ($C_{reservoir}$=∞), but note that the cost of intermediate mode is significantly higher DC output resistance than either of these. The implication of the intermediate mode's high DC output resistance is that it is best suited to constant current loads such as regulators or pre-amplifiers where the signal current is so small that it cannot modulate the supply. Although intermediate mode has a high DC output resistance, it is efficient because adding a real series resistance to the output of a capacitor input supply to produce the same output voltage would waste power.

Note that because intermediate mode uses an undersized reservoir capacitor, that capacitor has a substantial ripple voltage across it (which isn't a problem), but it means that the following choke must also have a substantial ripple voltage across it, implying significant AC current in addition to the DC load current. Thus, just like the choke input power supply, we must ensure that our choke has an adequate current rating to cope with the sum of these currents without saturating.

It would be nice if intermediate mode had one equation for predicting the required reservoir capacitor value for a particular HT voltage and another for the peak choke current, but we contravene the assumptions made earlier for capacitor input smoothing and do not consider the DC component leaving the rectifier. The only way of determining the required value of the reservoir capacitor is to determine it experimentally – start with a 5 μF reservoir and adjust it up or down until the required HT voltage is achieved at the expected load current. Having found the capacitor value, the choke peak current can be determined. Even the author does not have a decade capacitance box having an adequate voltage rating (and 1–10 μF range), let alone a large selection of chokes, but the experiment is easily done as a computer simulation. PSUD2 is ideal for the task, but beware that intermediate mode may take considerable time for its output voltage to reach its final value, so a long simulation time is essential (50 s is generally adequate).

PSUD2

PSUD2 by Duncan Munro is a power supply simulation freeware that has become the de facto standard for simulating the traditional linear power supplies used in valve amplifiers. The software can simulate all combinations of LCR supplies and gives a graphical display of chosen voltages or currents. PSUD2 is easily used, so rather than explain its full operation, the author will simply touch on two important points.

Most commercial mains transformers operate just above the knee of the *B*/ *H* curve, and this means that ripple current is limited by core saturation to four to six times DC load current. Simulations like PSUD2 necessarily use a simple transformer model that can't take account of saturation, so it typically predicts a slightly higher ripple current (perhaps seven times DC load current). When the real-world transformer saturates and limits ripple current, the effect is to very slightly lower the DC load voltage compared to the PSUD2 prediction.

PSUD2 has a feature that seems trivial, yet allows unprecedented understanding of the supply being simulated. PSUD2 can check the low frequency stability of a supply. Checking low frequency stability is important because, as we will see in a moment, it is easy to accidentally design a supply that rings like a subsonic bell. Low frequency stability is checked using the 'stepped load' command, which is invoked using the 'edit' command on the load to select a constant current load (see Figure 5.26).

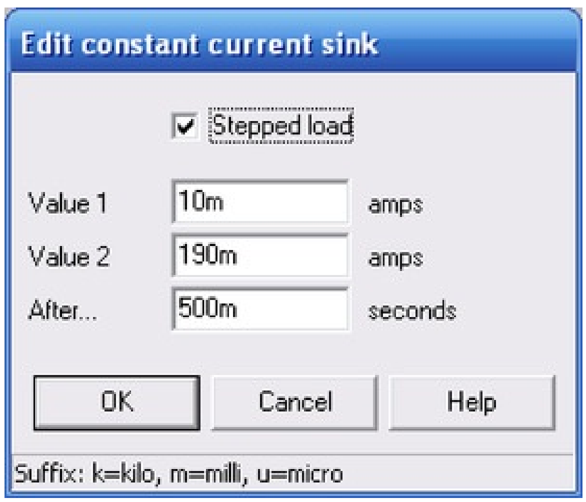


Figure 5.26 Invoking the stepped current feature in PSUD2.

A step in current of 10:1 will certainly provoke any resonances, and these will be seen as capacitor voltage ringing. Note that the time when the step occurs can be set, so it should be some time after the supply has completed its switch-on transient. Thus, a multi-section filter with current taps at each section should set stepped loads at each current tap but with different timings (1 s intervals are useful), and then monitor each capacitor voltage. Ringing at one section often leaks into other sections, so when an unstable section's load steps, it will cause particularly bad ringing, revealing the problematic section. If we over-damp the filters, the power supply will be slow to respond to changes in load current, so we ideally want a little less than critical damping but no oscillation. Thus, each of the programmed current steps should cause capacitor voltages to change with a fast exponential curve but without ringing (see Figure 5.27).

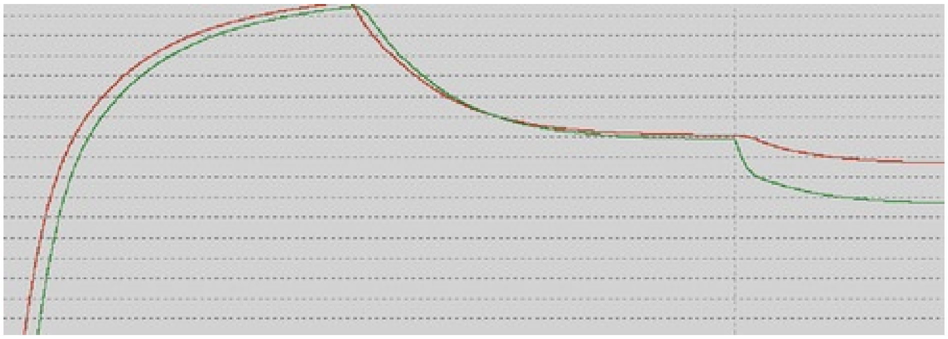


Figure 5.27 A well-designed supply responds to current step changes without ringing.

However, should one of the *LC* sections have too high a *Q* (caused in this instance by reducing the series resistance of its choke from 400 Ω to 40 Ω), ringing occurs which is lightly coupled into the voltage of the other tap (see Figure 5.28).

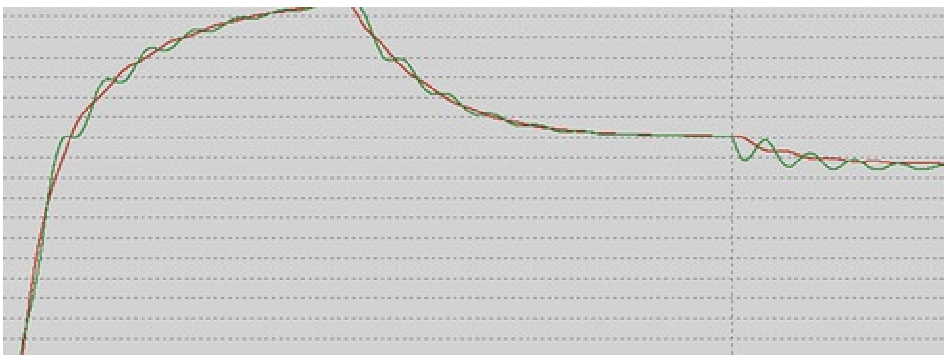


Figure 5.28 An unstable section causes low frequency ringing.

Although a 220 nF capacitor is frequently adequate for snubbing high frequency ringing of the choke in a choke input supply, it may not be large enough to damp the Low Frequency resonance formed by the combination of the choke and the following reservoir capacitor, and this is best checked using PSUD2. The exact value can be quite critical; when set to step from 20 mA to 80 mA after 0.9 s and viewed for 1 s after a reporting delay of 1 s, changing the snubber from 500 nF to 600 nF completely eliminated the ringing in one HT supply. Although larger values ensure freedom from Low Frequency ringing, they also slug the response of the supply to a current step, so it’s worth using a snubber only a little larger than the theoretical required value. In practice, the theoretical value is unlikely to be a preferred value, so the next preferred value up will do nicely.

We could iteratively optimise a design by randomly changing all values in PSUD2 until we obtain the required result, but this is likely to be very slow. The earlier analysis of rectification and *LC* filters greatly speeds optimisation because it allows us to make initial estimates of all required values and put them into PSUD2. Further, the analysis also gave us an understanding of which peak currents and voltages should be checked in PSUD2 to avoid overloading components and how to avoid low-frequency instability.

Broadband Response of Practical LC Filters

So far, our investigation of rectification and filtering has focussed on the behaviour at ripple frequency and its harmonics, but we now need to broaden our outlook to include behaviour from DC to low radio frequencies. To attenuate low (≈100 Hz) frequencies significantly, an *LC* filter with a large inductance is required, which inevitably has internal shunt capacitance. Conversely, the filter capacitor has series inductance and resistance (ESR), and these hidden components mean that any practical *LC* filter has a frequency response that may be divided into four main regions. (Although surprisingly smooth, the following graph is a result of practical measurements of an *LC* filter. See Figure 5.29 .)

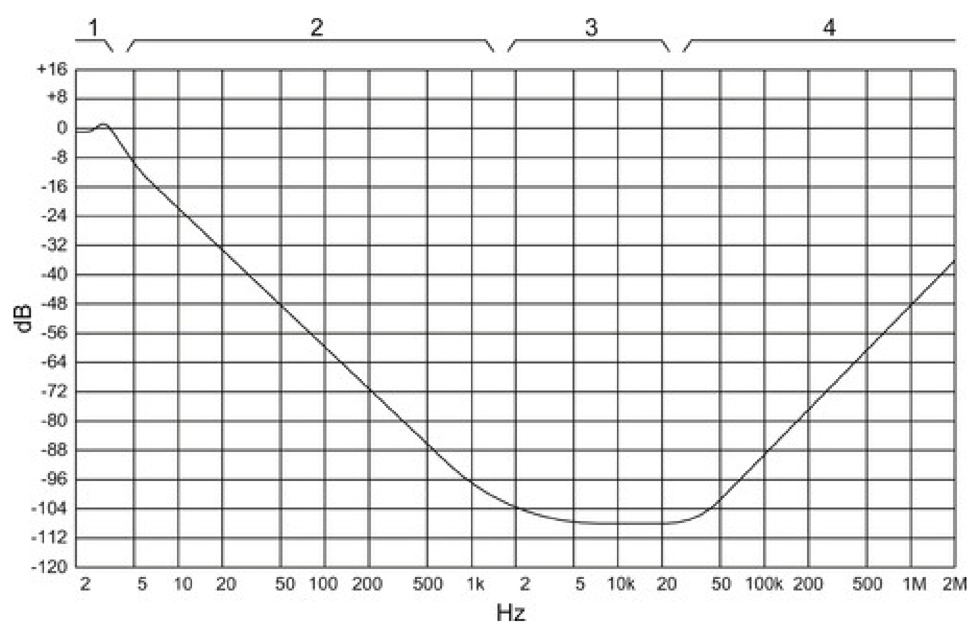


Figure 5.29 Measured frequency response of *LC* filter (20 H 50 mA, 120 μF 400 V polypropylene).

Region 1

This is the only region we can directly control, so it is well worthy of investigation. Apart from losses due to DC resistance, the low-pass filter does not attenuate frequencies below the Low Frequency resonance:

$$f_{\text{res(Low Frequency)}} = \frac{1}{2\pi\sqrt{LC}}$$

We aim to position the (hopefully) subsonic resonance as low as possible by making L and C large because every octave by which we can lower $f_{\text{res(Low Frequency)}}$ produces an additional 12 dB of filtering. If $f_{\text{res(Low Frequency)}}$has Q>0.707, an Low Frequency peak results in the response of the filter, so it is useful to check Q:

$$Q = \frac{1}{R_{\text{DC}}} \cdot \sqrt{\frac{L}{C}}$$

where

L=inductance of choke

R_{DC}=DC resistance of choke

C=capacitance of filter capacitor.

Ideally, the resonance should be critically damped (Q=0.5), which can be achieved by adding external series resistance to the choke. Strictly, the load resistance across the capacitor also damps the resonance, and this may be transformed into a notional extra choke series resistance using:

$$R_{\text{notional}} = \frac{L}{CR_{\text{load}}}$$

However, the damping effect of the load resistance is usually negligible. Even worse, a series voltage regulator is a constant current, or infinite, AC load to the smoothing circuitry, so it adds no damping whatsoever.

A typical traditional example: A choke input supply might use a 15 H choke having 260 Ω internal resistance coupled to an 8 μF capacitor, resulting in $f_{\text{res(Low Frequency)}}$=15 Hz and Q=5.3. This Q is too high, and $f_{\text{res(Low Frequency)}}$ is too near the audio band, but the additional 2.5 kΩ series resistance required to achieve critical damping would waste HT voltage and greatly increase power supply output resistance. A better alternative would replace the 8 μF capacitor with a 120 μF polypropylene, since this would give $f_{\text{res(Low Frequency)}}$=3.75 Hz and Q=1.36, and this Q might be acceptable. Note that reducing Q by increasing C is not as effective as increasing R because C is inside the square root term of the Q equation, so when simulating in PSUD2, expect the need to increase filter capacitance significantly if you can’t tolerate increased choke resistance.

Region 2

The reactance of the choke doubles for each octave rise in frequency, whilst the reactance of the reservoir capacitor halves, producing the familiar 12 dB/octave slope. PSUD2 works in the time rather than frequency domain, so we don't see this directly, but the reported values of ripple tell us how well we are filtering at our chosen ripple frequency.

Note that PSUD2 has no understanding of the next two concepts, so both concepts must be checked manually.

Region 3

The choke's internal shunt capacitance begins to take effect. Once the reactance of the shunt capacitance is equal to the inductive reactance of the choke, the choke resonates, so this region may be defined as beginning at $f_{\text{res(high frequency)}}$. Above this self-resonant frequency (3–15 kHz for a typical HT choke), the choke's shunt capacitance forms a potential divider with the filter capacitance whose loss is constant with frequency:

$$\text{Loss (dB)} \approx 20 \log\left(\frac{C_{\text{smoothing}}}{C_{\text{choke}}}\right)$$

Region 4

The series inductance of the filter capacitor becomes significant, and this forms a hidden high-pass filter in conjunction with the shunt capacitance of the choke, so the output noise of the practical filter rises at 12 dB/octave.

All of the previous concepts can be simplified by considering an idealised *LC* filter response to be made up of three straight lines having freedom to move either vertically or horizontally (see Figure 5.30).

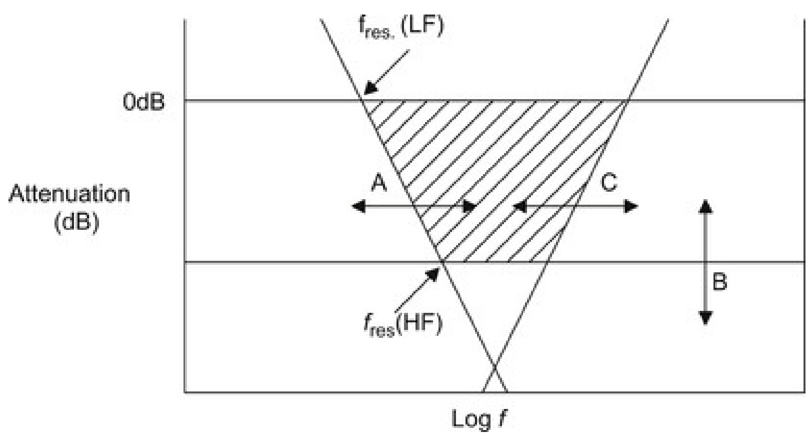


Figure 5.30 Conceptual model of universal *LC* filter.

- Line A falls at 12 dB/octave, and slides horizontally to the left as the product of choke inductance and filter capacitance increases.

• Line B falls vertically as the ratio of filter capacitance and choke shunt capacitance increases, intercepting line A at the choke's self-resonant frequency. Capacitance between adjacent winding layers of the choke could be reduced by vertical sectioning or by interposing earthed electrostatic screens.

• Line C rises at 12 dB/octave, and slides horizontally to the right as filter capacitor series inductance falls, intercepting line B at the capacitor's self-resonant frequency. Most modern capacitors have series inductance between 10 nH and 100 nH, so if this value is known, their self-resonant frequency can be calculated using the standard resonance equation. Conversely, if their self-resonant frequency is known, series inductance can be calculated. Sadly, the two preceding statements are not true for an electrolytic capacitor – the capacitance of an electrolytic capacitor at self-resonance is likely to be between 25% and 50% lower than its DC value.

Unfortunately, filter capacitor ESR also complicates the issue and can easily mask the plateau of line B. The intersection of lines B and C can be viewed as the response of an *LC* filter, but plotted upside down. Thus, to avoid degrading the plateau, the *Q* of the capacitor's self-resonant frequency should be >1/√2 (unusually, we *want* the filter's response to be underdamped at its resonant frequency). Remembering:

$$Q = \frac{1}{R}\sqrt{\frac{L}{C}}$$

Rearranging and inserting our *Q* requirement:

$$\mathrm{ESR}_{\mathrm{required}} < \sqrt{\frac{2L_{\mathrm{series}}}{C}}$$

Thus, a 100 μF filter capacitor having a typical series inductance of 20 nH would require ESR<18 mΩ.
The modelled filter compares the effect of the measured parameters of three different filter capacitors in combination with a Woden 51700 choke (20 H, 365 Ω and 95 pF) (see Figure 5.31 and Table 5.5).

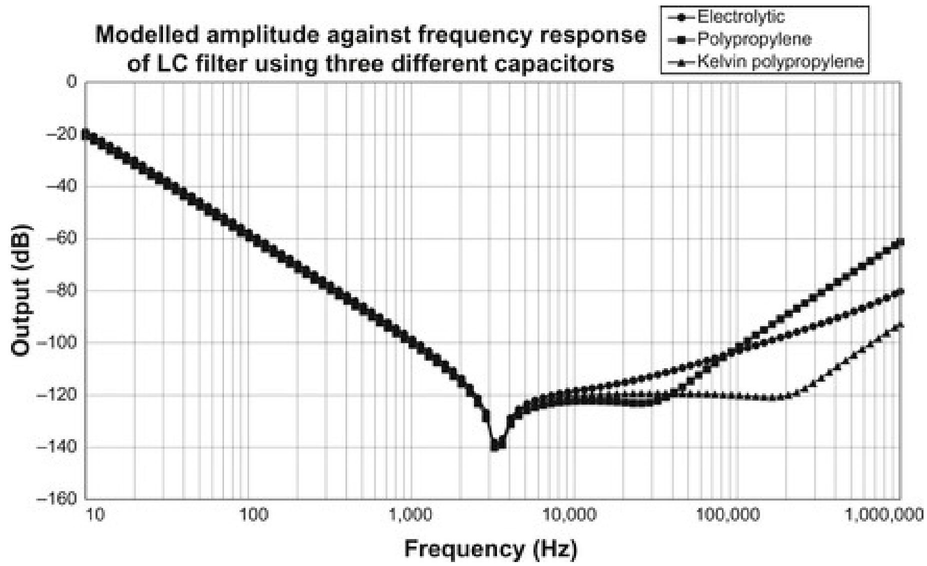


Figure 5.31 Modelled comparison of amplitude against frequency response of *LC* filter using three different capacitors.

Table 5.5 Capacitor Values and Parasitic Components of the Modelled Filter of Figure 5.31

	C (μF)	ESR (mΩ)	L_{series} (nH)
(A) 100-μF 450-V 159 series BC Components PCB mounting snap-in aluminium electrolytic capacitor	95	300	16
(B) 120-μF 400-V Ansar metallised polypropylene	120	35	200
(C) 100-μF 400-V Ansar metallised polypropylene with Kelvin connection	100	7	≈5.5

As expected, the electrolytic capacitor fails to achieve the filtering plateau due to its significant ESR. Capacitor B had the potential to offer excellent broadband filtering, but its ESR and series inductance were badly compromised by an ill-considered mechanical requirement to have both leads (known as tails) exit from the same end, enforcing an additional 120 mm length on one tail.

Capacitor C with its Kelvin connection was from the second batch specially made for the author by suppression devices of clitheroe but will doubtless become a standard part. Note that not only does capacitor C easily achieve the filtering plateau, but also that its minimal series inductance maintains that plateau to the highest frequency. Each end of the capacitor has two independent tails from the tin/zinc layer connecting to the plates: one tail is used for the source of current and the other for the load (see Figure 5.32).

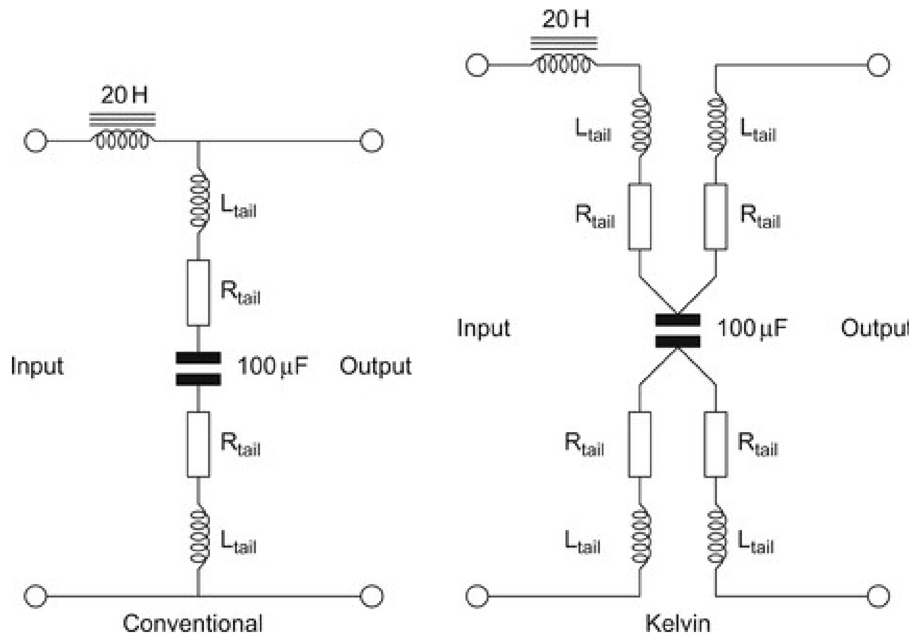


Figure 5.32 Kelvin connection moves lead inductance to a position in the circuit where it no longer matters.

The significance of the Kelvin connection is that from the point of view of filtering incoming interference, tail inductance (typically 0.75 nH/mm) and resistance are no longer relevant, and filtering is determined purely by the residual inductance and resistance of the foils. This connection is also useful when the capacitor is used as a reservoir capacitor because it enables ideal separation of ripple current from load current. Most importantly, if a plastic capacitor has to be used (usually because of the required voltage rating) Kelvin connection offers a significant reduction of mains-borne interference between 10 kHz and 1 MHz at insignificant additional cost.

The modelled Kelvin connection shows a significant improvement, but does it match reality? (see Figure 5.33).

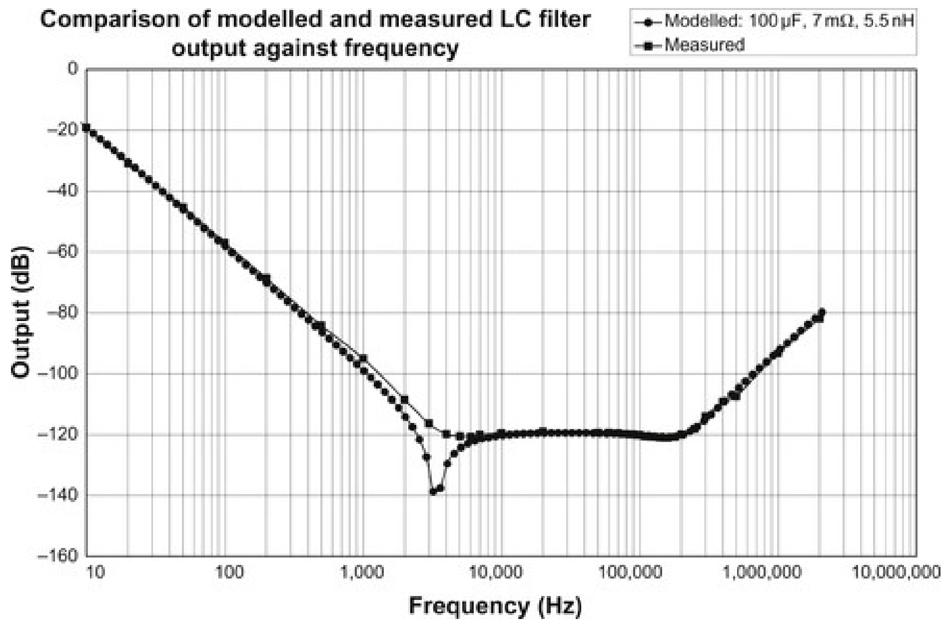


Figure 5.33 Comparison of modelled and measured *LC* filter using capacitor having Kelvin connection.

As can be seen, the match isn't quite perfect, with meter noise causing a shallower null at the choke's self-resonant frequency than modelled, but the general correspondence is excellent, and matching the model to the measurement was the only practical way of determining the Kelvin capacitor's series inductance and ESR. More importantly, the hypothesised improved high-frequency filtering of the Kelvin connection was confirmed.

Summarising, the ideal practical *LC* filter achieves the plateau of line B and maintains it to a high frequency. The limiting factors are choke self-resonant frequency, filter capacitor ESR and L_{series}, so the ideal *LC* filter uses a capacitor having a Kelvin connection.

Estimation of Wide-Band LC Response

The significance of the *LC* model is that once we know the self-resonant frequency of the iron choke, we also know where the filter's attenuation becomes constant. Further, since we know that the line down to the self-resonant frequency falls at 12 dB/octave (or 40 dB/decade, if you prefer), we know the maximum high-frequency attenuation. It is easy to measure a choke's self-resonant frequency using an oscillator and an oscilloscope, but that takes time, and we might need a quick 'down and dirty' estimation. The author measured the self-resonant frequency of a number of HT chokes and plotted this against their inductance (see Figure 5.34).

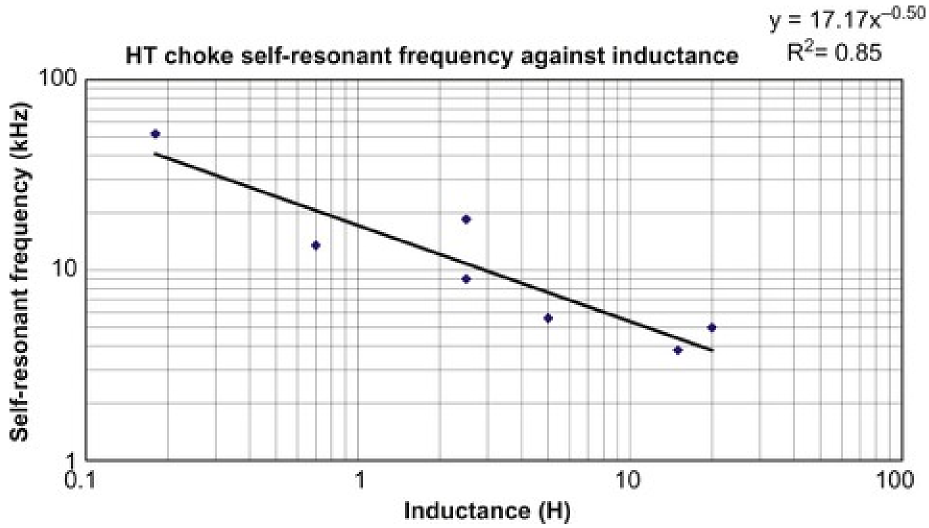


Figure 5.34 Self-resonant frequency against stated inductance for a number of iron-cored power supply chokes.

The graph shows that we can estimate a choke's self-resonant frequency as being 20 divided by the square root of its inductance. As an example, if we were wondering about the suitability of a 16 H choke, we could estimate its self-resonant frequency as being 5 kHz. Further, we could note that from 100 Hz to 5 kHz is roughly five-and-a-half octaves, so we would know that the maximum attenuation compared to 100 Hz would be $5.5 \times 12 \approx 66$ dB. If we had already simulated the choke and accompanying capacitor in PSUD2, we would know the expected ripple at 100 Hz, so we would now have an estimate of the noise at the line B plateau. Obviously, if we knew the exact values of all the parasitic components in our *LC* filter, we could model it in T/spice. However, it takes time and a little care to determine these values to any degree of accuracy, so if we are happy with a result accurate to ±6 dB, an estimation is fine.

Sectioned RC Filters

We might have carefully designed the first stage of an HT power supply (choke or capacitor input) with the available parts so that it produces 2 V_{pk-pk} of ripple, yet we might need ripple <1 m V_{pk-pk}, but could afford to drop some DC voltage. Thus, we need a filter that can attenuate the ripple by a factor of >2,000. Since an *RC* filter is a potential divider, the attenuation is R/X_c (provided that this ratio is reasonably large). Suppose that in our example, we can tolerate 2 kΩ of resistance, so X_c=2 kΩ/2,000=1 Ω. Since the ripple frequency is 100 Hz, we find the required capacitance using:

$$C = \frac{1}{2\pi f X_c} = \frac{1}{2 \times \pi \times 100 \times 1} = 1{,}590\ \mu\text{F}$$

This is a very large capacitor and represents a brute force solution to the problem. A more efficient alternative is to make a filter out of a cascade of sections each using a smaller resistor and capacitor (see Figure 5.35).

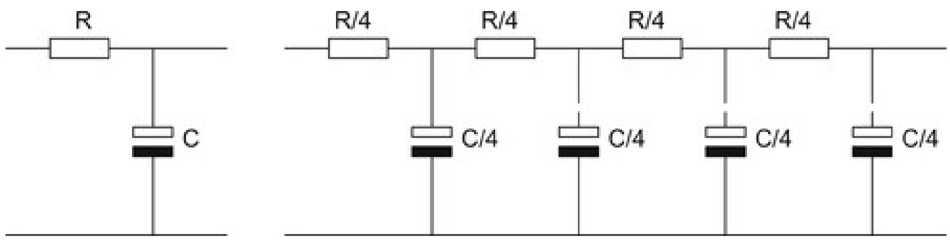


Figure 5.35 Sectioning the RC filter leaves total resistance and capacitance unchanged, but increases attenuation because ultimate slope increases from 6 dB/octave to 24 dB/octave.

The problem is to determine how many sections are ideal. Fortunately, Scroggie [12] (writing as ‘Cathode Ray’) has already investigated this problem and produced Table 5.6 .

Table 5.6 Cascaded Section Coefficients (after Scroggie)

No. of sections	$2\pi fCR$ (R_{total}/X_C)	Attenuation	*RC* in kΩ·μF per section	
			100 Hz	120 Hz
1	16	16	25.5	21.2
2	45.6	130	18.1	15.1
3	90	997	15.9	13.3
4	149	7,520	14.8	12.4
5	223	56,400	14.2	11.8
6	311	420,000	13.8	11.5
Some values in Table 5.6 differ from the original reference because Scroggie did not have the benefit of a spreadsheet to accurately calculate his values.				

To understand Table 5.6 using our previous example, we need attenuation >2,000, so the first number of sections that can exceed this in the attenuation column is n=4. If the resistance must total 2 kΩ, each section must be 2 kΩ/4=500 Ω. To find the individual capacitance required, we use the 100 Hz *RC* column. The capacitance required is 14.8/0.5=29.6 μF. In practice, we would probably use 470 Ω resistors and 33 μF capacitors. The key point is not just that the four 33 μF capacitors are likely to be much cheaper (and smaller) than a single ≈1,590 μF capacitor, but that the sectioned filter promises almost four times the attenuation. PSUD2 can analyse multiple *RC* filters, so the combination of Scroggie’s table and a PSUD2 simulation can be very effective.

Alternatively, you might have a large bag of 22 μF capacitors, and enough room to use four, but must use 2.5 kΩ of series resistance. How can the capacitors be best used to attenuate 100 Hz ripple? Connecting the four capacitors in parallel gives a total capacitance of 88 μF, so the ratio of R_{total}/X_c=138. Inspecting the

R_{total}/X_c column, 90 is the first solution exceeded by 138, so three sections should be used. Each resistance is thus 2.5 kΩ/3=833 Ω. If we only use three sections, our total ratio R/X_c is reduced by a factor of 3/4, so it falls to 104, but this is still optimum with three sections and gives an attenuation of 997, whereas using four parallel 22 μF capacitors with a 2.5 kΩ series resistor would only have given an attenuation of 138. Our factor of improvement is 7, yet we have used one fewer capacitor (saving space).

Regulators

The best way of improving a power supply is to use a voltage regulator. A voltage regulator is a real-world approximation of a Thévenin source; it has a fixed output voltage and an output resistance that approaches zero. A true Thévenin source implies infinite current capacity, whereas the supply that feeds a regulator has limited current capacity. It is therefore important to realise that the regulator can only simulate a Thévenin source over a limited range of operation, so we must ensure that we remain within this range under *all* possible operating conditions.

All voltage regulators are based on the potential divider. Either the upper or the lower leg of the divider is made controllable in some way, and by this means, the output voltage can be varied (see Figure 5.36).

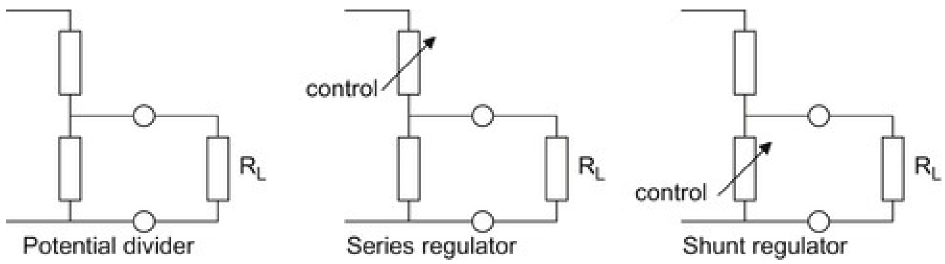


Figure 5.36 Relationship between voltage regulators and the potential divider.

If the upper element is made controllable, then the regulator is known as a *series* regulator because the controlled element is in *series* with the load. If the lower leg of the divider is controlled, then the regulator is known as a *shunt* regulator because this element is *shunted* by the load. Shunt regulators are usually inefficient compared to series regulators, and their design has to be carefully tailored to their load, but they have the advantage that they can both source and *sink* current.

The Fundamental Series Regulator

The fundamental elements of a series voltage regulator are shown in Figure 5.37 .

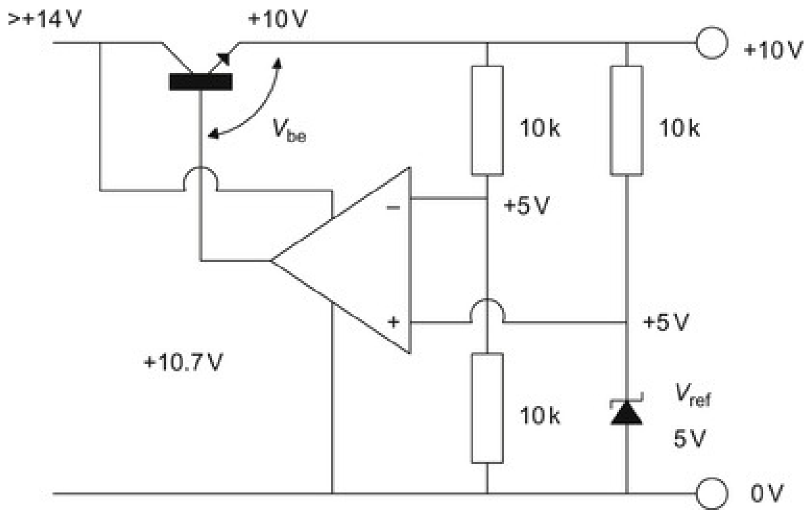


Figure 5.37 Fundamental series regulator.

The circuit shown uses semiconductors, but a valve version could equally well be built. The error amplifier amplifies the difference between the reference voltage and a fraction of the output voltage and controls the series pass transistor such that a stable output voltage is achieved.

The circuit depends for its operation on negative feedback. We saw in Chapter 1 that when feedback is applied, input and output resistances change by ratio of the feedback factor $(1+\beta A_0)$. Voltage regulators rely on shunt-derived feedback reducing the output resistance of the system by the ratio of the feedback factor.

Suppose initially that the regulator is working and that there is 10 V at the output. By potential divider action, there must be 5 V on the inverting input of the operational amplifier. The voltage reference is holding the non-inverting input at 5 V. The series pass transistor is an emitter follower fed by the error amplifier, and has 10 V on its emitter, so the base must be at 10.7 V.

Suppose now that the output voltage falls for some reason. The voltage at the midpoint of the potential divider now falls, but the voltage reference maintains 5 V. The error amplifier now has a higher voltage on its non-inverting input than on its inverting input, and its output voltage must rise. If the voltage on the base of the transistor rises, its emitter voltage must also rise. The circuit therefore opposes the reduction in output voltage.

Since the same argument works in reverse for a rise in output voltage, it follows that the circuit is stable and that the output voltage is determined by the combination of the potential divider and the reference voltage. If we redraw the regulator, we can see that it is simply a non-inverting amplifier whose gain is set by the potential divider and that it amplifies the reference voltage (see Figure 5.38).

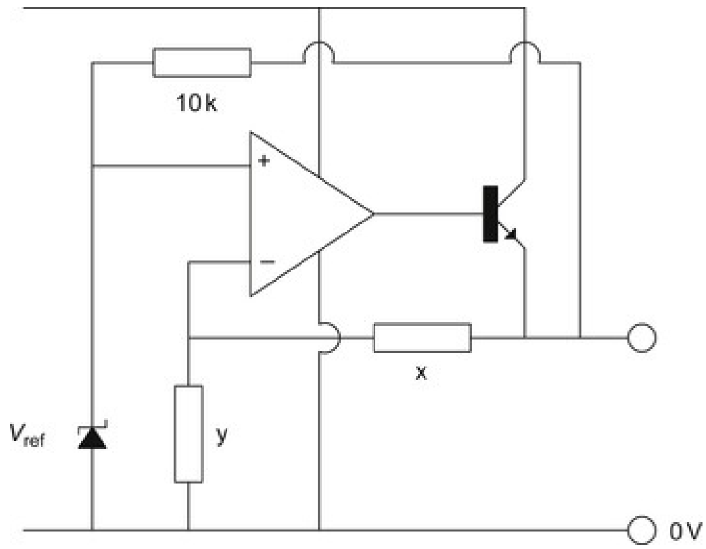


Figure 5.38 Series regulator redrawn to show kinship to non-inverting amplifier.

By inspection, the output voltage is therefore:

$$V_{out} = \frac{x + y}{y} \cdot V_{ref}$$

Since the error amplifier simply amplifies the reference voltage, any noise on the reference will also be amplified, and we should feed it from as clean a supply as possible. Although the argument seems like a snake chasing its own tail, if we feed the reference from the output of the supply (which is clean), then the reference will be clean, and the output of the supply will also be clean. It might be thought that supplying the current for the reference voltage from the output voltage would cause instability, but this is not a problem in practice.

It should be noted that all regulators need an input voltage higher than their output voltage. The minimum allowable difference between these voltages before the regulator fails to operate correctly is known as the *drop-out* voltage (because the regulator 'drops out' of regulation). With this particular design it is only a few volts, but drop-out voltage for a valve version could be 40 V or more.

The Two-Transistor Series Regulator

The two-transistor series regulator is a very common and useful circuit (see Figure 5.39).

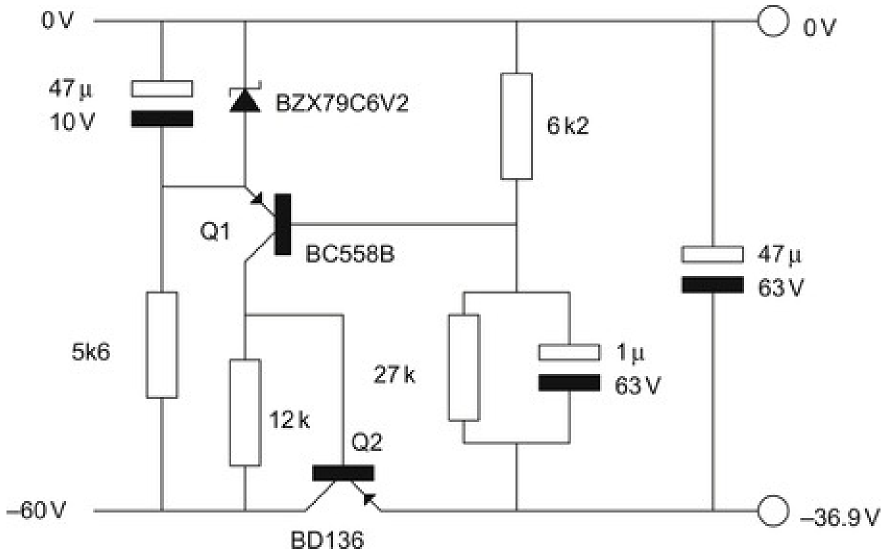


Figure 5.39 Basic two-transistor negative regulator.

This circuit is popular because of its extreme cheapness, but despite that, its performance is really quite good. Q2, the series pass transistor, is fed from the collector of Q1, a common emitter amplifier. The emitter of Q1 is held at a constant voltage by the voltage reference, whilst its base is fed a fraction of the output voltage by the potential divider. If the output voltage rises, Q1 turns on harder, drawing more current, its collector voltage (connected to the base of Q2) falls, causing the emitter voltage of Q2 (which is the output voltage) to fall, thus counteracting the initial error.

This circuit is ideal for use as a bias voltage regulator in a power amplifier because we often need to drop more volts than an IC regulator would tolerate.

As presented, the circuit can only supply 50 mA of output current because the base current for Q2 is stolen from the collector current of Q1. If we increased the collector current of Q1, Q2 could steal more, and output current could be increased, but a better solution would be to replace Q2 with a Darlington transistor, which would need less base current. Alternatively, Q2 could be a power MOSFET, but it would need a gate-stopper resistor of ≈100 Ω soldered directly to its gate pin.

The Zener diode passes 12 mA, which is quite sufficient to ensure that it operates correctly, and has a stable output voltage with minimum noise. A 6.2 V Zener has been chosen because it has lowest temperature coefficient and lowest slope resistance, but it still produces some noise, so it is bypassed by the 47 µF capacitor.

The Speed-Up Capacitor [13]

This capacitor is connected across the far resistor of the potential divider. Its purpose is to increase the amount of negative feedback available at AC, and thereby reduce hum and noise. Since any linear regulator can be considered to be composed of an op-amp enclosed by a feedback loop, a generic graph may be drawn (see Figure 5.40 a).

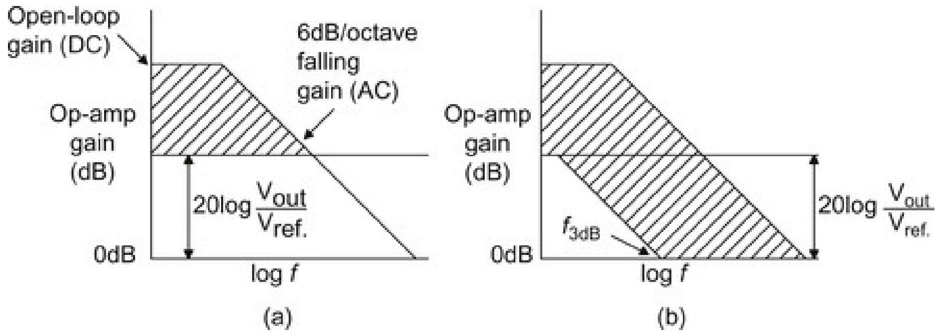


Figure 5.40 The effect of the speed-up capacitor on ripple rejection.

The op-amp gain is a combination of DC open-loop gain and gain that falls at 6 dB/octave with frequency. The gain within the hatched area is available for attenuating incoming ripple, so ripple reduction is maximised by:

- Maximising DC open-loop gain
- Maximising low frequency corner frequency (741: $f_{corner} \approx 20$ Hz; 5534: $f_{corner} \approx 1$ kHz)
- Minimising the ratio of DC output voltage to reference voltage.

Although we need the op-amp to have the correct gain at DC to set the output voltage correctly, the gain below the hatched area is wasted. The speed-up capacitor aims to recover some of this wasted gain (see Figure 5.40 b).

At first sight, it would seem that f_{-3dB} should be placed sufficiently low that all of the previously wasted gain is recovered. However, an oversized speed-up capacitor slugs the response of the regulator to changes in load current.

The maximum value for this capacitor is found by first calculating the AC Thévenin resistance that it sees:

$$r_{\text{Thevenin}} = \frac{1}{(1/h_{ie}) + (1/x) + (1/y)}$$

Remembering that

$$h_{ie} = \frac{h_{fe}}{g_m}$$

and also that

$$g_m = 35 I_c$$

we find that for this circuit, with h_{fe}=200 and I_c=2 mA, h_{ie}≈2.9 kΩ. So the Thévenin resistance seen by the capacitor is ≈1.8 kΩ.

We would like the capacitor to have a significant effect on the lowest ripple frequency to be attenuated, which is 100 Hz (120 Hz US). The potential divider chain and capacitor is a step equaliser whose effect on the regulator is similar to that used for the RIAA 3,180 µs/318 µs pairing in Chapter 7 . We could make the reactance of the capacitor at the lowest ripple frequency equal to the Thévenin resistance at the tapping of the potential divider, which would mean that an infinitely large capacitor could only improve ripple reduction by a further 3 dB:

$$C = \frac{1}{2 \times \pi \times 100 \times 1{,}800}$$

$$C = 870 \text{ nF}$$

The nearest value is 1 µF. This is quite a small capacitor, and the author has seen many similar circuits with oversized capacitors and, indeed, built one himself. The subjective effect of the oversize capacitor was to create a bass boom that was incorrectly thought at the time to be due to room acoustics when it was really due to the sluggish recovery of the power supply to transients.

At the opposite limit, we could set the reactance of the capacitor relative to the total resistance of the divider chain. The smaller capacitor would only give a 3 dB improvement in hum compared to a chain without a capacitor, but its low frequency transient response would be better than a regulator using a larger capacitor.

The value of the speed-up capacitor is a compromise between hum reduction and regulator low frequency transient response, so there isn't a 'correct' answer here other than that the capacitor should be small. You might even want to determine its final value by listening because different loudspeakers (with different low frequency damping) can prefer different values.

Compensating for Regulator Output Inductance

The regulator also has a capacitor across its output. As shown in Figure 5.40 , the gain of the error amplifier falls with frequency due to Miller effect and stray capacitances, so the amount of gain available for reducing output impedance falls. If (1+ βA_0) has fallen, then the output impedance must rise, and the effect is that output impedance rises with frequency. A perfect Thévenin source in series with an inductor would look identical, and for this reason the output of regulators is often described as being inductive at high frequencies. The output capacitor maintains a low output impedance at high frequencies.

A Variable Bias Voltage Regulator

We often need a bias voltage regulator to be variable between certain limits. In this example, we will look at a grid bias regulator needed for an 845 directly heated triode. Perusal of RCA anode characteristics (*circa* 1933) indicated a grid bias voltage of −125 V, but modern valves do not match the original curves exactly, and we must equalise anode currents in this (push–pull) output stage to avoid saturating the output transformer with an unbalanced DC current and causing distortion. A range of ±25 V on either side of the nominal −125 V seems reasonable, but how do we design a regulator to fulfil this requirement?

Fortunately, since it supplies a part of the circuit where the signal voltages are very high (up to 90 V_{RMS}), the regulator need not have an impeccable noise performance, and avalanche diodes are perfectly acceptable (see Figure 5.41).

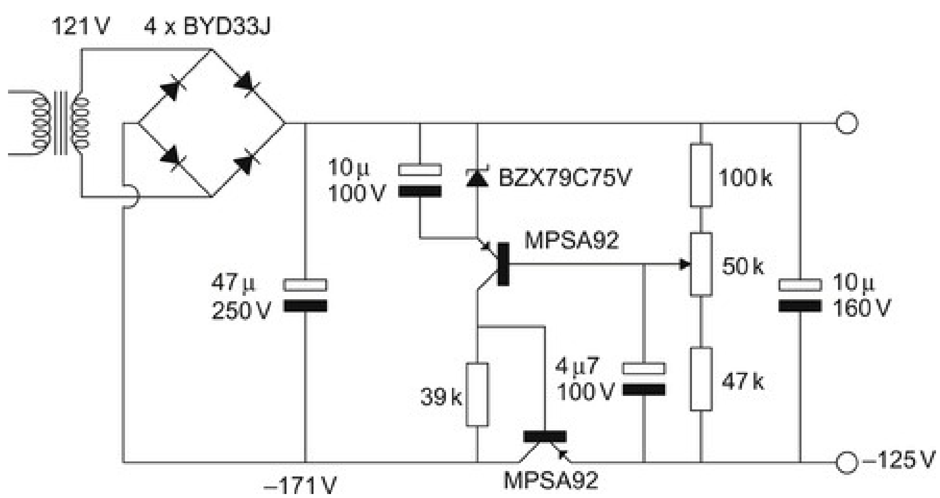


Figure 5.41 Adjustable −125 V bias regulator.

A higher-voltage reference allows the finished circuit to have better regulation, but we must still allow a reasonable voltage between the collector and the emitter of the control transistor. In practice, a reference voltage of about half the maximum output voltage is usually a good choice, and 75 V reference diodes are available.

The reference diode holds the emitter of the transistor at −75 V, and V_{be}=0.7 V, so the base of the transistor will be held at a fixed potential of −75.7 V. Since the base of the transistor is connected to the wiper of the potential divider, the wiper must also be held at −75.7 V, no matter what output voltage is set. We can now calculate the required attenuation of the potential divider for the two extreme design cases:

$$\frac{100\ \text{V}}{75\ \text{V}} = 1.323$$

$$\frac{150\ \text{V}}{75.7\ \text{V}} = 1.9815$$

By choosing a convenient value for the variable resistor in the middle of the potential divider, we now have enough information to calculate the required resistors on either side. A low value of variable resistor would require a large current to flow in the potential divider, whereas too high a value will cause errors due to the (small) base current drawn by the transistor. A good engineering principle is that the potential divider chain should pass roughly 10 times the expected base current, so a 50 kΩ variable resistor was a convenient standard value for this example.

When the wiper of the variable resistor is set to produce the largest voltage at the regulator output, it is connected directly to the grounded resistor (*x*), and vice versa. Using the standard potential divider equation, for −150 V:

$$\frac{x + y + 50}{x} = 1.9815$$

And similarly for −100 V:

$$\frac{x + y + 50}{x + 50} = 1.323$$

We now have two equations that can be solved, either simultaneously or by substitution, to give the values of the fixed resistors *x* and *y*. In this particular case, the values fell out very conveniently to give x=100 kΩ and y=47 kΩ, where *x* is the upper potential divider resistor and *y* is the lower potential divider resistor.

As a final note, bias supplies and their regulators draw very little current from their transformer, so a low current winding is used. Low current windings tend to have very poor regulation, and because their full-load current is rarely drawn, so the actual voltage at the theoretical −170 V point could easily be −180 V when measured.

The 317 IC Voltage Regulator

Although the two-transistor regulator is the ideal choice for a bias regulator because of its high voltage drop capability, once we need higher currents at lower voltages its limitations quickly become apparent.

It is perfectly possible to build a voltage regulator using a handful of components including an operational amplifier, a voltage reference, various transistors, resistors and capacitors. With care, the circuit can be made to work almost as well as an IC regulator and only costs about three times as much. We need not feel guilty about using IC regulators.

The 317 is a standard device that is made by all the major IC manufacturers [14] . Linear Technology [15] makes an upgraded version of the 317, the LT317, but the only difference is that the guaranteed tolerance of the voltage reference is tighter. A commercial design could therefore set its output voltage using fixed resistors rather than a variable resistor, thus saving money, because not only are variable resistors expensive to buy, they also have to be adjusted (which costs money). We do not often have to worry about such considerations, so the standard 317 is fine.

The 317 incorporates all of the fundamental elements of a series regulator in one three-terminal package, and we only need to add an external potential divider to produce an adjustable regulator (see Figure 5.42).

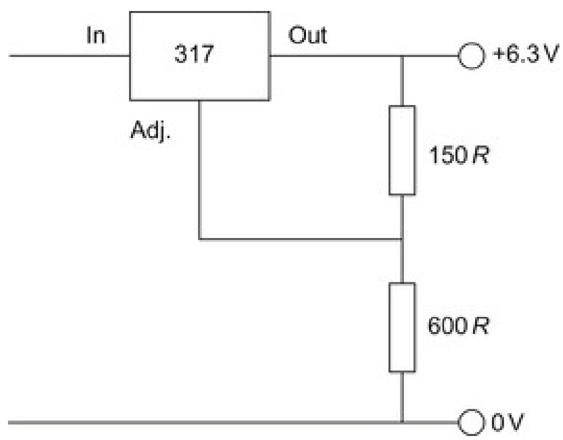


Figure 5.42 Basic 317 regulator circuit.

One end of the voltage reference is connected to the OUT terminal, whilst the other is an input to the error amplifier. The other input of the error amplifier is the ADJ terminal. The 317 therefore strives to maintain a voltage equal to its reference voltage (1.25 V) between the OUT and ADJ terminals. All we have to do is to set our potential divider so that the voltage at the tap is V_{out}=1.25 V, and the 317 will do the rest.

In datasheets for the 317, you will invariably find that the upper resistor of the potential divider is 240 Ω. The reason for this is that the 317 must pass 5 mA before it can regulate reliably. If the potential divider passes 5 mA, then this ensures that the device is able to regulate even if there is no external load.

The 317 sources ≈50 μA of bias current to the opposite rail from the ADJ pin, which therefore flows down the lower leg of the potential divider. Normally, this is negligible, but if you are designing a high voltage regulator, and choose a lower potential divider current, this will need to be taken into account.

The manufacturers' data sheets generally show a regulator with the ADJ pin bypassed to ground by a 10 μF electrolytic, which improves ripple rejection from 60 dB to 80 dB at 100 Hz. This is the speed-up capacitor that we added to the two-transistor regulator, but because the reference voltage is tied to V_{out}, rather than ground, the speed-up capacitor connects to ground, rather than V_{out}.

We could therefore use the method derived earlier to check whether 10 μF is a reasonable value of capacitor. The ADJ pin is an input to an operational amplifier, so we can treat it as infinite input resistance, and we are only concerned with the external resistor values. If we were to use an upper resistor of 240 Ω, and a 2.7 kΩ lower resistor to set an output voltage of ≈15 V, then the maximum value would be 7.2 μF, so a 10 μF electrolytic capacitor is a reasonable choice, although the author would probably prefer 6.8 μF if he had one in stock.

Just like the two-transistor regulator, the output of the 317 is inductive, and the manufacturer's output impedance curves suggest that the output impedance is equivalent to ≈2.2 μH in series with 2.7 mΩ when set to produce 10 V, so they recommend a 1 μF tantalum bead output bypass capacitor, as shown in the equivalent circuit (see Figure 5.43).

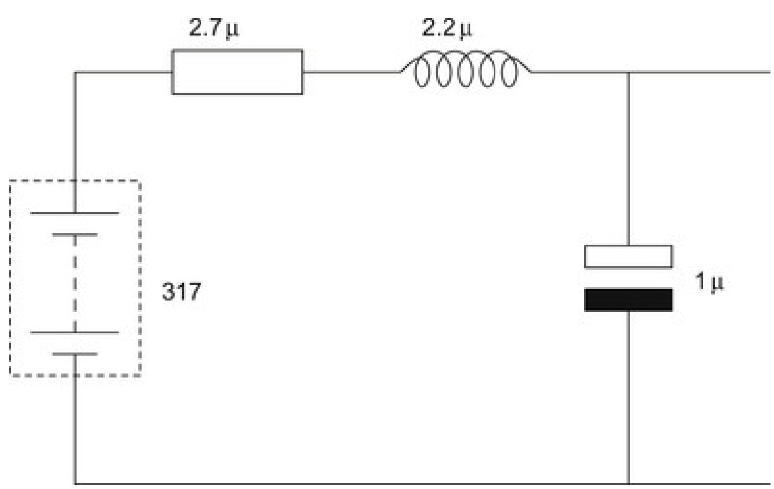


Figure 5.43 AC Thévenin equivalent of 317 plus 1 μF bypass capacitor.

Assuming that the tantalum capacitor had zero ESR (!), the only damping resistance would be the 2.7 mΩ of the 317, so we have an underdamped resonant circuit, and we can calculate its *Q*:

$$Q = \frac{1}{R}\sqrt{\frac{L}{C}} = \frac{1}{2.7 \times 10^{-3}} \cdot \sqrt{\frac{2.2 \times 10^{-6}}{1 \times 10^{-6}}} = 550$$

Wiring resistance will reduce this *Q* considerably, but it will not reduce it to *Q*=0.5, which would be critically damped. This would not matter greatly because we would be unable to excite the circuit from the output (any external excitation would be short-circuited by the capacitor). If we now concede that the capacitor is not perfect, we may be unlucky enough to be able to excite the resonance, and the circuit could become unstable. Rearranging the formula, 3 Ω critically damps the resonance, and the IC manufacturers recommend 2.7 Ω in series with the tantalum capacitor.

The author measured a couple of 1.5 μF tantalum bead capacitors randomly picked from his parts bin and found one had an ESR of 4.8 Ω, whilst the other measured 2.7 Ω – negating any need for the series 2.7 Ω resistor. It does seem that tantalum bead capacitors have very variable ESR not just between manufacturers, but also between samples within batches, so data sheets either need to be read carefully or should be measured before use. An aluminium electrolytic capacitor plus series resistor would be a cheaper and less variable alternative.

The 317 as an HT Regulator

Because the 317 is a floating regulator, there is no reason why it should not be used to regulate a 400 V HT supply. However, because the 317 can only tolerate 37 V from input to output, it needs support circuitry for protection [16] (see Figure 5.44).

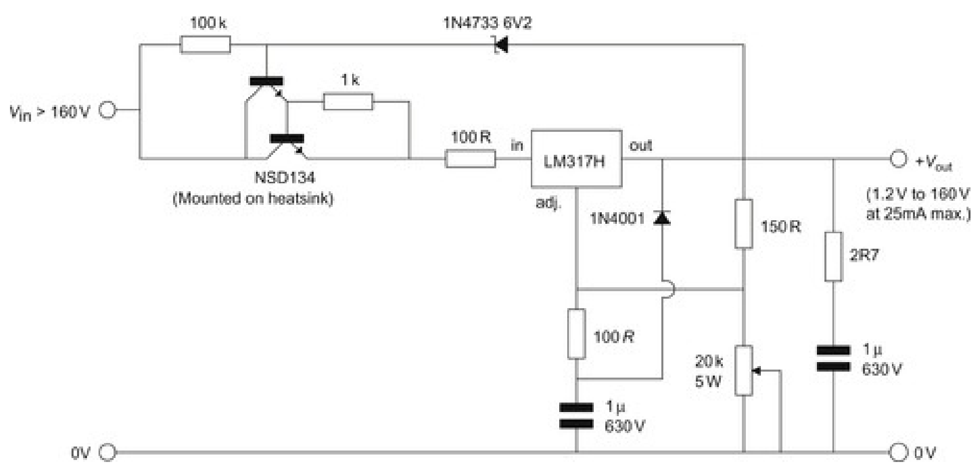


Figure 5.44 Maida high-voltage regulator (reproduced by kind permission of National Semiconductor).

The 317 is preceded by a high voltage transistor Darlington pair whose sole aim in life is to protect the 317 by maintaining $6.2\ \text{V} - 2\ V_{BE} - 0.1\ I_{LOAD(mA)}$ between its IN and OUT terminals. Unfortunately, the voltage drop across the 100 Ω resistor in this particular circuit means that the 317 drops out of regulation once load current >20 mA. The Darlington pair can easily cope with variations in mains voltage, but it should not be thought that this circuit is proof against a short-circuit when used at typical valve voltages.

Accidentally short-circuiting a regulator of this type with an oscilloscope probe results in an almighty bang, and all the silicon is destroyed. The author *knows*.

The lower arm of the potential divider is bypassed, but has a resistor in series with the capacitor to improve low frequency transient response by raising the lower f−3dB frequency of the step equaliser, and a diode has been added allegedly to discharge the capacitor in the event of an output short-circuit (although the author's experience was that it didn't actually help).

Variants of the Maida circuit are very popular as HT regulators, and the author has used many over the years (see the Crystal Palace amplifier in Chapter 6), but the circuit is fragile, and we will see later that it is possible to do better.

Valve Voltage Regulators

Valve voltage regulators have always been very rare, and we will now see why (see Figure 5.45).

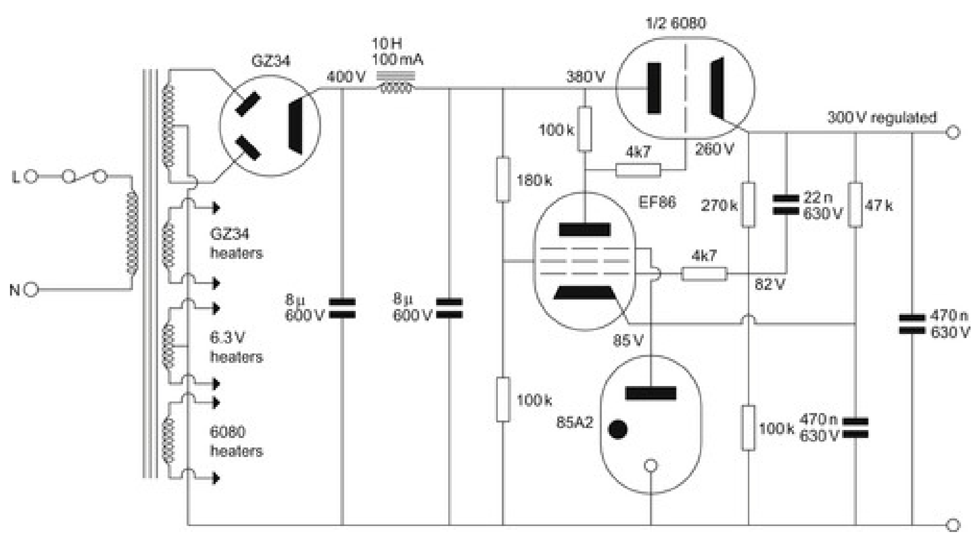


Figure 5.45 Basic valve voltage regulator.

The circuit is directly analogous to the two-transistor regulator; it simply has valves and higher voltages. The silicon reference diode has been replaced by a gas reference that holds the cathode of the EF86 stable at 85 V, whilst the grid is fed from the potential divider. The series pass element is a 6080 double triode ($P_{a(max)}$=13 W), which was specifically designed for use in series regulators and can pass high currents at low anode voltages.

A valve rectifier is used, and in deference to its limited ripple current capacity, an 8 µF paper/foil capacitor has been chosen for the reservoir, although a polypropylene capacitor (≤60 µF for this rectifier) would be physically practical. This results in considerable ripple voltage which is filtered by the following *LC* combination.

Unless the g_2 resistor value is carefully chosen, the performance of the regulator is slightly compromised by feeding g_2 of the EF86 from the raw supply, but if fed from the regulated supply, there is a danger that the circuit might simply sulk and not switch on. The gain of the EF86 is ≈100, and above ≈100 Hz this gain is available for reducing the output resistance of the 6080, whose g_m≈7 mA/V, so r_k≈200 Ω (including the effect of the external 100 Ω R_a). The regulator therefore achieves an output resistance of ≈2 Ω.

Each valve has its cathode floating above ground, implying three separate heater supplies (gas references are cold cathode valves). The 6080 data sheet specifies $V_{hk(max)}$=300 V, so if $V_{regulated}$<300 V, the 6080 heater could be powered from a grounded heater supply. The EF86 could be supplied by a grounded heater supply, but this is putting quite a strain on the heater to cathode insulation; stressing heater/cathode insulation is not recommended because it reduces valve life expectancy and increases their noise.

The EF86 is somewhat noisy (2 µV), but this noise is swamped by the 60 µV noise from the 85A2 (both Mullard-stated values). Even the better valves such as the 85A2 are notorious for *voltage jumps*, an effect whereby the reference voltage jumps by typically 5 mV if the operating current changes. Sadly, smaller random jumps occur even with a constant current, as shown by the captured trace of a Mullard M8223 150 V special quality gas stabiliser operating at 20 mA (see Figure 5.46).

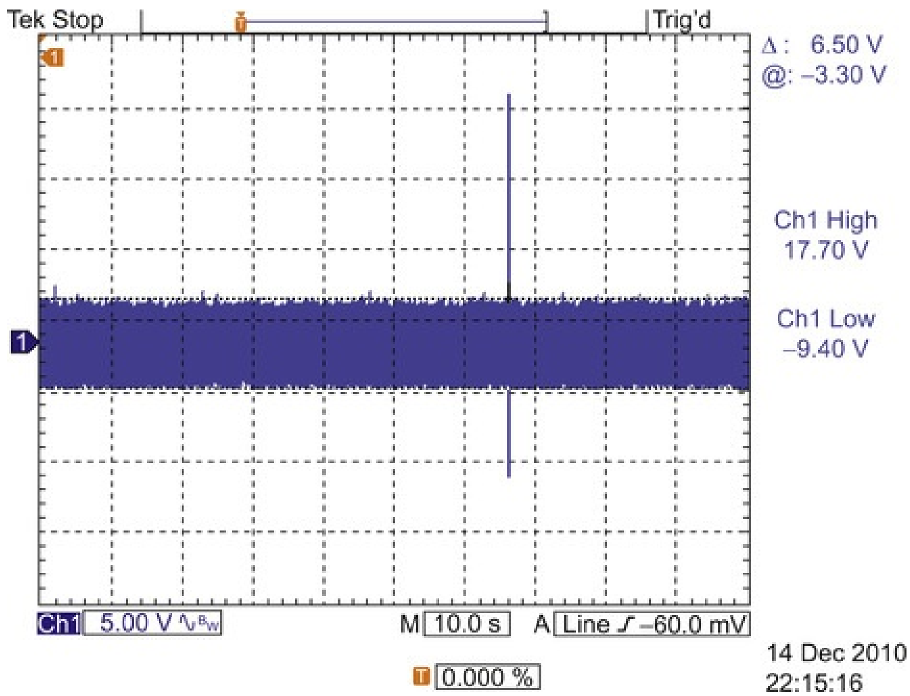


Figure 5.46 Voltage steps in M8223 gas reference.

The oscilloscope trace requires careful interpretation because it is preceded by 75 dB (≈×5,600) of AC coupled gain, so information can only be gleaned from edges, not DC levels. The trace shows a positive pulse due to a positive jump closely followed by a negative pulse due to a negative jump. As measured by the oscilloscope, the amplitude of the positive pulse is 17.7 V, but as the NE5532 op-amp preceding it has ±18 V rails, it can be assumed that it was clipped and that the original jump exceeded 3.2 mV. However, the negative step at −9.4 V would not have been clipped, and as the oscilloscope was in 'peak detect' mode, its amplitude can be assumed to be correctly captured at −1.7 mV. Summarising, the burning voltage jumped up by >3.2 mV and almost immediately fell back by 1.7 mV to a new slightly higher burning voltage than before the two jumps.

Maximum stability is achieved by stabilising the valve at the manufacturer's preferred operating current, but if this current changes, even if it returns to its original value, the valve takes time to recover its original stability.

Although a gas reference may operate at 150 V, it needs extra energy to strike the glow discharge. There are three possible source of this extra energy:

- Provide a sufficiently high (current limited) striking voltage. (Mullard 150C4 in total darkness needs 225 V.)
- Allow the photons in the ambient light to provide the extra energy. (Mullard 150C4 striking voltage drops to 185 V.)

- Add a radioactive gas such as ^{85}Kr that emits beta particles (electrons) having an average energy of 200 keV.

The significance of these three sources is that if the last two are not available, a gas reference may require a higher striking voltage than expected. Krypton 85 has a half-life of only 10.756 years, so after 40 years it will have decayed to 7.6% of its original activity, significantly diminishing its striking assistance. Moral: Don't rely on a manufacturer's specified striking voltage if there's any reason to think they used radioactive assistance to lower it. The GEC QS1215 90 V stabiliser's data sheet states that an (unspecified) radioactive material is used to maintain the same maximum striking voltage (115 V) irrespective of light or darkness, but the Mullard 90C1 has similar specifications with no mention of radioactivity. Despite these theoretical caveats, the author's sole QS1215 only needed a striking voltage of 107.13 V in total darkness – well within specification.

Optimised Valve Voltage Regulators

Oscilloscope design presents many challenges because a bandwidth of DC to at least 20 MHz is required. Valve oscilloscopes required stable, quiet HT supplies, so their voltage regulators were carefully optimised and heater voltages were stabilised against mains voltage variation by control circuits involving a saturable inductor in series with the mains winding of the heater transformer [17].

Any voltage regulator can be improved by increasing the gain of its error amplifier. A single triode has the lowest gain, but a pentode (or cascode) has higher gain. If even greater gain is required, a pair of stages can be cascaded (more than two stages would be impractical because phase shifts would almost certainly turn the regulator into a power oscillator). Because the error amplifier amplifies DC, drift must be minimised, so the first stage of a high-gain regulator *must* be a differential pair, and a dual triode is convenient. The second stage is much more flexible, and could be another triode differential pair, or a single-ended stage using either a triode or a pentode.

Using a Pentode's g_2 as an Input for Hum Cancellation

If a pentode is used as the second stage, g_2 can be considered to be an inverting input. If the correct proportion of raw HT ripple could be injected at this point, it would cancel at the anode, resulting in a regulator with no hum at its output, but the tactic is not without its problems:

• For the pentode to operate correctly, g_2 must be at the correct DC potential. This is usually derived from a potential divider across the (clean) output of the supply. A large-value resistor can then be connected from the raw HT, and its value adjusted until ripple is cancelled. The exact value of this resistor is awkward to calculate because we do not usually know the value of $\mu_{(g_2 - a)}$, so its value is usually determined by experiment. Values could be anywhere in the range from 150 kΩ to 1.5 MΩ.

• Although valve manufacturers specified most parameters quite tightly, we now rely on an unspecified parameter, and there is no guarantee that valves made by different manufacturers that meet all the specified parameters will match our unspecified parameter. As an example of this problem, the 1970s four-tube EMI2001 colour camera set beam current by controlling g_2, so whenever a new tube was ordered (£1500 a shot in 1986), it was necessary to specify that the tube was to be used in an EMI2001 colour camera. Similarly, Tektronix stocked selected tubes (valves), not because they were better than any others, but because they were guaranteed to work correctly in *their* circuit.

• Variations between valves mean that the cancellation is not perfect, but any remaining ripple is easily mopped up by the loop gain of the error amplifier.

Increasing Output Current Cheaply

The majority of circuitry within oscilloscopes and audio is Class A, so it draws a very nearly constant current. One of the functions of a regulator is to regulate output voltage against changes in load current, but if the current is almost unchanging, much of the regulating ability is being wasted. As an example, the regulator might face a load with a quiescent current of 100 mA but that could rise to 150 mA, or drop to 50 mA under certain circumstances. We could design the regulator to be able to pass 150 mA, but this would need a bigger series pass valve. Instead, we could bypass the series pass valve with a resistor that allowed 50 mA to flow directly into the load. The series pass valve now only has to pass 100 mA under full load. When the load requires only 50 mA, this is provided entirely by the bypass resistor, and the regulator is in danger of dropping out of regulation, so this condition sets the limit for the maximum current that may be bypassed by a resistor.

Adding the bypass resistor slightly increases ripple because it injects raw HT into the clean circuit, but because the output resistance of the regulator is likely to be <1 Ω, potential divider action greatly reduces the added ripple. As an example, the following circuit incorporates both of these modifications (see

Figure 5.47). The regulator showcases some other tricks that improve performance.

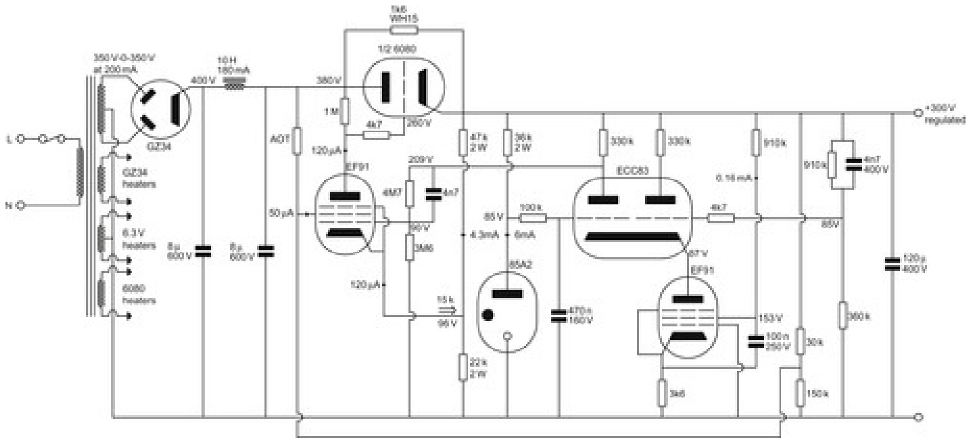


Figure 5.47 Optimised valve voltage regulator.

As previously mentioned, gas references produce noise, but because we have chosen to use a differential pair, the gas reference now drives a high-impedance input, so we can add a filter to reduce noise. The capacitor previously across the reference has been removed because of the danger of it causing oscillation when excited by voltage jumps (this was previously damped by r_k of the valve). Additionally, the current through the gas reference has been stabilised at the manufacturer's preferred operating current, so jumps should be minimal.

The ECC83 differential pair has its anodes at 209 V, and although it would just be possible to directly couple this voltage to the grid of the EF91 pentode, its cathode would be at ≈213 V, which would not only cause problems with V_{hk}, but would reduce gain because of the necessarily high value of R_k. To reduce this problem, V_k has been reduced to a similar voltage to the cathodes of the ECC83, allowing them to share a heater supply. We could simply insert a cathode resistor to ground, but a potential divider across the regulated output can set the required voltage and give a much lower Thévenin output resistance (15 kΩ versus 800 kΩ). The significance of this resistance is that it reduces the gain of the stage, so we want as small a resistance as possible to maintain maximum open-loop gain in the regulator.

To couple an anode of the ECC83 to the EF91, a potential divider is needed to drop the voltage from 209 V to 90 V, thus we sacrifice ≈7 dB of DC open-loop gain. However, the sacrifice is worthwhile because gain is recovered faster by dropping V_k (and reducing local feedback) on the EF91 than it is lost by the potential divider. Ultimately, the choice of V_k is usually determined by V_{hk}

considerations. Nevertheless, we could recover the gain at AC by bypassing the upper resistor with a capacitor, although we would need to check that the circuit was still stable.

The ECC83 differential pair has a constant current sink tail. If we were making symmetrical split rail supplies, we could simply take a large tail resistor to the negative supply, but a single supply needs the constant current sink.

Finally, because of the greatly increased open-loop gain, the regulator has a much lower output DC resistance than before (<10 mΩ), so it needs a commensurately large bypass capacitor to maintain low output impedance at higher frequencies. The author's recent capacitor measurements show that only plastic capacitors have a sufficiently low ESR to shunt 10 mΩ, but 120 µF 400 V plastic capacitors are bulky and expensive.

As can be seen, much can be done to improve the basic valve voltage regulator, but the penalty is considerable cost and complexity.

Regulator Sound

Single-ended amplifiers (whether pre-amplifiers or power amplifiers) supplied from a regulator force the error amplifier to track the musical waveform. This is because the amplifier draws a current proportional to the music, and the regulator's error amplifier strives to maintain a constant voltage in the face of this changing current. At high frequencies, the output shunt capacitor is a short-circuit and maintains a low output impedance, but at low frequencies it is the error amplifier that must do the work and cope with the (musical) current waveform. The quality of the regulator is therefore inevitably audible. Nevertheless, regulator defects are still an order of magnitude below passive supply defects.

Power Supply Output Resistance and Stereo Crosstalk

In Chapter 2 , we designed simple stages and made the explicit assumption that our power supply had zero output resistance at all frequencies from DC to light. Feedback regulators have low, but not zero, output resistance, so this parameter is often specified. As examples of typical 300 V regulators at low frequencies, a well-implemented simple valve regulator can achieve ≈2 Ω, an optimised valve regulator ≈10 mΩ and a typical Maida regulator ≈80 mΩ.

The question is, how important is power supply output resistance? Or, more specifically, how high can it be before it becomes a problem? All gain stages use their load resistance R_L to convert a change of signal current into a signal voltage. If the power supply has AC output resistance r_{supply}, then this resistance is in series with R_L and slightly increases the gain of the stage:

$$A_v = \frac{\mu(R_L + r_{supply})}{(R_L + r_{supply}) + r_a}$$

If we first assume that r_{supply} is constant with frequency, then there's no problem for a single stage. However, if we share this supply between a pair of stereo channels, the signal current of one channel's stage crosstalks into the other, via a pair of potential dividers (see Figure 5.48).

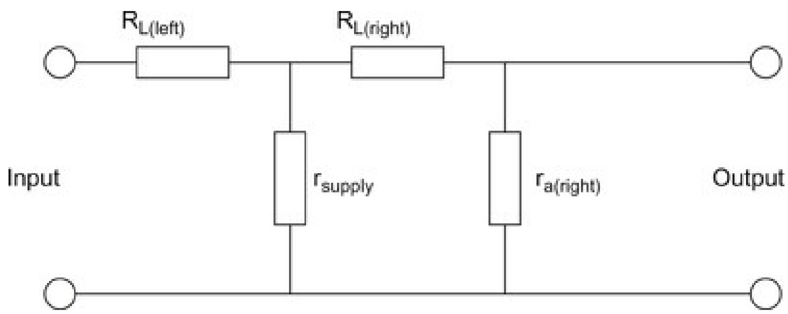


Figure 5.48 The two potential dividers that determine stereo crosstalk when a common supply is used.

Rigorous calculation of this cascade of potential dividers is unnecessary because we know r_{supply} must be quite small, so we simply determine individual potential divider attenuations, and then multiply them together. As an example, a stereo pair of common cathode E88CC stages might have R_L=33 kΩ, r_a=6.6 kΩ and r_{supply}=2 Ω:

$$\text{Crosstalk} = \frac{r_{supply}}{r_{supply} + R_L} \cdot \frac{r_a}{r_a + R_L} = \frac{2}{2 + 33{,}000} \cdot \frac{6{,}600}{6{,}600 + 33{,}000} = -100\ \text{dB}$$

Crosstalk at −100 dB is certainly adequately low, but if we increase r_{supply} to 200 Ω, crosstalk deteriorates to −60 dB, which although perfectly acceptable for an RIAA stage (cartridge crosstalk rarely exceeds −35 dB, even in the mid-band) would undoubtedly be criticised elsewhere despite the fact that a 26 dB difference in level is sufficient to slew the apparent position of the source firmly to one loudspeaker.

However, we should note that because the factor R_L appears in both denominators, the most effective way of reducing stereo crosstalk is by maximising R_L, rather than minimising r_{supply}. Note also that r_{supply} is likely to be inductive due to falling regulator error amplifier gain with frequency, implying crosstalk that deteriorates with frequency.

Power Supply Output Resistance and Amplifier Stability

When we calculated stereo crosstalk, the second potential divider described how well the stage could reject a signal from the power supply, so it was the Power Supply Rejection Ratio (PSRR) term we first saw in Chapter 2 , whereas the first potential divider described how well the power supply could attenuate an injected signal. Whenever individual stages are interconnected to form a system, each stage requires power, which must ultimately be derived from a common source (even if that common supply is the AC mains entering the building). Thus, the stereo crosstalk concept can be extended to include coupling between two entirely different stages via the power supply.

The significance of common power supply coupling between two entirely different stages is that one might be the input to the other, so if the output of the second is coupled to the first via the common power supply there might be sufficient loop gain and phase shift to allow oscillation. If the common power supply had zero output resistance, there could be no coupling, but this is not possible. Moreover, the common power supply is unlikely to have a constant resistance and is more likely to have a complex impedance that changes with frequency.

Traditional electronics used an RC ladder network as a means of progressively attenuating power supply ripple from the (less sensitive) output stage to the (most sensitive) input stage. The smoothing capacitors are an open circuit at very low frequencies, so they no longer effectively attenuate power supply coupling from the output stage back to the input stage, and if there is enough low-frequency gain in the amplifier, the system can become a blocking oscillator [18] . This low-frequency (≈1 Hz) phenomenon was known classically as *motorboating*, but marginal stability probably went unnoticed much of the time, because loudspeakers of the time had very stiff suspensions and their cones were rarely visible. The fault was in the power supply, yet the traditional 'cure' was to reduce the value of one of the amplifier's coupling capacitors.

Rather than having an RC ladder network, we could connect all the stages to a single regulator. Provided that the regulator has zero output resistance, there can be no coupling from one stage to another via the common power supply. In practice, a feedback regulator always has a complex output impedance that rises with frequency (hence the 2.7 mΩ+2.2 μH inductance of the 317) in order to maintain stability of its error amplifier. Thus, a feedback regulator has low enough output impedance at low frequencies to prevent motorboating, but its rising impedance at high frequencies might allow the circuit it powers to burst into high frequency oscillation instead.

A guaranteed way of breaking the loop is to add a regulator per stage because this adds the PSRR of each regulator (typically 60 dB) to the PSRR of the stage (anywhere between 0 dB and 60 dB). Solid-state electronics could cheerfully add a ±15 V regulator to each stage at minimal cost, but HT regulators are more difficult; a clutch of Maida regulators would be a significant investment, and the idea of multiple valve regulators is mindboggling (although that didn't stop Glen Croft in 1984). Thus, we need something cheap, simple and robust.

The Statistical Regulator

This circuit acquired its name because it applies simple statistical rules in a novel way, resulting in a simple regulator producing extremely low noise and having very high ripple rejection.

As we saw earlier, all regulators are based on the potential divider; series regulators control the upper arm and shunt regulators the lower arm. It's a good engineering maxim that if a large improvement is needed, it's easier to obtain it in a number of small bites than one large one. We have to work hard to achieve 80 dB of attenuation in a single regulator, but obtaining two attenuations of 40 dB is easy. Thus, we could cascade regulators in the same way that Scroggie showed that cascaded filters achieved high attenuation more efficiently than a single brute force filter, but this multiplies the drop-out voltage by the number of regulators. What would be better would be to make a regulator that controls both arms of the potential divider simultaneously. The individual arms need not be especially good because their combination would make a very good regulator.

It is probably possible to make an adjustable high voltage regulator having a feedback loop controlling series and shunt elements simultaneously that doesn't explode at switch-on. Whether it would be stable when connected to a real load is another matter. Fortunately, we won't need our regulator to be adjustable because we will know the exact voltage and current required. Thus, we do not need a potentially unstable control loop – we could optimise the two arms independently and run them open-loop.

The ideal upper arm would have infinite slope resistance, whereas the ideal lower arm would have zero slope resistance. We saw in Chapter 2 that we could make a very good constant current sink using two DN2540N5s and three resistors, so this would make an excellent (and simple) upper arm. The lower arm needs to be a constant voltage and at its very simplest could be a string of gas references or Zener diodes (see Figure 5.49).

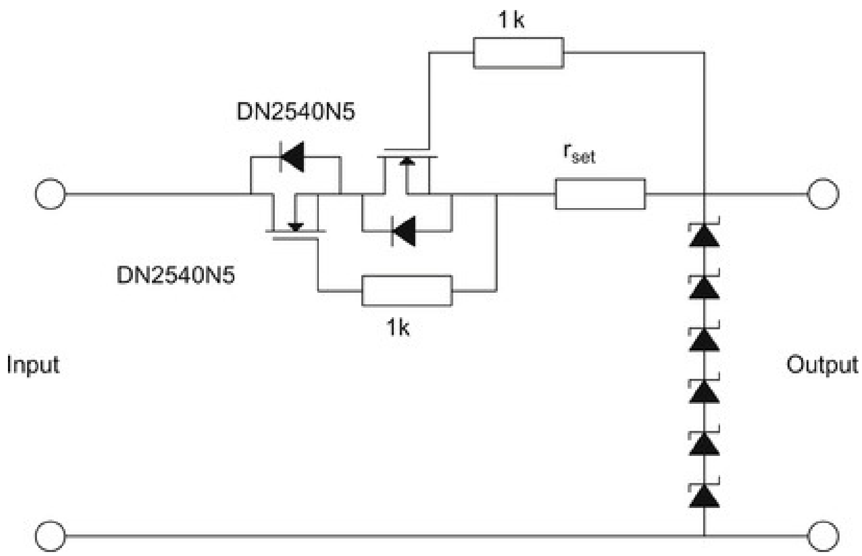


Figure 5.49 This regulator controls both upper and lower arms of the potential divider.

Suppose a load needed 195 V at 15 mA. To allow for signal variation in load current and to ensure that the voltage references operate correctly, the lower arm should pass a minimum of 10 mA. Thus, the constant current sink must be designed to pass 25 mA. Curve tracer measurements of a DN2540N5 at 25 mA and 100 V suggested $g_m \approx 148$ mA/V and $r_{ds} \approx 41$ k, resulting in $\mu \approx 6{,}000$. For that particular device, V_{gs} for 25 mA≈1.45 V, so a 56 Ω programming resistor would be needed. The cascode constant current sink multiplies the value of its programming resistor by the product of both μ, but the lower DN2540N5 is forced to operate at a very low voltage, badly compromising both g_m and r_{ds}, drastically reducing μ. Careful AC measurement suggested $r_{slope} \approx 30$ MΩ with an uncertainty of ±12% for a DN2540N5 cascode constant current source set to 25 mA.

Suppose that the lower arm was a series chain of Zener diodes. We saw in Chapter 3 that a composite device made of multiple 5.6 V Zeners was particularly quiet. To achieve 195 V (actually 196 V), we would need 35 diodes, and a total slope resistance of ≈400 Ω at 10 mA is likely.

We can now use the potential divider equation to determine the attenuation caused by the constant current sink and Zener string:

$$\frac{V_{out}}{V}(\mathrm{dB}) = 20\log\left(\frac{r_{lower}}{r_{lower} + r_{upper}}\right) = 20\log\left(\frac{400\ \Omega}{400\ \Omega + 30\ \mathrm{M}\Omega}\right) \geq 97\ \mathrm{dB}$$

97 dB is a very impressive attenuation, although it will vary considerably between different samples of DN2540N5s and improve substantially as current increases (Zener slope resistance falls with current whilst JFET g_m rises).

Nevertheless, such a satisfyingly large number means we need not determine it accurately, we can just assume that it will be 'good enough' and turn our attenuation to practicalities.

Thirty-five 5.6 V Zener diodes is a lot, and the knee-jerk reaction is to reject such a solution. However, a little more thought suggests that it is entirely practical, if a little unconventional. The primary disadvantage is not cost (Zeners are cheap) but the tedium of soldering all those diodes in series (and every one the right way round). Zigzagging the diodes backwards and forwards across a tag strip is ideal because despite the fact that the total wire length of all 35 Zeners on the author's tag strip ≈680 mm, and has a calculated ≈1 μH series inductance, its time constant of 5 ns is insignificant (see Figure 5.50).

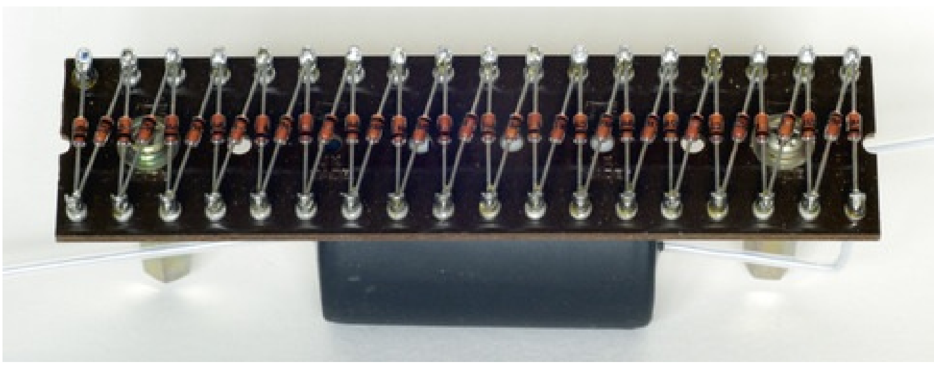


Figure 5.50 A composite Zener is easily constructed on tag strip.

Short-circuiting the output of the regulator puts the entire input voltage across the DN2540N5 cascode. Surprisingly, the author's prototype regulator survived this inadvertent abuse, but it is essential that the upper DN2540N5 has an adequate heatsink to cope with the short-circuit dissipation.

Shunt regulators also run the risk of self-immolation if their load is removed because this forces the shunt element to dissipate the design current plus load current. Both the BZX55 and the BZX79 series are rated at 500 mW, so for a 5.6 V diode that means a maximum current of 89 mA.

Bypassing the Composite Zener

Although the composite Zener produced very low noise measured over a 22 Hz to 22 kHz bandwidth, it produced 20 dB more when this filter was removed, which was a larger increase than initially expected, and the culprit is mains interference via the 12 pF output capacitance C_{OSS} of the DN2540N5. If we assume a slope resistance of 30 MΩ for the cascode constant current source (CCS) and 400 Ω slope resistance for the 195 V composite Zener, we can draw a diagram (see Figure 5.51).

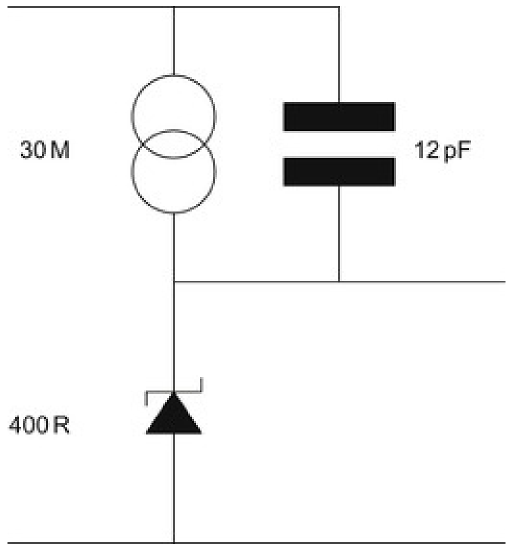


Figure 5.51 DN2540 C_{OSS} short-circuits the statistical regulator's CCS.

Looking at the diagram, we observe that C_{OSS} short-circuits the 30 MΩ slope resistance, and if we calculate their time constant (360 µs) this implies that mains noise rejection falls at 6 dB/octave from 442 Hz upwards. The solution is to convert the circuit into an oscilloscope probe by adding a 0.9 µF capacitor across the lower resistance to form another 360 µs time constant. Our potential divider now has equal time constants top and bottom, implying constant 97 dB attenuation with frequency. In practice, the 12 pF C_{OSS} estimate is likely to be unreliable, but the calculation gives us a useful lower bound to the required value of bypass capacitor, and 4.7 µF would ensure that attenuation did not deteriorate as frequency rose. Since the entire circuit is reasonably simple, it was possible to make a model including Zener and capacitor series inductance and resistance, and drive it from C_{OSS} in parallel with 30 MΩ to simulate the cascode CCS (see Figure 5.52).

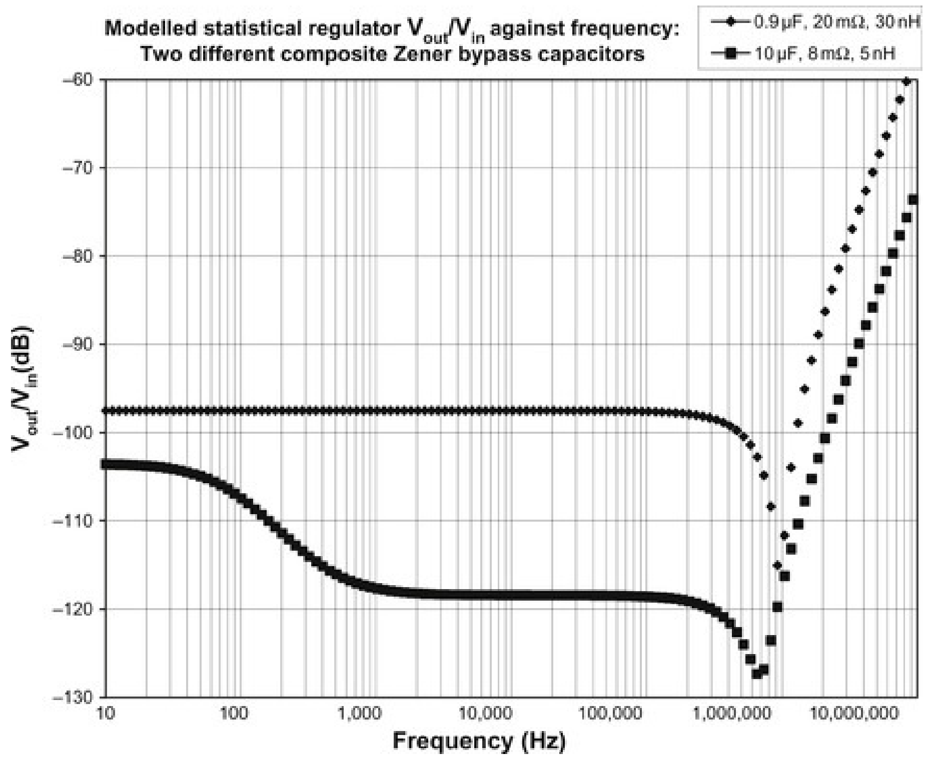


Figure 5.52 Modelled comparison of statistical regulator bypass capacitors.

As expected, the 0.9 µF capacitor allows constant attenuation with frequency, and the 1 µH series inductance (due to wire length) of the composite Zener is irrelevant. As with the *LC* filter, the value of bypass capacitance determines the depth of the attenuation plateau, its ESR nulls depth at the capacitor's self-resonant frequency, and its series inductance determines behaviour above self-resonance. The model compares a typical 0.9 µF plastic capacitor having ≈15 mm tails at each end with a 10 µF polypropylene capacitor having a four-wire Kelvin connection that renders tail inductance irrelevant. We now have a complete set of component values for our supply (see Figure 5.53).

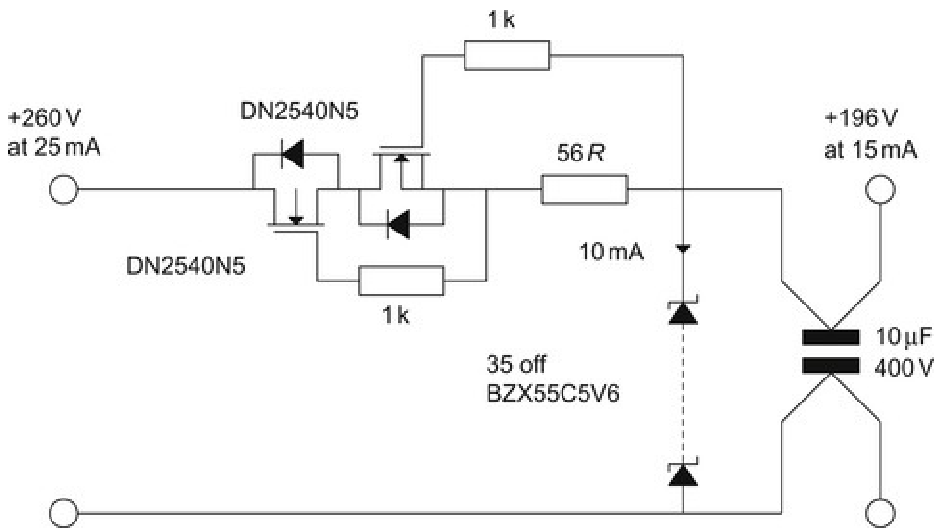


Figure 5.53 The complete statistical regulator.

As can be seen, the statistical regulator is very simple, yet measurements show that it is extremely good. Even better, it occupies little space so it can be positioned at the optimum point – adjacent to the load. Since it is best suited to pre-amplifiers, it is best to site the raw supply (transformer, rectifier, smoothing) remotely.

Optimising the Statistical Regulator

It is an engineering rule of thumb that Zeners should operate at a current of 10 mA or more (see Figure 5.54).

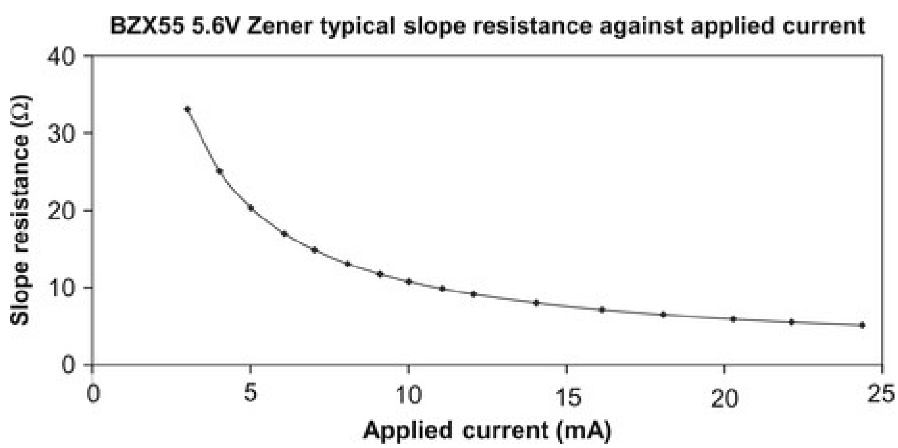


Figure 5.54 Typical BZX55 5.6 V Zener slope resistance against applied current.

The graph was produced by measuring Zener voltage against applied current for the composite 195 V Zener, fitting an equation of the form $V = a \cdot \ln(I) + b \cdot I$ to the curve, differentiating that equation to produce a slope resistance equation and dividing its coefficients by the number of individual Zeners. The graph is therefore an average of 35 BZX55 5.6 V Zeners from one batch, and although exact coefficients may vary from one batch to another, the general trend will remain. The slope resistance of the author's batch of BZX55 5.6 V Zeners could be found from:

$$r_{\text{slope}(\Omega)} = \frac{96}{I_{\text{DC(mA)}}} + 1.2$$

Thus, slope resistance falls from 20 Ω at 5 mA to 11 Ω at 10 mA and 6 Ω at 20 mA. The 1.2 Ω constant is probably the resistance of the fine wires needed to connect to the silicon die.

The DN2540N5 is a power JFET, and its g_m at low currents (<10 mA) is quite poor, significantly reducing the slope resistance of the cascode CCS, so it should pass a minimum current of 20 mA.

We saw in Chapter 3 that the noise generated by a DC reference was inversely proportional to the square root of applied current. Thus, although 10 mA of Zener current is sufficient to obtain a usefully low Zener slope resistance, doubling to 20 mA reduces Zener noise voltage by 3 dB. Since the current rating of the composite Zener is a constant 89 mA (500 mW/5.6 V) irrespective of total voltage, 20 mA is perfectly tenable, leaving 69 mA for load current. It would be wasteful to sink 20 mA of Zener current just to provide 1 mA of load current, so this implies that the statistical regulator is best suited to load currents of between 20 mA and 70 mA.

References for Elevated Heater Supplies – the THINGY

An elevated LT supply is required for any circuit having a cathode significantly above 0 V because leakage currents from heater to cathode $R_{hk(hot)}$ otherwise develop a noise voltage across the AC resistance seen looking into the cathode.

There are two ways of reducing the leakage current and therefore noise voltage:

- Don't use valves with poor $R_{hk(hot)}$. A valve tester or purpose-made jig can be used to select good examples. Because the most common cause of poor $R_{hk(hot)}$ is fluff or dust contamination during construction, it can often be burnt off by increasing heater volts by 2/3 and monitoring $R_{hk(hot)}$ without drawing any anode current. The resistance will begin to fall, and the moment it stops changing, switch the heater off, and allow the heater to cool. With luck, when tested again, $R_{hk(hot)}$ will be significantly improved. Note that raising the heater voltage easily damages an oxide-coated cathode, but if the valve was unacceptable anyway, you have nothing to lose.

• If the DC voltage across the leaky insulation was zero, the leakage current would be zero, and so would the noise.

In this example, it has been assumed that there are two heater supplies: one for valves with cathodes at ≈0 V and the other for valves with cathodes at ≈130 V, so if we were to follow the RCA recommendation, we would require heaters elevated by +40 V and +170 V. Although these voltages will not supply any current, they need a reasonably low AC source resistance and adequate filtering (see Figure 5.55).

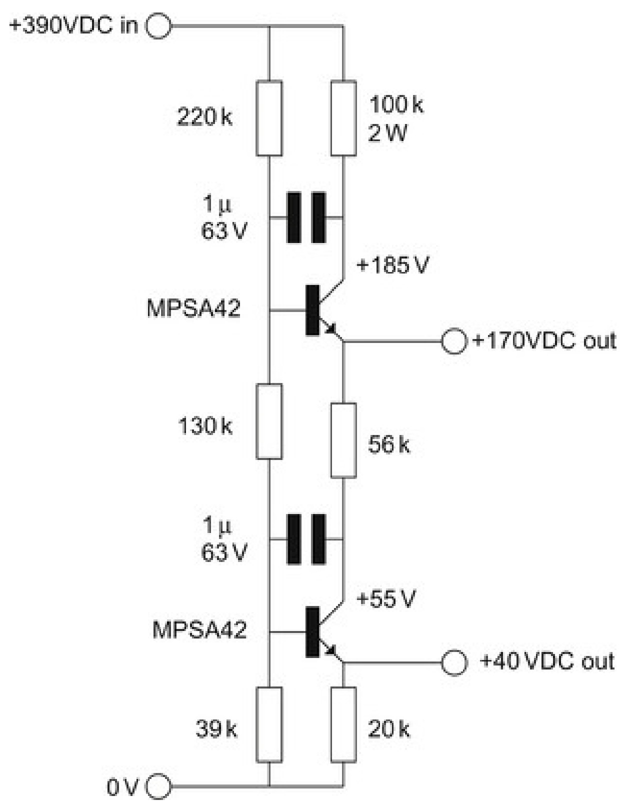


Figure 5.55 The THINGY, superimposing smooth DC on heater supplies.

The circuit is connected across the output of the HT supply, and it is a pair of emitter followers whose output voltage is set by a tapped potential divider. The circuit is utterly non-critical of component values and is easy to design/modify. Since we are dealing with reasonably high voltages, we will neglect V_{be} and consider the output voltages to be the same as the voltages at the tappings of the potential divider. If we neglect base current and arbitrarily set the current passing down the potential divider to 1 mA, then each resistor drops 1 V per 1 kΩ of resistance. Thus, if we want 40 V at the lower output, then 39 kΩ will be near enough for the lowest resistor. If the upper output is to be at 170 V, then the drop across the middle resistor is 170 V−40 V=130 V, and a 130 kΩ resistor will do nicely. If the HT voltage is 390 V, then the upper resistor must drop 390 V−170 V=220 V, so a 220 kΩ resistor is required.

Although the circuit only applies a potential to the external circuit, and does not source any current, each transistor must pass some collector current, but this current is not critical, and anywhere between 1 mA and 2 mA is fine. If we set I_c=2 mA, then the emitter resistor of the lowest transistor needs to be 40 V/2 mA=20 kΩ.

We could connect the collector of this transistor directly to the emitter of the upper transistor, but adding a collector load resistor improves the circuit's noise rejection and reduces power dissipation in the transistor. The resistor value is not in the least critical, but if we set V_{ce}=15 V for the lower transistor, then its collector must be at 40 V+15 V=55 V. The emitter of the upper transistor is at 170 V, so the voltage across the collector load resistor must be 170 V−55 V=115 V. Since the resistor passes 2 mA, its resistance must be 115 V/2 mA, and 56 kΩ is quite close enough. The advantage of including the collector load resistor is that it reduces V_{ce} (which reduces dissipation) and improves filtering.

The upper transistor also needs a collector load resistor. If we again assume V_{ce}=15 V, the collector of the upper transistor must be at 170 V+15 V=185 V. The HT is 390 V, so the upper collector load resistor drops 390 V−185 V=205 V. It passes 2 mA, so its resistance is 205 V/2 mA, and a 100 kΩ resistor will do nicely. 205 V across a 100 kΩ resistor dissipates 0.42 W, so a 2 W component is required.

Filtering is achieved by placing the filter capacitor not from base to ground, which would require a high voltage component, but from base to collector. For the lowest transistor, gain to the collector $A_v = -R_C/R_E$=56 kΩ/20 kΩ=−2.8, and the Miller effect therefore multiplies this capacitor by a factor of 3.8, so the effective value is 3.8 μF. Input resistance at the base of the transistors is approximately the Thévenin output resistance of the resistor chain, and the filter cut-off frequency is therefore 1.5 Hz. The lower emitter follower sees two cascaded 1.5 Hz filters, so noise is further rejected. The value of capacitance is not the least critical.

There is no reason to stop at just two outputs; extra output voltages can easily be derived by cascading more sections. Each section adds extra filtering, so you might choose to add a section just to improve noise rejection. Output resistance is less than 2 kΩ, although supplementing each transistor with another, to form a Darlington pair, would lower this output resistance.

The author was stumped for some time in attempting to name this circuit, but eventually realised that it is a Transistorised Heater Insulation Noise Grounding Yoke (THINGY), which is what people have been calling it anyway.

Note that *all* of the circuitry within an elevated LT supply is at least at the elevated voltage and that it therefore represents a shock hazard if touched. Even though the circuitry only contains components rated at a low voltage, elevated supplies should be treated with as much caution as HT supplies.

Common-Mode Interference

All of the previous discussion (rectification, smoothing/filtering, regulation) has been concerned with producing a source of DC power between two terminals with as little noise or interference as possible, and although not explicitly stated, this was differential-mode noise and interference. In this section, we will investigate common-mode interference where the difference in voltage between the supply's terminals may well be zero but both terminals are bouncing up and down in unison with respect to earth.

Common-mode interference often passes unnoticed. The problem becomes apparent when not only is common-mode interference present, but there is a mechanism for converting it into differential-mode.

Heaters and History

Golden age pre-amplifiers had AC heaters and suffered hum by modern standards. Sometimes this was due to leaky heater/cathode insulation, but mostly it was due to capacitance between heater wiring and signal wiring. The first solution was to centre tap the heater supply so that the heater wires had equal and opposite voltages, then to tightly twist the heater wiring. In this way, although there was capacitance from each heater wire to a given sensitive point, provided the twist was tight enough the two capacitances were equal (balanced), so the interfering currents cancelled. Because there's a practical limit to how tight the twist can be, the two capacitances cannot be perfectly equal, but provided that capacitances were minimised by pushing heater wiring into the corners of the chassis and only bringing it up to the valve at the last moment, the imbalance capacitance that caused hum was minimised.

The next step was to replace AC heater supplies with DC, and this produced an immediate measurable improvement in 50-Hz or 60-Hz hum that was often essential in microphone amplifiers. Remembering that the coupling mechanism to the signal electronics is capacitive, high frequencies will be coupled more easily, and rectification produces a spray of high-order harmonics of mains frequency. DC heater supplies need careful filtering if they are not to substitute a rattling buzz for the original hum; a simple rectifier and large reservoir capacitor are not sufficient.

Filtering low-voltage, high-current supplies was expensive in the 1950s, so the problem was eased by a low current heater variant of the input valve (UF86 needs 100 mA at 12.6 V as opposed to the functionally identical EF86 needing 6.3 V at 200 mA). A tape recorder of sufficient quality to require a low-hum

microphone amplifier probably had a push–pull loudspeaker amplifier that drew ≈100 mA from its HT supply, so a common trick was to power the UF86 from the HT supply return. The rarer, more expensive solution was to provide a dedicated heater supply with CRC filtering.
Any three-terminal IC regulator added after a rectifier and reservoir capacitor produces a close enough approximation to DC that differential-mode interference from heater wiring is eliminated.

How Common-Mode Heater Interference Enters the Audio Signal

Common-mode heater interference is a problem for small-signal valves because the interference current is coupled from the heater directly to the enclosing cathode via C_{hk} (and R_{hk} if it is poor). The interference current develops a noise voltage across the cathode impedance $r_k // Z_k$, which appears as a signal at the cathode, is summed with the wanted signal and amplified by the valve. There are two distinct scenarios:

- In a single-ended stage, the cathode will be (should be) coupled to ground via a low AC impedance Z_k. This low impedance might be the slope resistance of an LED or a large capacitor. But neither component can bond the cathode emissive surface directly to earth, so the non-zero length of the wiring has inductance which reduces coupling at RF, as does diode slope resistance, capacitor ESR and wiring resistance.
- In a differential pair, the cathode unavoidably has quite a high resistance to ground (via the anode load resistors divided by the μ of the valve – not the cathode resistance) and cannot form a useful CR filter in conjunction with C_{hk}. We are forced to rely on the (usually quite poor) RF balance of the differential pair to reject RF, so a differential pair is likely to be more sensitive to heater-borne interference than a single-ended stage.

Mains Transformers and InterWinding Capacitance

Since we cannot remove the path within the valve for coupling common-mode interference from the heater supply to the audio signal, we must prevent common-mode interference from reaching the heater supply.
A typical mains transformer has 1 nF of capacitance between adjacent windings, so rather than a spike being stepped up or down by the winding ratio, it is coupled unattenuated via this capacitance. Thus, a 1 V voltage spike that is an insignificant proportion of the 300 V HT can be coupled to a 6.3 V winding

where it is very significant. HT rectifier diodes inevitably generate spikes when they switch, and the author has observed HT switching spikes on the 6.3 V_{AC} LT supply of a Leak Stereo 20 chassis due to interwinding capacitance in the shared mains transformer.
Sensitive heater supplies require a separate transformer from high-voltage rectification (HT supplies, HT rectifier heaters and bias supplies).

Reducing Transformer InterWinding Capacitance

The capacitance of a parallel plate capacitor is directly proportional to the area of the plates and inversely proportional to the distance separating them, so if we could reduce the area between the primary and secondary and increase their separation, we would reduce the interwinding capacitance. (The effect of attempting to adjust the dielectric constant would be minimal.)
As an example, the Maplin 100 VA EI transformer kit is of split bobbin construction with the primary wound in one half of the bobbin and the other half left empty for the user to wind the secondary. The area of the interwinding capacitor on a split bobbin transformer is simply the area of the divider between primary and secondary less a little at the corners because the copper windings tend to curve as they bend round each corner (see Figure 5.56).

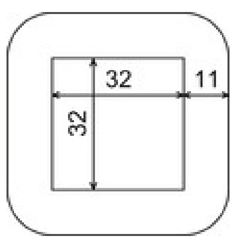


Figure 5.56 Interwinding capacitance in a split bobbin transformer.

The capacitive area of the divider is ≈1,800 mm ^{2}. Conversely, if the divider was to be removed and the bobbin filled with layer-wound windings, the area between the primary and secondary would be ≈7,200 mm ^{2}, increasing the capacitance by a factor of 7,200/1,800=4.
Turning to the thickness between the plates, a typical polyester transformer insulation tape might be 0.055 mm thick, but although a single thickness would theoretically be rated at 3 kV breakdown voltage, a transformer manufacturer would be aware that practicalities could require four layers of this tape between the primary and secondary layers to guarantee adequate dielectric strength to adequately insulate mains voltages, giving a total interwinding thickness of 0.22 mm. Conversely, the divider in the split bobbin transformer is 1.07 mm thick,

so it reduces capacitance by a factor of 1.07/0.22=4.9.
Taken together, the reduced plate area and increased separation of the example split bobbin transformer reduce interwinding capacitance by a factor of 4×4.9≈19.5, implying an interference reduction of ≈26 dB.
Before mains-borne interference can reach the interwinding capacitance, it must pass through the transformer's leakage inductance, so these two components form a resonant network. If we assume (for the purposes of comparison) that the resulting current develops a voltage across a 1 Ω earth bond resistance between the heater winding and the chassis, we can model the three main transformer constructions. The author measured the leakage inductance and primary to secondary interwinding capacitance of three commercial 100 VA transformers, and then modelled how well common-mode interference transferred to the resistance (see Figure 5.57).

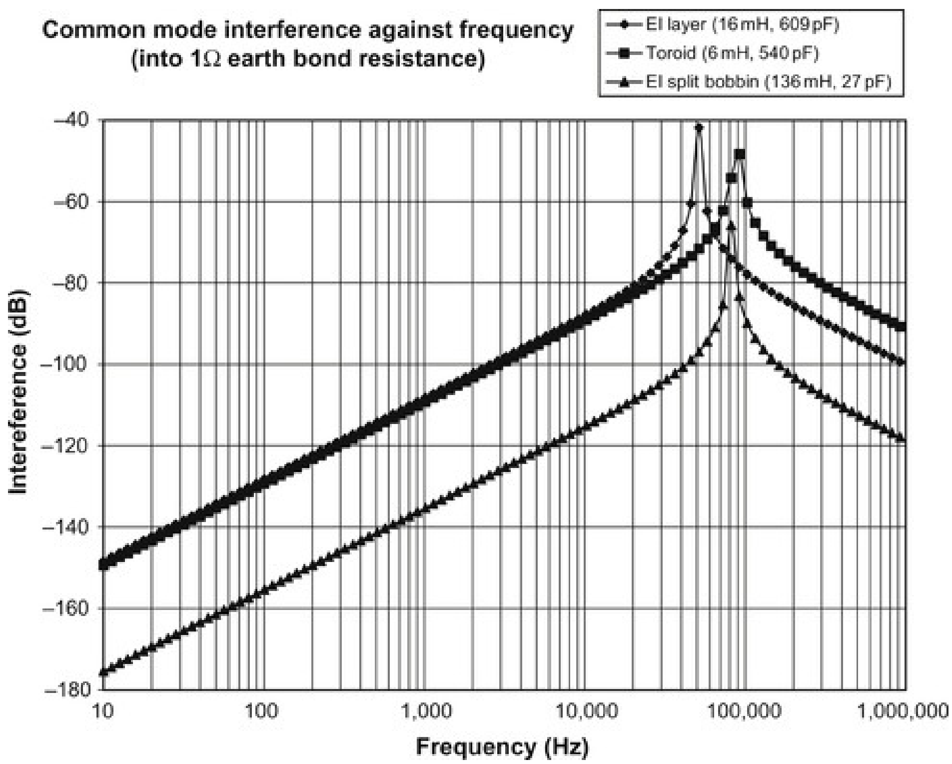


Figure 5.57 Modelled mains interference transmission for the three dominant mains transformer constructions.

Surprisingly, the large differences in leakage inductance are immaterial, and interference at audio frequencies is determined purely by interwinding capacitance, with the split bobbin example passing 27 dB less interference than the layer wound or toroid.

Post-Transformer Filtering

We have seen that segregating HT and heater transformers and choosing split bobbin construction (mandatory in Australia) reduce common-mode interference but further filtering may be necessary.

Even if a 6.3 V heater winding already has a centre tap, we can add a very cheap LR filter simply by connecting a 47 Ω resistor from each end of the winding to chassis – the filter uses the leakage inductance of the transformer.

We can make an explicit common-mode filter by adding series inductance to each leg of the heater supply and capacitance from each leg to the chassis. Since we are trying to filter common-mode noise, rather than differential-mode, we wind a bifilar choke on a small ferrite core and do not worry about core saturation because the currents in the two coils create equal and opposite fields, which cancel, so the core does not see any net magnetisation. Commercially made common-mode chokes are readily available.

Practical Issues

Although we now have enough theory to determine component values and ratings in each of a power supply's individual blocks, there are some practical considerations that should be addressed before we design a complete supply.

Transformer Regulation

Transformer manufacturers quote secondary voltage at rated current and also regulation as a percentage:

$$\text{Regulation }(\%) = \frac{V_{\text{off-load}}}{V_{\text{rated current}}} - 1$$

A typical 50 VA transformer might have a regulation of 13%, so a 6 V 50 VA transformer could be expected to produce 1.13×6 V=6.78 V off-load, and this is the voltage PSUD2 needs.

Perhaps more significantly, if we need 6.3 V, we can achieve it by under-running a 6 V transformer. We start by rearranging the previous equation:

$$V_{\text{rated current}} = \frac{V_{\text{off-load}}}{1 + (\text{regulation }(\%)/100)}$$

We now say that rather than using the entire current rating, we will use a fraction of it:

$$V_{\text{required current}} = \frac{V_{\text{off-load}}}{1 + (\text{current fraction} \times \text{regulation }(\%)/100)}$$

Thus, for our 50-VA 6-V example, the full-load current would be 8.33 A, but if we only needed 4 A, our current fraction would be 4 A/8.33 A=0.48, so:

$$V_{\text{required current}} = \frac{6.78\text{ V}}{1 + (0.48 \times 13/100)} = 6.38\text{ V}$$

Thus, we can achieve 6.3 V from a standard 6 V transformer and have 80 mV in hand for voltage drop across the heater wiring. This trick of using an oversized standard transformer can be a very useful way of avoiding the expense of a custom-wound transformer.

HT Capacitors and Voltage Ratings

The capacitors for 300 V HT supplies are easily obtained because the switched mode supplies in computers need 385 V capacitors to cope with the 340 V resulting from rectifying 240 V mains. But if we need a higher HT voltage, perhaps 430 V for a pair of EL34s, then a 450 V rated capacitor would be overstressed if mains voltage rose by 6% (as it is legally allowed to do). There are two choices: we either use a higher voltage capacitor, which will usually be paper or plastic film and generally only available in quite low values, or connect ‘ *n*’ *equal-value* electrolytic capacitors in series to obtain a composite capacitor having a total voltage rating multiplied by ‘ *n*’ but total capacitance divided by ‘ *n*’.

Because the capacitors are connected in series, the current passing through the capacitors must be equal, so each capacitor receives an identical charge ($Q = It$).

If their capacitances are equal, then the voltage across each one of them must be equal ($Q = CV$).

Unfortunately, even if the capacitances are equal, the leakage currents in each individual electrolytic capacitor are unlikely to be equal, so the voltage across each capacitor will not be equal. To equalise the voltages, and prevent one capacitor from exceeding its rated voltage, each capacitor should be bypassed by a resistor so that the resulting potential divider chain forces the voltages to be equal (see Figure 5.58).

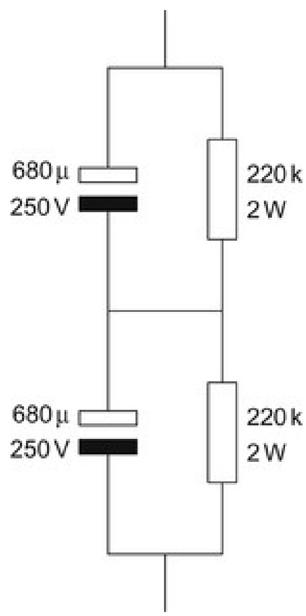


Figure 5.58 Bleeder resistors equalise capacitor potentials.

The divider chain should pass at least 10 times the expected leakage current of the capacitors to ensure correct operation. Typically, a 220 kΩ 2 W resistor suffices.

A technically better method is to use separate HT windings and rectifier/smoothing circuits, and place the resulting floating DC outputs in series to obtain the required HT voltage. This ensures that each capacitor cannot exceed its rated voltage, but the mains transformer is now more complex (read expensive) (see Figure 5.59).

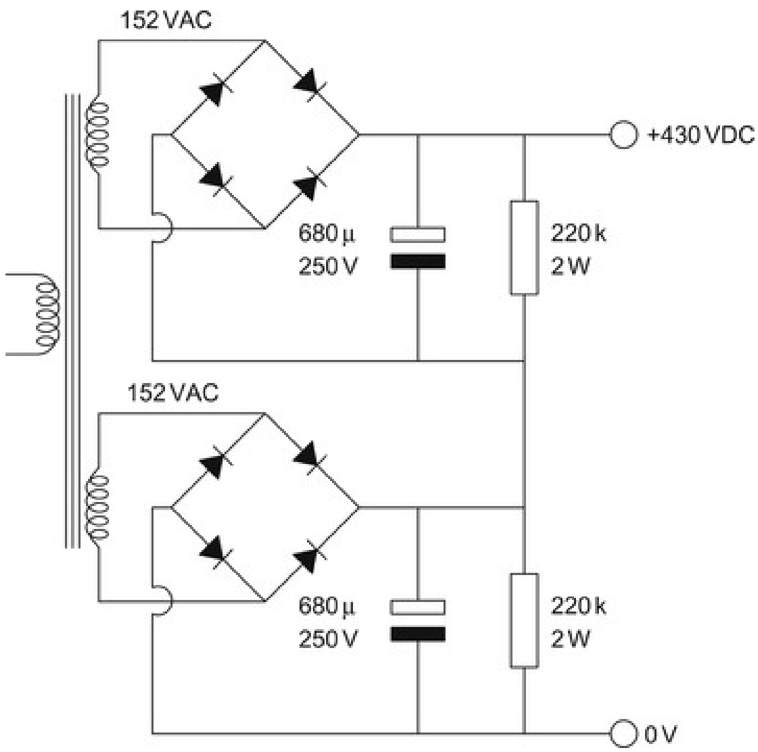


Figure 5.59 Achieving HT>340 VDC with electrolytic capacitors.

Can Potentials and Undischarged HT Capacitors

Both of the previous schemes for producing a composite HT capacitor of high voltage rating resulted in one capacitor with its negative terminal stood away from ground potential. This is significant because the can of an electrolytic capacitor is connected either to the negative terminal or at a potential very close to it. Cans at an elevated voltage must not only be insulated from the chassis (flanged plastic capacitor clamps that prevent the can touching the chassis are therefore best), but the capacitor must also be properly insulated from the user.

HT supplies represent a formidable shock hazard, and it is essential that provision is made for fully discharging the reservoir and smoothing capacitors when the equipment is switched off. The HT supply therefore needs a purely resistive discharge path to 0 V at some point, and the simplest way of providing this is to connect a 220 kΩ 2 W resistor across the reservoir electrolytic, which not only discharges the capacitor, but also (provided that there is a return path) discharges subsequent HT capacitors.

The Switch-On Surge

If we do not use a valve rectifier, the HT switches on instantly, and suddenly applying 400 V to an electrolytic capacitor stresses both dielectric and electrolyte, so we should look to see if there is some way of prolonging life. (Given that computers typically last only five years before being consigned to landfill, physical longevity of their power supplies is scarcely an issue, so they don't worry much about switch-on transients.)

When we apply rectified AC to the reservoir capacitor, we may be unlucky enough to switch at the instant that the cycle is at its peak voltage. The instantaneous transition from 0 V to 325 V ($dV/dt \approx \infty$) applied to the capacitor causes a theoretically infinite current to flow because:

$$I = C \cdot \frac{dV}{dt}$$

If, however, we could always switch at the zero-voltage point, then although dV/dt for a sine wave is at a maximum at this point, it is not infinite, and the inrush current is reduced. Devices capable of performing this switching are known, predictably, as zero-voltage-switching relays and the Crystal Palace amplifier in Chapter 6 uses one to soften the inrush to its main HT supply.

However, the author is beginning to harbour suspicions that these relays might cause switching spikes as the mains crosses 0 V and switching is handed between the two back-to-back thyristors that make up the internal triac, and he's certainly seen some odd effects on logic lines as a zero-voltage-switching relay switched an inductive load, so in low-noise applications perhaps it's time to revert to hard vacuum rectifiers and simply switch mains to the HT transformer (that also has the rectifier heater winding) using a conventional switch or electromechanical relay.

Mains Fusing

Although UK mains plugs have a fuse in the line, it's there to protect the lead, not the equipment.

The first component the line should meet when it enters a piece of equipment is an appropriately sized fuse that protects against fire in the event of an over-current fault within that equipment. If we use one piece of equipment to switch and distribute mains to other equipment, we must fuse its internal electronics separately from the mains distribution. 230 V appliances in North America require double-pole fusing (a fuse in each line) because the 230 V is actually a two-phase system of 115 V–0–115 V (centre tap to earth), so although a single fuse would protect against an over-current fault from line to line, each line must be protected against line to earth faults.

Fuses are manufactured to be 'fast' (F) or 'timed' (T). 'Timed' used to be called 'anti-surge', and these fuses are intended to survive a short inrush current considerably higher than their sustained current rating. Almost all electronic equipment draws an inrush current at switch-on as reservoir capacitors are charged or toroidal transformers kick, so we always need 'timed' fuses, but how do we determine their rating?

Fuse manufacturers' data sheets have plenty of graphs that refer to I^2t ratings. What they're pointing out is that if a fuse wire has a certain mass, it must require a defined quantity of heat energy to raise its temperature to melting point so that it ruptures. Recall that power= I^2R and that energy= Pt, and I^2t ratings suddenly make sense. At this point, the hopeful author reached for his Hall effect current probe, used his digital oscilloscope in 'single sequence' mode to capture a mains inrush current at switch-on, set the oscilloscope's 'math' function to Ch1×Ch1 (to give I^2), positioned cursors at the beginning and end of the pulse, gated a 'mean' measurement on 'math' by the cursors, and then manually multiplied that mean value by the time between the cursors to arrive at an accurately measured value of I^2t.

Unfortunately, further reading of fuse data sheet reveals what we already knew subliminally. Fuses aren't terribly accurate, and as their I^2t rating falls closer to the measured value, we're more and more likely to suffer nuisance blowing. Understandably, fuse manufacturers express the situation more formally in terms of probabilities, but their advice amounts to: Guess/measure an I^2t rating, try a fuse of that rating and if it blows too often without there being a fault, try the next current rating up.

Mains Switching

Mains switching is similar to mains fusing but may have a complication added by the incoming mains connector.

For a single phase system, a single switch in the line conductor is safest provided that connector polarity is unambiguous (IEC mains connector), but if the unpolarised two pin figure-of-eight small appliance connector is used, double pole switching (one switch in line and one in neutral) is necessary to ensure that no matter which way round the connector is inserted, line passes no further than the switch when in its 'off' position. Multiple phase systems (North American 230 V, German 380 V) are unlikely to be encountered in audio but would require switching of each phase.

It might be thought that a single phase system should be double pole switched irrespective of mains connector, but consider a piece of equipment having an earthed metal chassis, powered via an IEC connector, double pole switched, and imagine that the switch in the neutral fails open circuit. The equipment loses power and superficially appears safe for investigation to determine the fault. However, line is still present on the mains transformers and there is a very real danger of the investigator suffering a shock from line to the earthed metal chassis.

The easiest way to ensure safety of mains fusing and switching is to use an IEC inlet having an integral fuse and switch.

A relay allows remote mains switching. Thus, a low voltage switch on a control unit could energise a remote relay to switch mains. Unfortunately, that implies that we need permanently applied standby power in order to be able to switch the relay. The author was previously perfectly happy to leave RIAA-stage heaters permanently heated in standby mode, thus also providing power for the relay, but recent electricity price rises have changed that. Nevertheless, you might feel that the convenience of a relay outweighs the cost.

There’s a nasty little problem to be aware of in some relays. At its simplest, an electromechanical relay is just a coil of wire round a soft magnetic core that attracts a hinged soft magnetic material known as an armature to provide the switching action. A simpler (cheaper) relay can be made by passing the switching current through the armature, so when we switch our mains, we connect the relay’s (electrically conductive) core to 240 V. The core’s surrounding coil is wound on an insulating plastic former, so we have just connected a significant capacitance from one side of our heater supply to the mains – negating all our design efforts at reducing common-mode noise. Specify a relay withstand voltage of 10 kV, and the problem is unlikely to occur. The relays to be suspicious of have a low profile rather than being square when viewed from the side. If in doubt, reach for a hacksaw and sacrifice one to science to determine its exact construction before building that type in and having to diagnose the problem later.

A Practical Design

Having investigated individual blocks, we are now in a position to be able to design a complete HT and heater power supply. An RIAA stage imposes the most demanding power supply requirements, so we will design two variants of a power supply for this use; we can then add/discard/modify blocks as necessary for other applications.

The first variant is for the balanced hybrid RIAA stage of Chapter 7 that requires an HT of 195 V at 48 mA and a single heater supply of 6.3 V at 1.2 A. We will need:

- HT regulation
- HT rectification and smoothing
- Heater rectification and smoothing
- Heater regulation
- Mains filtering.

We can now draw a block diagram (see Figure 5.60).

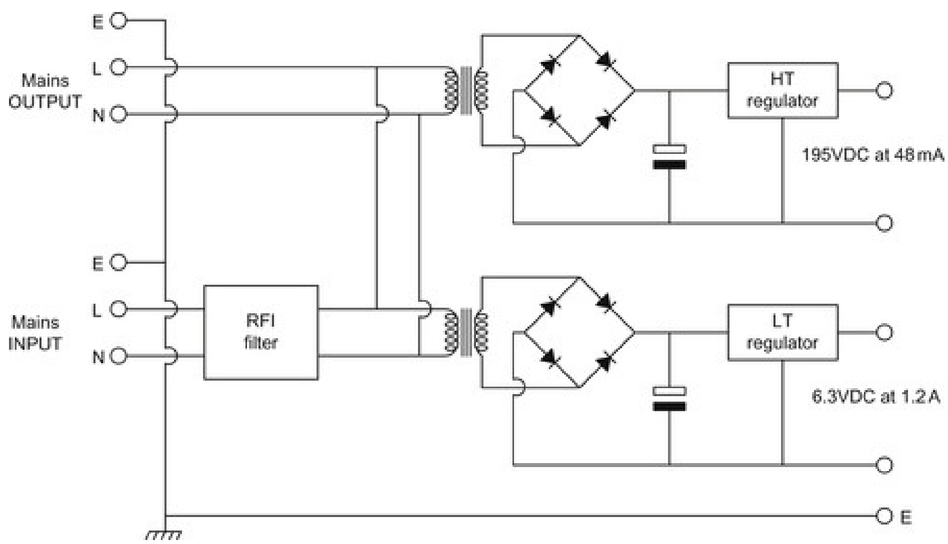


Figure 5.60 Preliminary block diagram of power supply.

HT Regulation

There are three possibilities:

- Valve regulator

• Maida (317) regulator

• Statistical regulator.

Valve regulators tend to be big. Techno-pretty, but big. An optimised valve regulator is bigger, prettier (even more glowing glass) and needs lots of heater supplies for all that glass. With care, an optimised valve regulator can achieve noise ≈1 mV $_{pk–pk}$. The Maida regulator is small, doesn't need heater supplies, but is electrically fragile, and produces noise comparable with a valve regulator but without the DC drift. The statistical regulator is slightly larger than the Maida regulator, but far quieter and more robust than the author expected.

The balanced hybrid RIAA stage uses a differential cascode input stage having limited PSRR, so the low noise of the statistical regulator makes it the natural choice (see Figure 5.61).

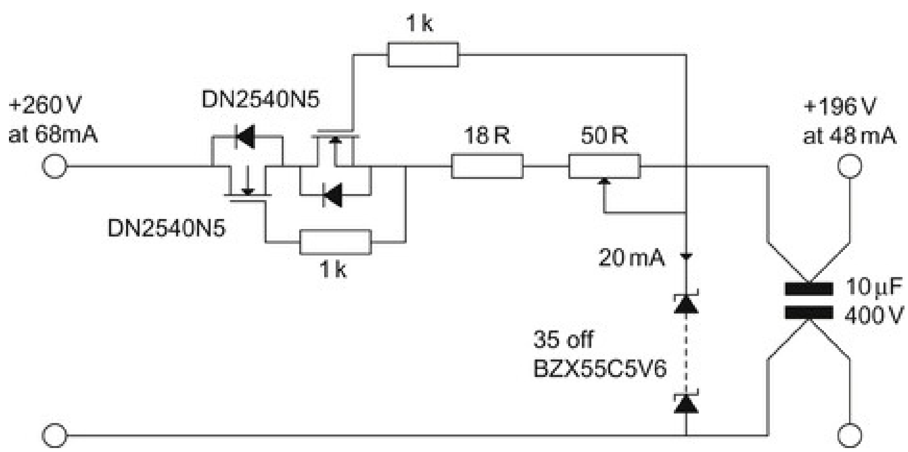


Figure 5.61 The statistical regulator configured for 195 V at 46 mA.

The cascode DN2540N5 CCS must be programmed to pass the load current plus the composite Zener current. The composite Zener could operate at 10 mA, but slope resistance and noise are significantly lower at 20 mA, so the CCS needs to be programmed for 68 mA. Field Effect Transistor (FET) device variation means that the required source programming resistance varies hugely, so a 50 Ω variable resistor is required – the 18 Ω resistor prevents accidental setting of excessive current. Although the programmed 68 mA current passes through both DN2540N5s, only the upper one has an appreciable voltage across it, so it will require a heatsink to dissipate 4–5 W, but we won't know the exact power to be dissipated until we have designed our rectification and smoothing.

We know that we must use 5.6 V Zeners, so 195 V/5.6 V means that we need 35 of them in our composite Zener. Under normal conditions, the composite Zener passes 20 mA, but in the event of a catastrophic fault causing the entire load to be disconnected, the constant current source will drive its entire programming current of 68 mA down the composite Zener. Since each Zener is rated at 500 mW, it can pass a maximum current of 0.5 W/5.6 V=89 mA, so a 68 mA fault current is tolerable.

We saw earlier that the composite Zener requires a bypass capacitor to set a time constant of at least 360 μs in conjunction with its slope resistance, so if we assume r_{slope}=6 Ω per Zener at 20 mA, then the composite Zener's slope resistance will be 210 Ω, and 360 μs would be achieved by 1.7 μF, but because of the likely variability of that 360 μs time constant, it would be safer to use 10 μF. The capacitor has 195 V across it, so a 250 V component would be fine, but the author had a batch of 10 μF 400 V Kelvin capacitors, so this is what he used.

HT Rectification and Smoothing (a PSUD2 Exercise)

Although the statistical regulator has excellent low frequency ripple rejection, rejection of interference >100 kHz is determined primarily by the inductance of its bypass capacitor which will be resonant at <1 MHz. Thus, it makes sense to choose a post-rectification filtering scheme that maintains its filtering to as high a frequency as possible, and an *LC* filter using a capacitor having a Kelvin connection is ideal. Having tested a prototype and proven the principle, the author had previously bought a batch of ten 100 μF 400 V metallised polypropylene Kelvin connection capacitors, so this is what he used.

The statistical regulator requires an absolute minimum of 10 V across its CCS before it will operate, but >20 V to ensure that C_{OSS} in the upper device falls to its asymptotic value of 12 pF. Thus, we need 195 V+20 V=215 V leaving the filter, but to allow correct operation in the face of mains voltage variation, it is wise to increase this by ≈10%, bringing it to 240 V.

Assuming a PSRR of 40 dB (due to the differential pair) at the input stage of the balanced hybrid RIAA stage, an 80 dB signal to hum ratio requires ripple on the 195 V HT to be <2.7 mV $_{pk–pk}$. The statistical regulator will certainly have >90 dB ripple attenuation, so this corresponds to tolerating 85 V $_{pk–pk}$ ripple leaving the *LC* filter, implying that filter design is not critical. In practice, 85 V $_{pk–pk}$ of ripple reaching the regulator would require raising its peak input DC voltage by 85 V from 240 V to 325 V to avoid drop-out, so it would be better to require that the *LC* filter limit ripple to <10 V $_{pk–pk}$. Thus, our *LC* filter must deliver 250 V with <10 V $_{pk–pk}$ ripple at a current of 68 mA, so we now need to switch on the computer and invoke PSUD2.

Since we need <100 mA, a hard vacuum rectifier is not only a tenable design choice, but also the most sensible, and 68 mA suggests the very cheap EZ80. We definitely want full-wave rectification, but a bridge rectifier with valves is awkward, so we should simulate using PSUD2's 'full-wave' option, and if we don't have a centre-tapped mains transformer of the required voltage, we simply make a hybrid bridge rectifier using the EZ80 and a pair of STTA512Fs.

Pleasingly, you will discover that having already chosen the 100 μF capacitor and EZ80 rectifier, it is very hard to design a 250 V 68 mA power supply incorporating an *LC* filter that produces as much as 10 V_{pk-pk} ripple. In practice, the main design limitation turns out to be preventing the choke's peak current from exceeding its rating, so make sure you monitor I[L1] as well as the output voltage. Remember also that the transformer voltage PSUD2 requires is not the manufacturer's rated full load voltage but the open circuit voltage.

The author had a 260–0–260 V 100 mA, 6.3 V 2 A Admiralty pattern mains transformer and a matching Admiralty pattern 20 H 120 mA choke, so the problem was to persuade the combination to produce the required DC voltage.

The open circuit voltage of the transformer from the 0 V to each 260 V tap was measured to be 282 V_{RMS}. Quite apart from the obvious safety considerations, this measurement needs to be made quite carefully – it's important to know that the transformer's primary is actually receiving its expected voltage. Suppose that the transformer's primary was stated to be 240 V but mains that day was 245 V; this would mean that the secondary voltage would be falsely 2% high. If you have two meters, use them and correct for mains voltage variation.

The DC resistance from 0 V to each 260 V pin was measured. One was 99 Ω and the other 89 Ω, and this disparity is common in layer-wound transformers because it is cheaper to layer-wind a centre-tapped transformer with one half of the winding on top of the other, so the average diameter of the outer winding is a little larger than that of the inner winding, resulting in a slightly higher copper resistance. Unless balanced by adding an external resistance to the inner winding, a ripple component at mains frequency appears at the output of the rectifier, which is not particularly well attenuated by an *LC* filter. Thus, we should add a 10 Ω resistor in series with the 89 Ω winding to make both 99 Ω.

PSUD2 requires transformer output resistance including reflected primary resistance. Primary resistance was 24 Ω, and since the open circuit voltage was 282 V_{RMS} when the input was 240 V_{RMS}, the turns ratio is 282/240=1.175. Impedances are transformed by the square of the turns ratio, so the reflected primary resistance is $24 \times 1.175^2 = 33$ Ω, and we add this to the 99 Ω secondary resistance to give 132 Ω, and put this value into PSUD2.

The DC resistance of the choke was 367 Ω, and the ESR of the 100 μF polypropylene capacitor is 6 mΩ, so these two values were also entered into PSUD2. The load was set to 'constant current' and 'stepped load' starting at 10 mA, and then 68 mA after 1 s. The simulation was set to run for 2,000 ms after a delay of 0 s. In the 'Options' menu, 'Allow warnings', 'Auto simulate', 'Dual axis' and 'Soft start' were all ticked. Finally, V[C1] was ticked in the result column (see Figure 5.62).

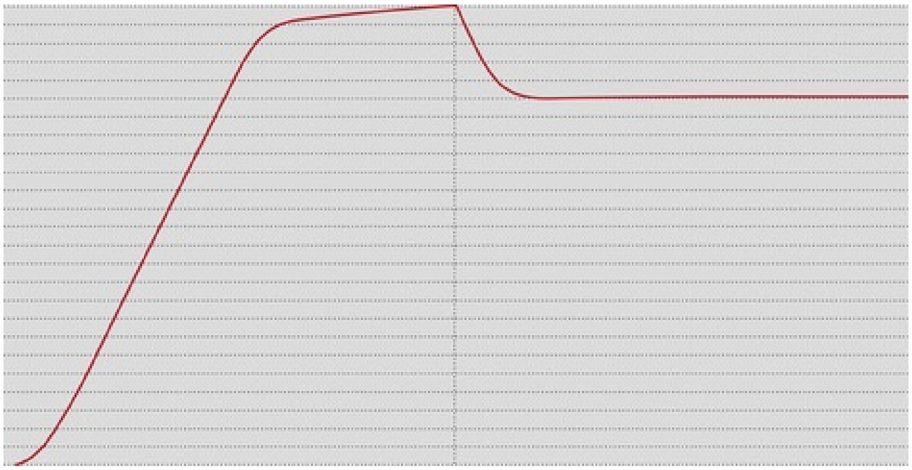


Figure 5.62 PSUD2 simulation shows insufficient voltage and a hint of low frequency undershoot.

As you will find if you run the simulation, not only could a choke input supply not achieve the required 250 V at 68 mA, but there was a very slight undershoot when the stepped current load changed from 10 mA to 68 mA (temporarily reduce choke resistance from 362 Ω to 62 Ω to see this effect more clearly). A capacitor input supply would certainly produce too much voltage (√2×282=399 V), so we need an intermediate supply.

We highlight the entire *LC* filter, right click and 'insert' a 'C' filter. PSUD2 automatically sets any new capacitor to be the same as the existing one, so we edit C1 to 5 μF; a 5 μF film capacitor is likely to have an ESR of ≈10 mΩ, so we can edit this value if we wish (it doesn't actually make any difference). We now change the result to V[C2] and find that the overshoot has disappeared, but that we have a higher voltage (310 V) than we want on C2. We edit the value of C1 until we obtain 250 V, requiring 1.8 μF. Unfortunately, 2.2 μF is the nearest E6 standard value and this produces 260 V, which is not ideal because it increases dissipation of the CCS's upper DN2540N5 from 3.6 W to 4.3 W, but this is still acceptable.

Now that we have the required output voltage, we can check internal voltages and currents:

• I_L peaks at 83 mA 0.5 s after switch-on but operates at 76 mA_{pk} during normal operation with the 68 mA DC load.

• V[C1] peaks at 383 V, so a 400 V component will not do − 630 V is needed, implying polypropylene dielectric.

• I[C1] peaks at +184 mA during normal operation, and this is no problem even for a metallised capacitor.

• It is quicker to check the fault value of V[C2] manually than use PSUD2, but try it anyway to prove it to yourself. V[C2] would rise to $282\times\sqrt{2}=399$ V with no load current, which is why a 400 V component was needed. (If the mains voltage was simultaneously high, the 400 V rating would be exceeded, but audio is not life-critical, so it is conventional to assume only one fault at a time.)

• I[D1] peaks at 253 mA, but increasing the load current by only 8–76 mA causes PSUD2 to flash up an over-current warning, so the EZ80 is marginal and an EZ81 might be better, although it would cause the output voltage to rise to 272 V.

Finally, we should change PSUD2's simulation time to 50 ms and its reporting delay to 5 s to investigate the ripple V[C2]. Subtracting the minimum voltage on the graph from the maximum, we find we have 205 $mV_{pk\text{–}pk}$ of ripple − 34 dB better than our 10 $V_{pk\text{–}pk}$ limit. The statistical regulator passes 20 mA through its composite Zener, implying a slope resistance of 200 Ω and therefore 103 dB ripple attenuation. In theory, the entire HT supply has the potential to produce its 195 V with only 500 nV_{RMS} ripple. In practice, what this means is that ripple will only be measurable if construction is flawed.

Now that the principles have been demonstrated, you can use a combination of the formulae given earlier in this chapter plus PSUD2 to design your own rectification and filtering that gives the required performance using components available to you.

Heater Rectification and Smoothing (a Manual Exercise)

When designing the HT supply, we started with the regulator and then progressed to rectification and smoothing. However, for the heater supplies, we know that we must use a split bobbin transformer in order to benefit from their greatly reduced inter-winding capacitance, but such transformers tend to have standard secondary voltages such as 6 V_{RMS} or 9 V_{RMS} and it is easier to test

standard transformer voltages for suitability than demand an exact voltage.
We ultimately need 6.3 V $_{DC}$ at 1.2 A, and this can be catered for by the 1.5 A 317. As before, regulator design begins with regulator drop-out voltage, and despite what the data sheets claim, a typical 317 needs >3 V across it to regulate cleanly. We therefore need a minimum voltage of 9.3 V before the regulator.
Assuming a capacitor input filter, 6 V $_{RMS}\times\sqrt{2}$=8.5 V $_{pk}$, which is insufficient, so we need a 9 V $_{RMS}$ secondary, which provides 12.7 V $_{pk}$. We will use a bridge rectifier, which always has two diodes in series, so it will drop ≈1.4 V across the rectifier, which brings the voltage down to 11.3 V.
If a rectified sine wave of 11.3 V $_{pk}$ leaves the rectifier, then this is the maximum voltage to which a reservoir capacitor of infinite capacitance could charge. A capacitor of finite capacitance will charge to this voltage on the peaks, but its minimum voltage must be:

$$V_{\text{minimum}} = V_{\text{peak}} - V_{\text{ripple}}$$

The absolute minimum voltage that we can allow is 9.3 V, so the maximum ripple voltage we can tolerate is 2 V $_{pk\text{–}pk}$.
Using our earlier equation that related ripple voltage to current:

$$C = \frac{It}{V} = \frac{1.2 \times 0.01}{2} = 6{,}000\ \mu\text{F}$$

The equation requires 6,000 μF, so we could use 6,800 μF, but this would not allow for any tolerance on capacitor value, or mains voltage variation, so 10,000 μF would be a safer choice, resulting in 1.5 V of ripple.
You could analyse this circuit in PSUD2, but you will find that entering the transformer resistance requires you to be able to measure a secondary resistance of the order of 40 mΩ, which can only be done reliably using a four-wire Kelvin connection (not available on most DMMs). Further, even if you are able to make such measurements, you will find that a ripple current of 9 A is predicted – much higher than the typical measurement of 6 A. Sometimes, manual calculation beats computer simulation.
If the bridge rectifier drops 1.4 V, and passes an average current of 1.2 A, then it must dissipate ≈1.7 W (this is a crude approximation because $I_{\text{average}} \neq I_{\text{RMS}}$, but we don't know the averaged ripple current I_{RMS} over one cycle). This is a significant amount of heat to be lost from the typical W02 1.5 A bridge rectifier package, so they invariably become very hot and eventually fail. It is thermally better either to use individual diodes such as the 3-A 1N54** series, or a 4 A bridge rectifier package.
An even better solution can be to use Schottky diodes, perhaps the 31DQ**

series, for the bridge rectifier. These have a slightly lower forward voltage drop, reducing diode dissipation, but the main justification for their use is that they switch off cleanly, without the current overshoot exhibited by junction diodes. As mentioned earlier, the overshoot is an impulse that excites resonances in the transformer/rectifier/reservoir capacitor system.

The output voltage from the rectifier is only 11.3 V_{pk}, so V_{RRM} for each diode need only be 12 V; 50 V is commonly the lowest available rating available and allows for mains spikes, so this will be fine.

Each diode in the bridge should be bypassed with a film capacitor; 100 nF 63VDC is a good choice, but almost anything will do, provided that the voltage rating> V_{RRM} for each diode.

Heater Regulation

We should first consider how much power the 317 must dissipate when the 1.2 A load current is drawn. Assuming 1.5 V_{pk-pk} ripple on the reservoir capacitor, the average voltage applied to the regulator is 10.6 V, so the voltage across the regulator is 4.3 V, and the regulator therefore dissipates ≈5 W. 5 W is a perfectly reasonable dissipation for the (20-W) TO-220 package of a 317T to dissipate provided it is thermally bonded to an aluminium chassis, so the older and much more expensive 317K TO-3 metal ‘power transistor’ package is unnecessary. The 1N4002 protects the regulator from reverse voltage.

The next step is to determine the value of resistors needed in the potential divider. Experience shows that the reference voltage tolerance on the 317 is actually very good, and that it is unnecessary to include a variable resistor to tweak the output voltage. You might have a different view on this, but an upper resistor of 180 Ω and a lower resistor of 750 Ω set 6.5 V that allows for the 0.2 V voltage drop down the typical 3 m loop length of 0.6 mm-diameter solid core twisted pair separating the regulator in the remote power supply from the valve pins.

The Thévenin resistance of the 180 Ω and 750 Ω combination is 145 Ω, so a 10 μF speed-up capacitor should be used to bypass the ADJ pin to ground. The manufacturers’ application notes recommend that the output of the 317 be bypassed to ground with a 1 μF tantalum bead capacitor via a 2.7 Ω resistor, but tantalum bead capacitors generally have sufficiently high ESR to make the explicit resistor unnecessary. If you decide to discard the 2.7 Ω resistor, either check the capacitor data sheet, or (best) measure the capacitor with an ESR meter.

We need a common-mode choke, and we could put it between the transformer

secondary and the rectifier, but this relies on critical balancing of the opposing magnetic fields due to the 6 A $_{pk}$ ripple current to avoid saturation, so it is better to put it between the reservoir capacitor and the regulator where it only passes 1.2 A $_{DC}$. The following 10 nF capacitors should have leads as short as possible to chassis. If you can match the 10 nF capacitors, so much the better for avoiding converting common-mode noise into differential-mode.
We can now draw a heater supply circuit diagram (see Figure 5.63).

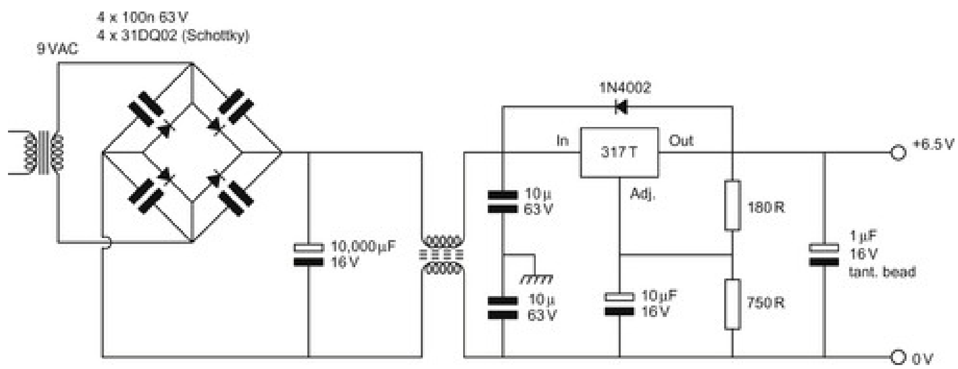


Figure 5.63 317 6.3 V 1.2 A heater supply.

Mains Filtering

The author has never been entirely convinced of the utility of mains filters, but they are a bit like a car seat belt – you wear it not because you expect to need it every day but because the day might come when you do.
Although the power consumed by electronic equipment may not be especially high, the ripple current (as we found earlier) can be much higher than the load current. Most commercial RFI filters are rated at 16 A or less, which may not be enough for audio. If we want an RFI filter, we probably need to make it ourselves, winding our own bifilar choke on a ferrite toroid (see Figure 5.64).

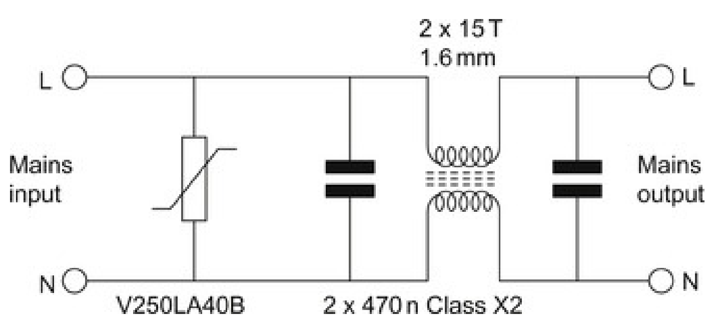


Figure 5.64 Mains filter.

A 130 J metal oxide varistor is connected across the mains to limit mains spikes. Manufacturers' data sheets for these devices disingenuously show that a 100 V spike riding on the peak of the mains waveform is clipped very effectively but don't show that a 340 V spike sitting on the 0 V crossing point would pass by unmolested. Still, the day might come when one is needed.

The series common-mode choke is terminated at each end by a pair of Class X2 capacitors. X2 capacitors are the only type of capacitors that may be legally connected between mains line and neutral (the reason is that they are specifically designed to fail safely). Many RFI filters also include a 4.7 nF Class Y capacitor from live to earth, and another from neutral to earth, but their utility is debatable for audio equipment as they can make the earth noisy.

If the HT mains transformer has an electrostatic screen, it should be connected directly to chassis. The heater transformer won't have an electrostatic screen because it is of split bobbin construction.

We can draw a full power supply circuit diagram (see Figure 5.65).

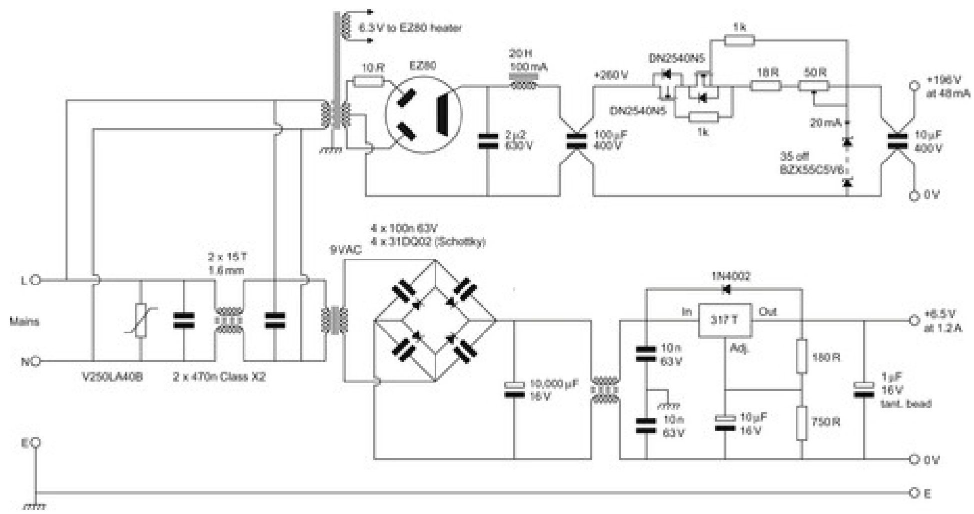


Figure 5.65 Complete power supply for balanced hybrid RIAA stage.

Adapting the Power Supply to the EC8010 RIAA Stage

The balanced hybrid RIAA stage required an HT of 195 V at 46 mA, but the EC8010 RIAA stage requires 390 V at 67 mA, and a pair of references (40 V, 270 V) at zero current. Rather than a single heater supply of 6.3 V at 1.2 A, the EC8010 RIAA stage needs four 300 mA at 12.6 V supplies tied to ≈40 V that can all come from one transformer secondary and a further two 300 mA at 12.6 V supplies tied to ≈270 V requiring another transformer secondary. The heater strings were configured in this way to enable use of a standard transformer having two secondaries of the same voltage. We will need:

- HT regulation
- Reference voltages
- HT rectification and smoothing
- Heater regulation
- Heater rectification and smoothing.

HT Regulation

We will again use the statistical regulator because of its low noise. 390 V might seem an alarming voltage to set from 5.6 V Zeners, but it's simply two 195 V composite Zeners in series (70×5.6 V Zeners). The increased current demand is potentially more of a problem as 67 mA of load current plus 20 mA Zener quiescent current requires 87 mA from the CCS, and if the load should fail, the CCS would sink all 87 mA into the composite Zener. The author was nervous that 87 mA was far too close to the 89 mA maximum rating, so he simulated the inside of an RIAA stage by covering a single Zener with a sheet of cardboard to prevent proper cooling, and then tested it (see Figure 5.66).

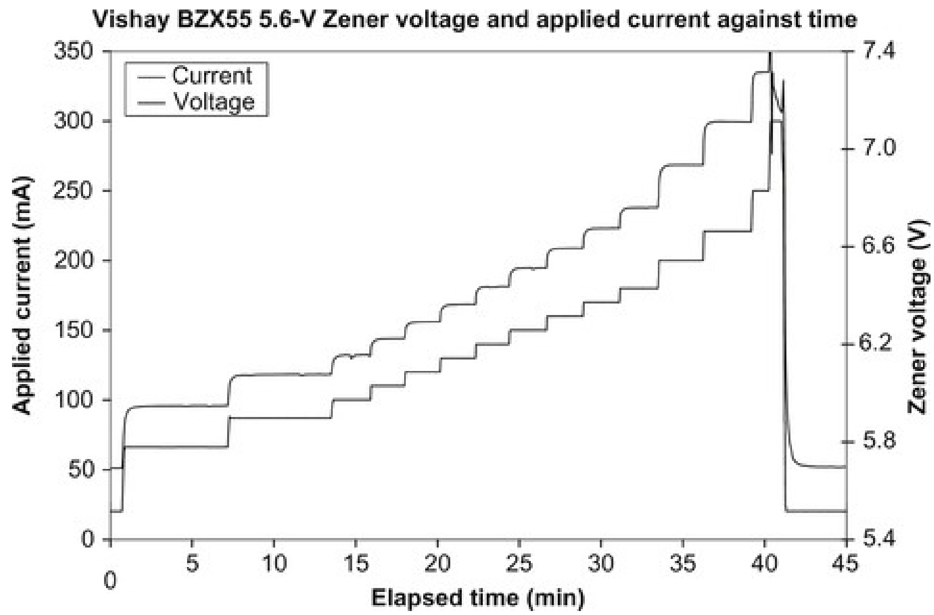


Figure 5.66 Stressing a BZX55 5.6-V Zener: voltage and applied current against time.

Measurements were logged at 1 s intervals using a pair of Agilent 34410A 6½ digit DMMs. The exponential curve following the transient on the Zener voltage (upper trace) is due to the wires and silicon within the Zener heating and increasing their resistance as a result of the increased applied current (lower trace). The Zener was run for a short time at the statistical regulator's design current of 20 mA, then increased to 68 mA to simulate a fault in the balanced hybrid RIAA stage, and then 87 mA to simulate a fault in the EC8010 RIAA stage. The Zener survived 5 min at 87 mA, so it was clearly time for some abuse. Current was increased in 10 mA steps and left for roughly 150 s each time. After half an hour of this, the author's patience was running thin, and he finally applied 300 mA, which *did* provoke a reaction (smoke from the cardboard and a spike on the graph), so he backed the current off to the original 20 mA. Astonishingly, despite discoloured lead-outs, once the Zener had cooled its voltage at 20 mA was within 0.1% of the original voltage.

There seems to be no justification for fearing damage to the composite Zener from a fault causing its current to rise to 87 mA.

Reference Voltages

The 40 V and 270 V references could be tapped off the statistical regulator's composite Zener, but having devised an astonishingly low-noise regulator, the author isn't about to ruin it by injecting mains interference via the inter-winding capacitance of a mains transformer, even if that capacitance would be divided down by a transistor's h_{fe}.

The safest solution is to use a THINGY configured for 40 V and 270 V and feed it from the raw supply before the regulator so that any mains interference that does manage to crawl back through the THINGY can be attenuated 100 dB by the statistical regulator.

HT Rectification and Smoothing (a PSUD2 Exercise)

We require 390 V, and we need at least 20 V across the statistical regulator's CCS, so if we allow 10% for mains variation, we need 451 V, say 460 V. This voltage is beyond what a single electrolytic capacitor could tolerate, and putting a pair in series would double typical series inductance from 18 nH to 36 nH, add 20 mm of wire at 0.75 nH/mm, and we would have ≈50 nH, guaranteeing poor RF filtering. Once again it makes sense to use an *LC* filter using a plastic capacitor having a Kelvin connection. The capacitor needs to be rated at 600 V, so 100 µF would be larger than that can be wound, but 47 µF is feasible.

A valve rectifier allows HT to be applied gently, and the critical parameter governing rectifier choice is now voltage rather than current rating. The traditional choice would have been a GZ34, but these are now far too expensive and a pair of 6CK3/6CL3 damper diodes is far cheaper. If we don't want to use a centre-tapped transformer, a hybrid bridge rectifier can be made by adding a pair of STTA512Fs to the 6CK3/6CL3.

For an 80 dB signal-to-hum ratio, the EC8010 RIAA stage requires ripple on the 390 V HT to be <1.7 mV $_{pk\text{–}pk}$ (slightly less tolerant than the balanced hybrid RIAA stage). The statistical regulator is certain to have >90 dB ripple attenuation, so this corresponds to it tolerating 53 V $_{pk\text{–}pk}$ ripple post-rectification. As before, this means we don't have to worry greatly about ripple and can concentrate on obtaining the required voltage using available components and ensuring Low Frequency stability.

When you experiment with PSUD2, you will quickly discover that it is difficult to achieve Low Frequency stability in an *LC* filter having a 47 µF capacitor. The inductance of the choke can be halved from 20 H to 10 H (reducing the *L*/ *C* ratio) to ease the problem, but Low Frequency stability can only be achieved by adding a capacitor before the choke. Even so, with typical transformer and choke resistances, only a capacitor of around 3–4 µF before the choke completely eliminates Low Frequency ringing. It seems we have designed another intermediate mode supply.

Given that low frequency stability specified the filter components quite tightly, the required 460 V $_{DC}$ output voltage can only be achieved by adjusting the transformer secondary voltage to 400 V $_{RMS}$ (off-load) (see Figure 5.67).

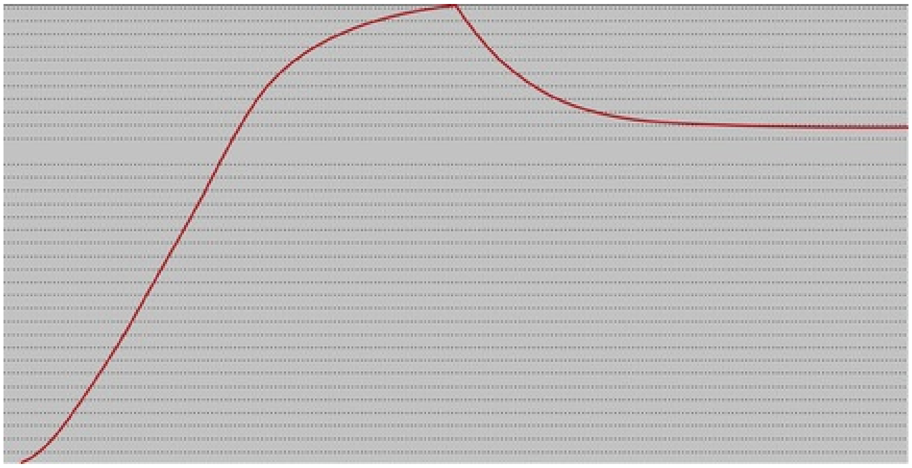


Figure 5.67 PSUD2 simulation achieves required voltage without low frequency ringing.

Having achieved the required DC voltage, we find that the ripple is 680 mV $_{pk–pk}$, so there will be no problem in achieving our 80 dB signal-to-hum ratio.

This is the point where we consider it a mild nuisance that PSUD2 works with off-load voltages because transformers are specified by their on-load voltage. There are two ways around this problem: either use the bar at the bottom of the screen to reveal the on-load RMS voltage leaving the transformer, or make a guess as to typical transformer regulation. Although the first method seems the best, it contains the hidden assumption that you guessed secondary and reflected primary resistance correctly. The author prefers to make the sweeping assumption that 50–100 VA transformers have ≈10% regulation, so that an off-load voltage of 400 V $_{RMS}$ corresponds to an on-load voltage of 372 V $_{RMS}$. This can be used as a rough check on the accuracy of your initial guess as to transformer secondary and reflected primary resistance – if you have guessed it reasonably well, the PSUD2 voltage will be about 3% lower than the 10% regulation guesstimate. (This discrepancy occurs because PSUD2 can't model core losses.)

Heater Regulation

The heaters in the EC8010 RIAA stage are run constant current rather than constant voltage and all six strings need 300 mA at an expected maximum voltage of 12.6 V. The 317T can be configured to produce a constant current very easily by placing a single resistor between OUT and ADJ that drops 1.25 V at the chosen current (see Figure 5.68).

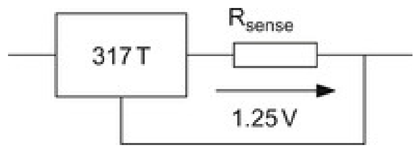


Figure 5.68 317 as a constant current regulator.

A 4.17 Ω current sense resistor is needed, and this can be achieved by a 4.7 Ω resistor in parallel with 36 Ω. The required power rating (0.33 W) of the 4.7 Ω resistor combined with 1% precision used to be a problem, but TO-220 package 4.7 Ω 1% 20 W resistors are now readily available. Each regulator can be made very simply by fitting the TO-220 resistor adjacent to the 317T on the heatsink and hard-wiring them and the 36 Ω resistor (no PCB needed).
When the heaters are cold, their resistance is low, so almost the entire reservoir capacitor’s voltage appears across each 317T, almost doubling its thermal dissipation, but after 20 s dissipation drops to normal (see Figure 5.69).

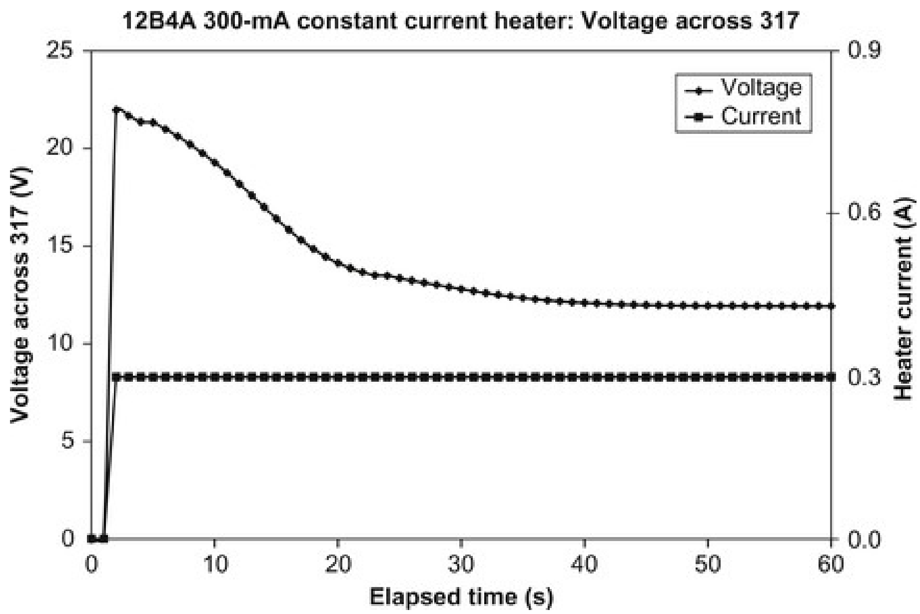


Figure 5.69 Cold resistance of valves causes large voltage drop across the 317 at switch-on.

Heater Rectification and Smoothing (a Manual Exercise)

The 317T needs >3 V across it to regulate cleanly and 1.25 V is dropped across the current sense resistor, so we need 4.25 V more than the expected 12.6 V across each heater string, implying 17 V minimum required across the reservoir capacitor.

Looking from the other direction, 2×18 V transformers are common, so they would produce 25.5 V_{pk} or 22.9 V if the mains drops by 10%. The bridge rectifier will drop 1.4 V, leaving 21.5 V, allowing 4.5 V of ripple. Each series chain requires 300 mA, so the four chains require a total of 1.2 A, and the other two chains require a total 600 mA. The 1.2 A supply requires a 3,300 μF capacitor for 3.6 V_{pk-pk} ripple, whereas 1,500 μF would suffice for 600 mA supply. Although it might be easier to use 3,300 μF for both, it would halve the ripple voltage on the 600 mA supply, raising the DC voltage and unnecessarily increasing the dissipation of the associated 317T.

The total DC current is 1.8 A from our raw 25.5 V_{pk}, so this is 46 W, requiring a 50 VA transformer having a split bobbin or foil electrostatic screen to minimise common-mode interference.

Now that we know the voltage on the reservoir capacitor, we can estimate the thermal dissipation of each 317T. If we treat the DC on the reservoir as being its peak voltage minus half the ripple, then:

$$V_{DC} = V_{pk} - \frac{V_{ripple}}{2} = 25.5\ \text{V} - \frac{3.6\ \text{V}}{2} = 23.7\ \text{V}$$

Each heater chain drops 12.6 V, and 1.25 V is dropped across the current sense resistor, leaving 9.85 V across the 317T, and $P = IV = 9.85\ \text{V} \times 0.3\ \text{A} \approx 3\ \text{W}$. It is convenient to mount all six 317Ts on a common heatsink (via insulating kits), so it must dissipate 18 W, and for a 20 °C temperature rise, that implies a thermal resistance of <1.1 °C/W.

We now have a full power supply circuit diagram (see Figure 5.70).

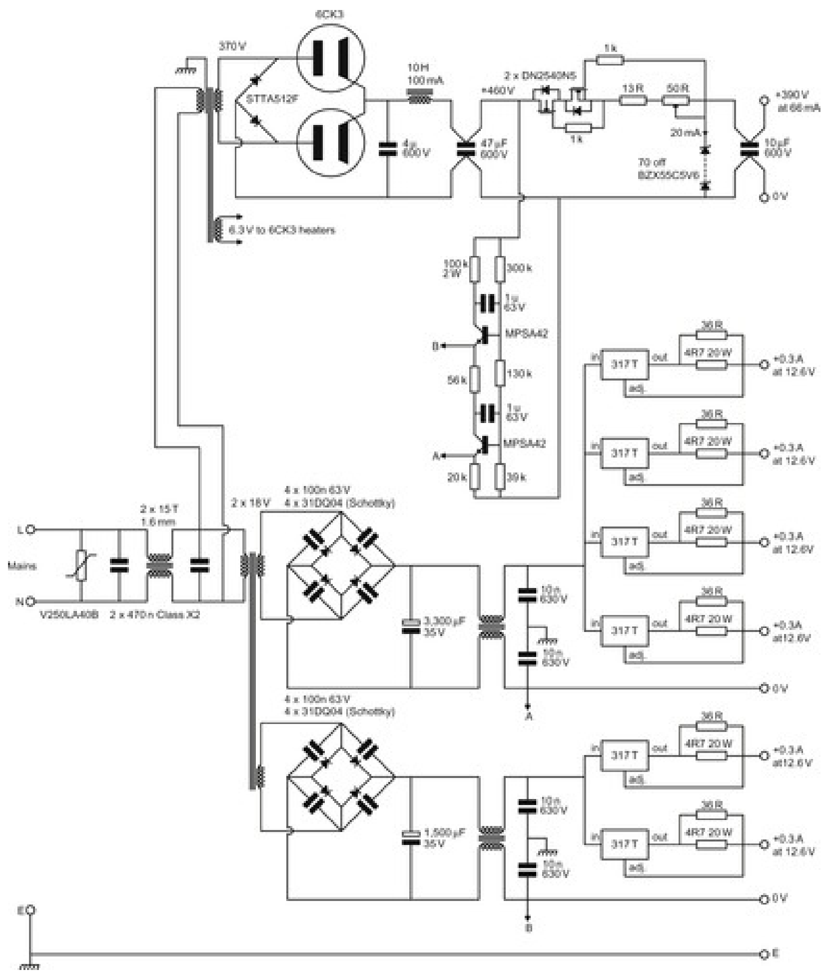


Figure 5.70 Complete power supply for EC8010 RIAA stage.

A power supply is a device that converts one voltage to another more convenient voltage whilst delivering power.

[1] Watts M. Hybrid bridge rectifier. Electronics World 1999;August:650.

[2] Yarwood, J, *An introduction to electronics*. (1950)Chapman & Hall; ;92–6.

[3] ESU872 half-wave mercury vapour rectifier data sheet. Edison Swan Electric; July 1949.

[4] Gas-filled rectifiers: general operational recommendations. Mullard data sheets.

[5] Vintage Hi-Fi: Krell KSA-50 power amplifier. Hi-Fi News & Record Review; May 2011:110–14.

[6] Schade, OH, Analysis of rectifier operation, *Proc IRE* **31** (July (7)) (1943) 341–361.

[7] Erzeugung einer Gleichspannung vom vielfachen Betrage einer Wechselspannung ohne Transformator. Bulletin des Schweizer Elektrotechnischer Verein 1920:11.

[8] Verfahren zur Erzeugung einer Gleichspannung mit Hilfe einer oszillierenden Spannung in einem beliebigen vielfachen Betrag derselben unter Verwendung von elektrischen Ventilen un Kondensatoren. Swiss Patent 93,107; 1922.

[9] Cockcroft, JD; Walton, ETS, Experiments with high velocity positive ions, *Proc Royal Society (A)* **136** (1932) 619–631.

[10] Cockcroft JD. Improvements in high voltage direct current systems. UK Patent 365,785; 1932.

[11] Cockcroft JD, Walton ETS. System for the voltage transformation of direct current electrical energy. US Patent 1,992,908; 1935.

[12] Smoothing circuits: (1) Resistance. Capacitance Cathode Ray. Wireless World 1949;October:389–93.

[13] Arthur, K; Thompson, J, *Power supply circuits*. 2nd ed. (1968)Tektronix.

[14] 3-Terminal regulator is adjustable. National Semiconductor Application Note 181. Linear Applications Handbook.

[15] '1990 Linear Databook' and '1992 Linear Databook Supplement'. Linear Technology.

[16] Maida M. High voltage adjustable power supplies. Linear Applications Handbook. National Semiconductor Linear Brief 47.

[17] Type 555 cathode ray oscilloscope instruction manual. Tektronix; 1961.

[18] Kobayashi M. Electric oscillator. US Patent 1,913,449atsugo; 1929.

Chapter 6. The Power Amplifier

The determining factor is the output stage. The solution adopted here dictates the topology of the remainder of the amplifier, so we will begin by investigating the output stage.

The Output Stage

Typical audio valves are high-impedance devices and can swing hundreds of volts, but deliver only tens of milliamperes of current. By contrast, a loudspeaker of typically 4–8 Ω nominal impedance requires tens of volts and amperes of current. The obvious solution to this problem is to employ an *output transformer* to match the loudspeaker load to the output valve or valves.

This is where the problems start. As was hinted earlier, transformers are rather less than perfect, and the ultimate quality of a valve amplifier is limited by the quality of its output transformer. Despite this, the transformer coupled output stage is a good engineering solution, and is used in most valve amplifiers (see later for Output Transformer-Less designs).

Valves designed specifically for audio use generally have optimised configurations that are detailed in the manufacturer's data sheets. Designing output stages for audio valves from first principles is reinventing the wheel, but an overview of the practicalities is most useful; therefore, we will indulge in a brief analysis of a currently fashionable topology.

The Single-Ended Class A Output Stage

A typical transformer coupled output stage is the familiar common cathode triode amplifier using cathode bias (see Figure 6.1).

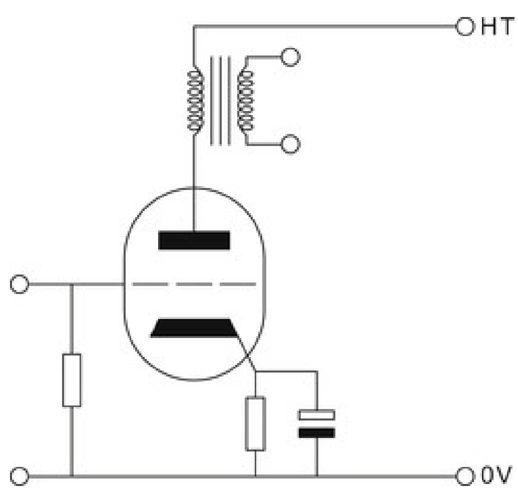


Figure 6.1 Single-ended transformer coupled stage.

When we investigated voltage stages, we used a loadline to choose the value of anode load, and generally optimised for linearity, rather than voltage swing; this time, we need to maximise power. For this example we will use an E182CC double triode, which might be useful as a headphone amplifier. We would normally set the operating point at the intersection between maximum continuous anode voltage ($V_a = 300$ V) and maximum anode dissipation ($P_a = 4.5$ W). But as there is a grid curve intersecting at $V_a = 295$ V, the operating point has been moved for convenience. For maximum output power, the optimum load for a triode is $2 \times r_a$. In our example r_a is 3.57 kΩ, so $R_L = 2 \times r_a = 7.14$ kΩ, and we plot this loadline (see Figure 6.2).

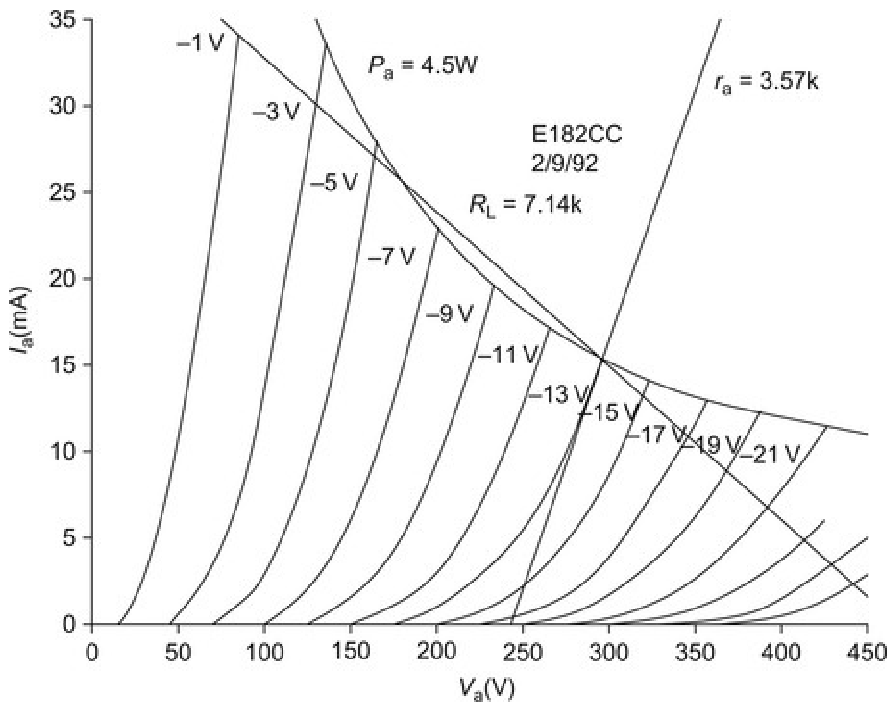


Figure 6.2 AC loadline for transformer coupled stage.

$V_{gk} = -1$ V is our positive limit from the bias point of $V_{gk} = -13$ V, therefore the negative limit will be $V_{gk} = -25$ V for a symmetric input voltage. This results in a peak to peak output voltage swing of 430 V to 85 V=345 V, or 122 V_{RMS} , which equals 2.1 W dissipated in the load. Under these conditions, 4.5 W is dissipated in the valve, giving an efficiency of 32%.

We should now observe some important points about the operation of this stage:

- The loadline strays into the region where $P_a \geq 4.5$ W. Since the stage is driven only with AC (it must be, as we could not otherwise transformer couple to the load), this is not a problem. This is because although on one half-cycle the anode dissipation is ≥4.5 W, on the other half-cycle it is less, and the thermal inertia of the anode will average the dissipation out at ≤4.5 W.

- We set the operating point of the valve at 300 V. If the transformer is perfect, then there will be no DC voltage dropped across the primary winding, and so the HT voltage must be 300 V. Yet we have allowed V_a to rise to 430 V, which is considerably above HT. This is possible because the transformer stores energy in the magnetic flux of its core. In theory, a perfect valve could swing V_a from 0 V to 2×HT, which is a very useful feature in a power amplifier.

- We carefully set our anode load at 7.14 kΩ, but in doing so we assumed that the loudspeaker was a resistor. Loudspeakers are not resistive, and the transformer is not perfect, so the actual load seen by the valve will not be a precise resistance, but a complex and variable impedance.

  The valve therefore sees an AC loadline that is an ellipse with its major axis roughly aligned with the theoretical resistive loadline. The gradient of the major axis is the resistive component, and the width of the minor axis indicates the relative size of the reactive component. This means that most of the calculations we can make for an output stage are informed guesses at best, and there is little point worrying about precise values.

- Because we wish to maximise the power in the load, we have to maximise the anode voltage swing, resulting in poor linearity. We could improve linearity by increasing the value of the anode load and plotting another loadline, but this will reduce available output power.

Although the linearity of single-ended stages is not good, the distortion produced is mostly second harmonic, which, as we observed earlier, is relatively benign to the ear. We can estimate the percentage of second harmonic distortion from the following formula:

$$\%D_{2\text{nd harmonic}} \approx \frac{V_{\text{quiescent}} - (V_{\max} + V_{\min}/2)}{V_{\max} - V_{\min}} \times 100\%$$

In our example, $V_{\max}$ =430 V, $V_{\min}$ =85 V and $V_{\text{quiescent}}$ =295 V, resulting in 11% second harmonic distortion at full output. Clearly,>10% distortion is not Hi-Fi, but the attraction of single-ended amplifiers is that their distortion is always directly proportional to level, and so at one-tenth output power, the distortion would be ≈1%, and so on. Since, most of the time, music requires very little power, it is often argued (oddly enough, by single-ended enthusiasts) that it is the quality of the first watt that is important, not the remainder that are rarely used. The distortion could be reduced by negative feedback, but this technique is almost universally shunned by the supporters of single-ended amplifiers, so single-ended amplifiers not only produce high distortion, but also tend to have an output resistance of half the assumed load resistance.

The Significance of High Output Resistance

The vast majority of modern loudspeakers use moving coil drivers in sealed or reflex boxes. The theory of the interaction between moving coil loudspeakers and their enclosures was introduced in a landmark series of papers by A.N. Thiele and R. Small in the *Journal of the Audio Engineering Society* in the early 1970s. A closed box is a second order high-pass filter, whereas a reflex box is fourth order, although it can be designed to look like third order. The crucial point is that Thiele and Small showed that the Q of the high-pass filter could be precisely set by the series contribution of voice coil resistance, crossover resistance, and amplifier output resistance, which is normally assumed to be zero. Typical single-ended amplifiers make a mockery of the zero output resistance assumption and cause the loudspeaker to produce a peaked bass response that the loudspeaker designer did not intend.

Reflex loudspeakers developed prior to Thiele and Small relied on the mechanical damping contributed by the suspension of the loudspeaker to determine bass response and made few assumptions about amplifier output resistance. Horns rely on the transformed air load for damping, so they too are tolerant of high amplifier output resistance. As a further bonus, both of these types of loudspeakers are sensitive, so they are popular with aficionados of single-ended amplifiers.

Bass is produced by moving a large volume of air, and requires a large, rigid, and consequently heavy, cone. Treble is produced by accelerating and decelerating a small area many times a second, requiring low mass. The requirements for bass and treble reproduction are contradictory, so most loudspeaker designers prefer to use optimised drivers for specific frequency bands supplied from an electrical filter known as a crossover. Nevertheless, others feel that the practice of multiple non-coincident drivers and crossovers is so fundamentally flawed that they attempt the design of full-range drivers. In practice, the treble response tends to be peaky and directional after 10 kHz, and the low-frequency resonance limits bass to ≈100 Hz, but this is a sizeable proportion of the audio band. Fortuitously, the motion of a full-range driver's low-mass cone is easily damped, and when mounted on an open baffle, the mechanical damping of the suspension is frequently sufficient. Thus, a sensitive full-range driver mounted on an open baffle can be an ideal match for a single-ended amplifier.

Transformer Imperfections

When we plotted the single-ended loadline, we treated the output transformer as being perfect, but we should now consider how its imperfections will affect the stage. Unfortunately, we pass a constant magnetising current ($I_{quiescent}$) through the primary of the output transformer. In order for the core not to saturate, which would cause odd harmonic distortion, we need a large gapped core. Another method of avoiding core saturation is to reduce the number of primary windings, which not only reduces the magnetising effect (ampere-turns, In) of the quiescent current but also reduces primary inductance.

Usually both methods have to be used, which results in a physically large transformer of low primary inductance at the operating point. Because the transformer is so large, it has correspondingly large stray capacitances, so high frequency performance is also compromised. Typically, the transformers used in this way are large, expensive, and have significantly reduced bandwidth compared to an (ungapped) push–pull transformer.

It might therefore be thought that the single-ended transformer-coupled power stage is a complete non-starter, but, curiously, this is not so. If we look at the hysteresis curve for transformer iron, this offers a clue to this topology's recent resurgence in popularity (see Figure 6.3).

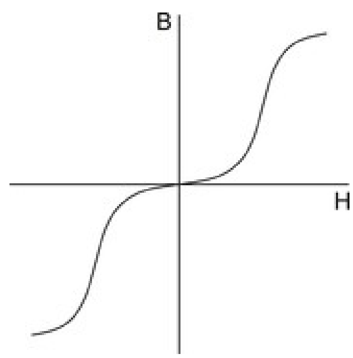


Figure 6.3 Exaggerated *B/H* curve of iron.

When used as a transformer, the hysteresis curve may be treated as a transfer characteristic showing the relationship between V_{in} and V_{out} . If there were no DC current flowing through the transformer, then an AC signal would swing symmetrically about the origin. If we look at the small-signal performance, we see that there is a kink in the characteristic around the origin where the slope of the curve is reduced. Since the core's permeability is proportional to the slope, the transformer has reduced primary inductance (L_p) at low levels. At low frequencies, reduced L_p reduces gain and increases distortion in the output stage. The cause of the kink is that the individual magnetic domains that make up the core have stiction in reversing the polarity of their magnetism. (The same effect is true in electrostatics, in that there is stiction in reversing electrostatic charges in polar dielectrics such as polyester and polycarbonate.) A recently popular solution known as *pinstriping* uses a mix of steel and mu-metal laminations to make up the core. Mu-metal has a much higher initial permeability, so it maintains high L_p at low levels, but saturates quite quickly, at which point the steel takes over, so pinstriping can improve the initial permeability of the core. Unfortunately, mu-metal is fragile and significantly more expensive than steel.

Alternatively, by passing the valve's quiescent current through the transformer, we ameliorate the problem of low initial permeability making the transfer characteristic more linear, and it is claimed that this is responsible for the excellent midrange detail in this breed of amplifiers. However, a more prosaic explanation is that voice coil inductance causes loudspeaker impedance to rise above 250 Hz, and that when driven from a source resistance of half the loudspeaker’s nominal impedance this typically results in a lift of 3 dB by 3 kHz.

Although the transformer has a low primary inductance, suggesting a poor bass response, a well-designed core is less likely to saturate at low frequencies, since it had to be oversize and gapped to accommodate the quiescent current. Because of this, L_p is nearly constant from full AC output power to zero AC output. Provided that the loudspeaker is carefully matched, a good subjective bass quality can be achieved, because it does not change significantly with level.

Unfortunately, we can make no excuses for the high frequency performance. The large, leaky output transformer has significant losses at high frequency, although excellent construction helps matters.

Single-ended amplifiers typically use true triodes with directly heated cathodes, such as 2A3, 300B, 211 and 845, rather than beam tetrodes or pentodes connected as triodes. Unfortunately, directly heated cathodes are prone to hum when fed with AC, or premature failure when fed with DC, because V_{gk} at one end of the heater/cathode is lower than the other, causing higher emission at the low V_{gk} end.

To sum up: single-ended triode amplifiers have good low-level performance but they require careful loudspeaker matching to make the most of their bass performance and low power (usually <10 W), and their cost per audio watt is high due to the expensive output transformer and esoteric valves. But they are simple to make.

Classes of Amplifiers

The 'class' of an amplifier refers to the proportion of quiescent anode current to signal current. Until now, we have only looked at Class A amplifiers, although the fact was not explicitly stated. If we relax that restriction, we will need some definitions. Efficiency is defined as:

$$\text{efficiency } (\eta) = \frac{\text{power out}}{\text{power in}}$$

Although the valve also requires power for the heaters, it is usual to define the 'power in' as being from the HT supply.

The following comparisons all assume that the load is coupled to the amplifier by a perfect output transformer because this enables the DC component of the valve's current to pass through the (reflected) load without loss (no resistive losses in a perfect transformer). Further, the output transformer stores energy in its primary inductance enabling the anode to swing linearly to twice the HT voltage.

Class A

Anode current is set such that even with maximum allowable input signal, current never falls to zero. In other words, the valve never switches off. (Maximum theoretical efficiency of a Class A amplifier is 50% for a sine wave output.)

Class B

There is zero quiescent anode current, and current only flows during the positive half-cycle of the input waveform. The valve is therefore switched *off* for the negative half-cycle of the input waveform, and considerable distortion of the signal occurs, since it has been half-wave rectified. Additional measures need to be taken to deal with this problem. (Maximum theoretical efficiency for sine wave output is 78.5% for a push–pull Class B amplifier.)

Class C

Anode current flows for *less* than half a cycle of the input waveform. This method is used only in radio frequency amplifiers where resonant techniques can be used to restore the missing portion of the signal, and results in even greater efficiency and distortion than Class B.

Radio frequency engineers refer to the *conduction angle* to specify the proportion of time in which anode current flows. Using this description, we see that Class A amplifiers have a conduction angle of 360°, Class B has 180° and Class C <180°. The transition between Class A and pure Class B is quite broad, and so there is an intermediate class known as Class AB (see Figure 6.4).

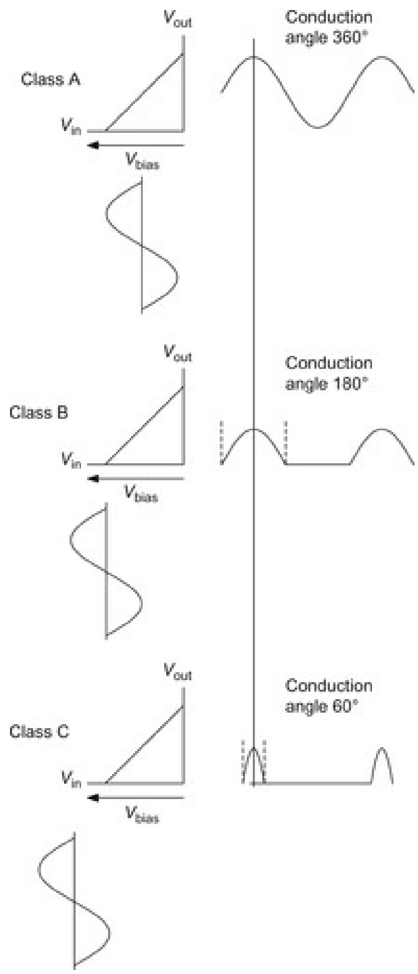


Figure 6.4 Relationship between anode current and input signal for Classes A–C.

In Figure 6.4 , the transfer characteristic of the output device is assumed to be perfect, so the input sine wave is simply reflected through the diagonal transfer characteristic to produce the output. In Class B operation, the bias voltage cuts off the valve, and it is only on positive half-cycles that the signal is able to switch the valve on. It will be noticed that the output waveform of the Class B stage is very similar to the power supply waveforms in Chapter 5 and, for the same reason, half-wave rectification is taking place.

Note that as bias voltage is increased negatively, the conduction angle falls.

In audio, we normally refer to currents rather than conduction angles, and there are subdivisions of classes defined by the *grid* current of the valve. (RF engineers are unable to do this since they invariably operate with grid current to maximise efficiency at the expense of linearity.)

*Class *1*

Grid current is not allowed to flow. Many of the larger (≥50 W) classic amplifiers were push–pull Class AB1.

*Class *2*

The input signal is allowed to drive the grid positive with respect to the cathode, causing grid current to flow. This improves efficiency, since the anode voltage can now more closely approach zero, which is particularly relevant to triodes. At the onset of grid current, the input resistance of the output stage falls drastically (possibly approaching $1/g_m$), and the driver stage needs a very low output resistance if it is to maintain an undistorted signal into this extremely non-linear load without distortion. Some modern single-ended amplifiers use high-μ transmitter triodes intended for Class B2 operation that pass very little anode current at $V_{gk}=0$, so they bias the grid positive to force the required quiescent current, and thus operate entirely in Class A2.

The Push–Pull Output Stage and the Output Transformer

We saw that the Class B stage introduced considerable distortion by half-wave rectifying the input signal. Clearly, this is a disadvantage for a Hi-Fi amplifier, since we require linearity.

Suppose, however, that we had two Class B valves, one fed directly with the input signal, and the other with an *inverted* signal. During time t_1 the upper valve conducts, whilst the lower is cut-off, and during t_2, the situation is reversed (see Figure 6.5).

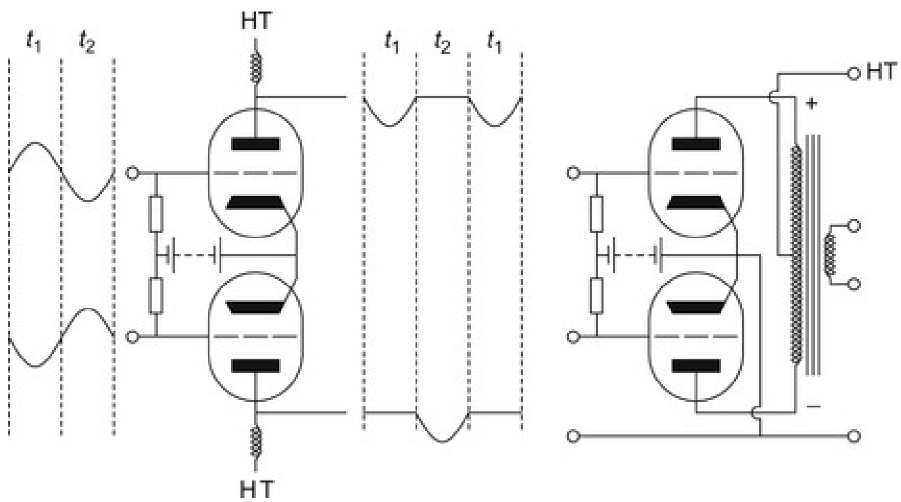


Figure 6.5 Summation of Class B signals in output transformer.

So far all that we have achieved is to ensure that any *one* valve is switched on, irrespective of incoming signal polarity. However, by inverting one output and summing it with the other in the output transformer, we can recreate the shape of the original input waveform. The inversion is performed by reversing the connection of one winding, and is marked on the diagram with + and – symbols.

Whether achieved by a transformer or by a direct coupled series amplifier, such as a White cathode follower, this form of connection is known as *push–pull* , and is the only way of approximating linearity in a Class B amplifier.

Unsurprisingly, this dissection of the signal and its subsequent restitching is rather less than perfect, and pure Class B is rarely used because of the distortion generated at the crossover region, where one device takes over from the other. In practice, some quiescent current is allowed to flow in an attempt to smooth the transition, resulting in Class AB operation. The theoretical optimum bias voltage for a Class AB amplifier is found by extending the linear part of the transfer characteristic until it intersects the V_{gk} -axis. However, practical devices do not operate linearly down to cut-off and then suddenly switch off, so individual differences between devices mean that the ideal point is ill-defined, and crossover distortion is not eliminated.

Push–pull output stages can also be used for Class A amplifiers, giving additional advantages.

Because of the reversal of one winding, the magnetising flux caused by the quiescent anode currents cancels (provided that they are equal). Because the transformer core only has to handle signal current, it does not need a gap and can be far smaller for a given power — this is the *main* reason for using a push–pull output stage in a Class A amplifier.

Since the core is small, it is important that the quiescent anode current of each valve *is* identical, otherwise DC magnetisation of the core will generate odd harmonic distortion. This can be done by having an adjustment for DC balance in the bias circuit, or by using pairs of valves with matched anode currents (see Figure 6.6).

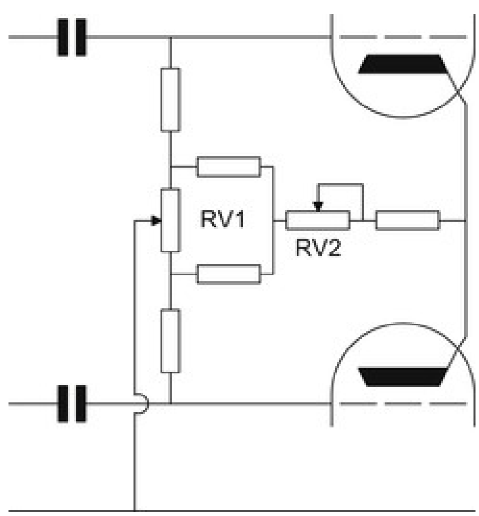


Figure 6.6 DC balance adjustment.

RV2 sets total anode current, whilst RV1 adjusts DC balance by biassing one grid more or less positively than the other.

If the core should become permanently magnetised (perhaps by failure of a valve), it will need to be degaussed, or it will generate additional (and unnecessary) distortion. This can be done by applying a sufficiently large alternating magnetic field to the core to saturate it both positively and negatively, and then reducing the field to zero over a period of about 10 s. In practice, the low remanence of typical transformer core materials means that this procedure is unlikely to be necessary.

A useful consequence of the reduced size of transformer is an improved High Frequency response due to the reduction of stray capacitances.

Not only does quiescent anode current cancel in the transformer, but power supply hum also cancels, since it is in-phase in each winding. Improved tolerance to power supply hum allows a cheaper power supply.

Additionally, even harmonic distortion, caused by unequal gain on positive and negative half-cycles, is cancelled, whilst odd harmonic distortion is summed. Since triodes generate primarily even harmonic distortion, this is useful, but pentodes generate primarily odd harmonics, and therefore require considerable (>20 dB) negative feedback to reduce their distortion to acceptable levels. Cancellation of even harmonic distortion is only achieved if each winding is fed identical signal voltage by its valve, so some amplifiers have adjustments for AC balance, whereas others specify pairs of valves matched for gain (see Figure 6.7).

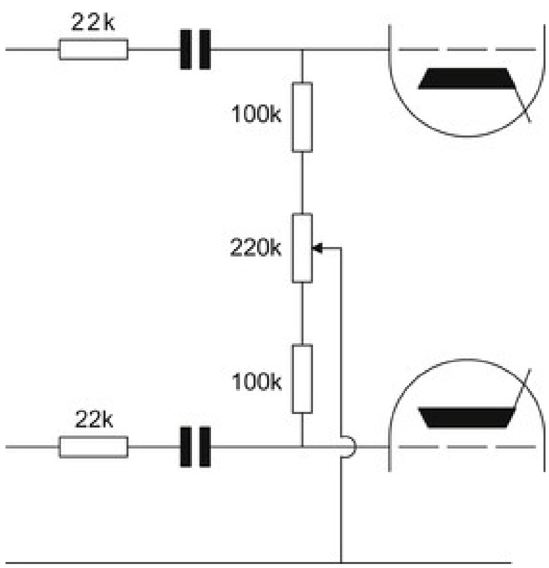


Figure 6.7 AC balance adjustment.

Modifying the Connection of the Output Transformer

We have mainly considered triodes, and given pentodes scant regard because of their odd harmonic distortion. But if we imagine the output transformer primary as a set of windings that could be tapped at any point, we see that for pentode operation g_2 would be connected to the centre tap (0%), whereas for a triode it would be connected at the anode (100%) (see Figure 6.8).

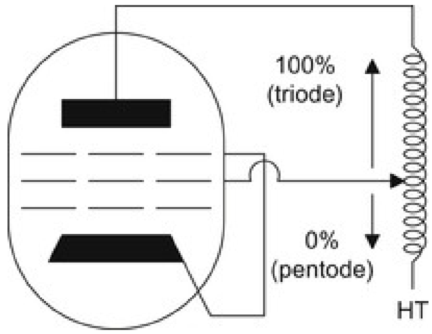


Figure 6.8 Blumlein or 'ultra-linear' output stage.

What would happen if we were to tap at an intermediate point? This question was asked in 1951 by David Hafler and Herbert I. Keroes [1] and an amplifier named *ultra-linear* became synonymous with an output stage that had been invented by Alan Blumlein [2] in 1937. Regarding tapping points, Mullard's EL84 and EL34 data sheets quite clearly state that 43% gave minimum distortion and 20% maximum power, although it was suggested [3] in 1958 that 20% gave minimum distortion for KT66. This form of distributed loading became almost universal in the final days of valve supremacy, since it combined the efficiency and ease of driving the pentode with much of the improved linearity of the triode. It should be noted that:

$$I_a \propto V_{g2}^{3/2}$$

And as a consequence, negative feedback at g_2 is not as linear a process as one might wish. Nevertheless, almost all power amplifiers using pentodes in the output stage use this scheme because it is far superior to pure pentodes.

Up until now we have placed the transformer in the anode circuit, but we could place the same transformer in the cathode circuit to form a cathode follower resulting in an extremely low output resistance from the valve. As an example, a pair of EL34s, connected as triodes, would each have an anode resistance of about 900 Ω, but used as cathode followers the driving resistance would be a tenth of this at 90 Ω. Reflected through the transformer, the output resistance seen by the loudspeaker would be a fraction of an ohm even before global negative feedback. Unfortunately, there are two crippling disadvantages to this topology.

Firstly, although the output stage is excellent, we have transferred its problems to the driver stage. Each output valve now swings ≈150 V_{RMS} on its cathode, and has a gain <1, requiring ≈500 V_{pk-pk} to drive it! This can be done, but it is not a trivial exercise to design the driver stage, since we must either use transformer inter-stage coupling, or a resistive anode load requiring a high HT voltage.

Secondly, the high voltage on the cathode of the output valves severely strains the heater/cathode insulation, and can cause premature heater failure. Connecting the heater to the cathode transfers the stress to the heater transformer but requires individual heater windings for each half of the output stage (to avoid shorting k_1 to k_2), and forces each valve to drive the transformer's inter-winding capacitance (≈1 nF). In the early days of television, the (expensive) display tube often failed because of leaky heater to cathode insulation, so a service engineer's fix that avoided the expense of a new tube was to tie the heater to the cathode and supply it from a new transformer rather than including it in the series heater string. But because the 3 MHz bandwidth video signal was applied to the cathode, the transformer needed a low inter-winding capacitance (typically 25 pF), so these were readily available. A similar transformer having multiple low capacitance windings would be ideal for cathode follower output valves.

Alternatively, there are a few valves (usually intended for use as the series-pass element in regulators) whose heater/cathode insulation can withstand 300 V, such as the 6080/6AS7G. Because this valve has such a low r_a , its optimum load resistance is fairly low, and its output voltage at full power is also quite low, reducing the strain on heater/cathode insulation. Unfortunately, μ is also very low, so the gain of the output stage is substantially less than unity, and the driver stage has to be quite special (see Figure 6.9).

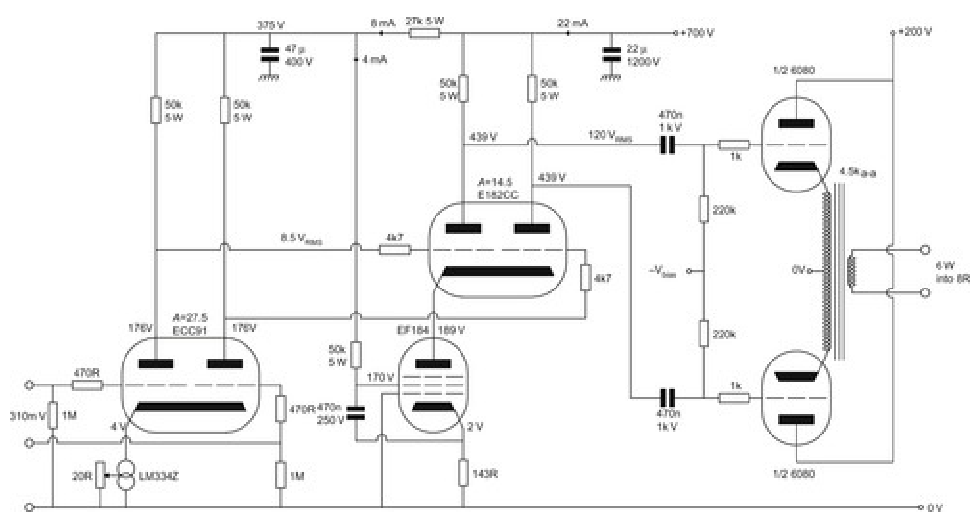


Figure 6.9 Amplifier with cathode follower output stage.

As can be seen, a complex power supply would be required simply to produce 6 W. Admittedly, the driver could cope with a number of 6080s in parallel, but the cure still seems worse than the complaint. The only reason that this design survived to the drawing stage is that the output stage should be quite tolerant of poor output transformers; conversely, a good output transformer would allow very good performance. The amplifier uses triodes throughout, whose distortion is predominantly second harmonic, but this is cancelled by push–pull action, so the amplifier relies on accurate balance rather than global feedback to reduce distortion, and has a balanced input.

Another form of distributed loading places part of the load in the cathode, and part in the anode; Peter Walker's Quad II amplifier in the UK used this technique to gain useful benefits from local feedback and distortion cancellation with reasonably relaxed driving requirements. McIntosh [4] in the USA took the technique to its logical conclusion by having equal anode and cathode loads bifilar wound to minimise the output transformer leakage inductance that exacerbated Class B crossover distortion. Sadly, the drive requirements are almost as severe as for the cathode follower, and insulation between the bifilar anode and cathode windings is crucial, so the technique has not been widely adopted (see Figure 6.10).

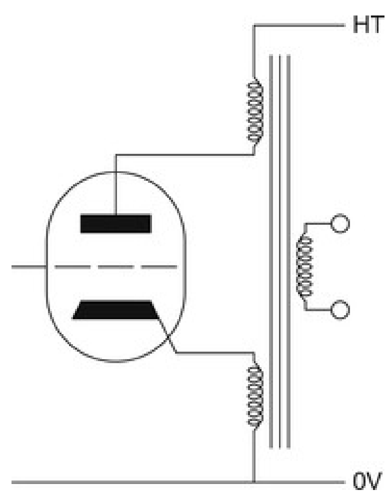


Figure 6.10 Quad II output stage (aka McIntosh configuration).

Each output valve controls its current through equal anode and cathode loads so any change in anode and cathode voltage is equal and opposite. Cross-coupling g_2 of each valve to the opposite valve's anode means that for a given valve, g_2 and cathode voltage changes track, implying pure pentode operation, and this is significant because:

- A pentode can swing its anode closer to 0 V, allowing it to deliver more power to the load than a triode.

- A pentode's g_2 prevents feedback from the anode to the control grid reducing gain. We often consider this feedback to be useful (it's what causes the low r_a of a triode), but if we want to apply external feedback, minimising this gain reduction becomes important. Thus, achieving pentode gain before applying transformer feedback maximises the amount of (distortion-reducing) feedback.

The commercial technique of bootstrapping the driver stage's HT to reduce its distortion and increase voltage swing is a form of positive feedback and makes amplifier high-frequency stability crucially dependent on output transformer design.

Output Transformer-Less (OTL) Amplifiers

Almost all the different output stage configurations were devised in an effort to reduce the adverse effect of the output transformer, so it is not surprising to find that there have been some designs that dispense with the output transformer. These are often known as Futterman [5] amplifiers (who patented the notion), or OTLs.

Driving low-impedance loads directly is not natural for a valve, so radical approaches are needed. High peak currents are required, so valves with robust cathodes are needed, which invariably were not designed for audio, and they therefore have extremely questionable linearity and consistency. Examples are the 6080/6AS7G double triode series regulator valve, and television line scan output valves such as the PL504 and PL519 pentodes. Efficiency is generally on the low side of appalling. Output stages invariably use paralleled White cathode followers with plenty of global feedback to reduce the output resistance (see Figure 6.11).

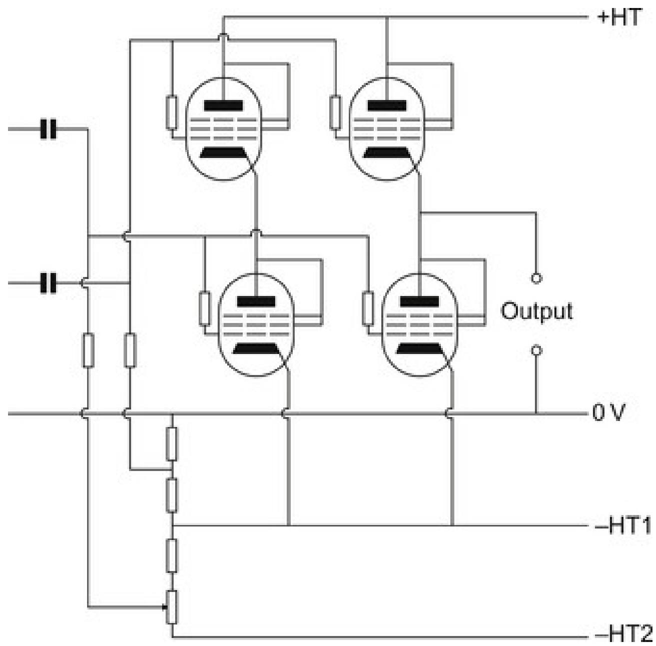


Figure 6.11 OTL output stage (paralleled White cathode followers).

These amplifiers are quirky in the extreme, yet some designers think that the problems of output transformers are so severe that they persist in making successful OTL amplifiers.

However, the one OTL application where no excuses need be made is headphone amplifiers. Not only do headphones tend to have a higher impedance than loudspeakers (32–300 Ω versus 4–8 Ω) thereby requiring less current, but the close coupling to the ear increases their apparent efficiency and further reduces the current required. The combination of these two factors makes an OTL headphone amplifier not only possible but also desirable, which accounts for the number of recent designs.

The Entire Amplifier

Having looked at the problems of the output stage, we can now consider the support circuitry in detail. We will mainly consider push–pull amplifiers, since they make up the majority of designs, although their design principles may well sbe perfectly applied to single-ended amplifiers. The output stage is insufficiently sensitive to be driven directly from a pre-amplifier, so it needs additional gain. If it is push–pull, it needs a phase splitter. Since linearity is unlikely to be ideal, we will probably need global negative feedback, which will further reduce gain, and this will need to be restored. A complete circuit might therefore comprise an input stage, a phase splitter, a driver stage and the output stage (see Figure 6.12).

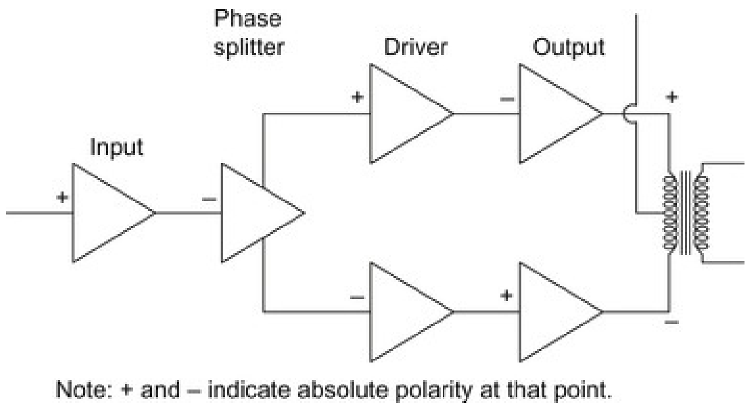


Figure 6.12 Block diagram of complete power amplifier.

Although a Class A output stage is a constant resistive load, a Class AB2 output stage heavily loads the driver stage when drawing grid current, and its driver stage would need very low output resistance and be able to source significant current to drive this load without distortion.

Unlike the output stage, and possibly the driver stage, the remainder of the stages in a power amplifier will be loaded by predictable resistive loads. It is therefore possible, and desirable, to design these stages with great care in order that they should not degrade the performance of the entire amplifier.

The Driver Stage

Unless the amplifier is quite low power, it will require a dedicated driver stage. We need good linearity, good output voltage swing and, unless we are prepared to add a cathode follower, low output resistance.

The differential triode pair is the ideal choice. The output stage probably requires about 25 V_{RMS} to each grid, and has an input capacitance of 40 pF, or more. An output resistance of 10 kΩ coupled to an input capacitance of 40 pF gives a high frequency cut-off of 400 kHz, which is perfectly acceptable. Since $r_a \approx R_{out}$, the high-μ valves, which tend to have a high r_a, are probably unsuitable.

In a properly designed power amplifier, the output stage should be the limiting factor, so we ought to design the driver stage to have at least 6 dB of overload margin. This requirement probably rules out our favourite valve, the E88CC. Very few commonly available valves satisfy our requirements (Table 6.1).

Table 6.1 Dual Triodes Suitable as Drivers

Type	r_a (kΩ)	Comments
SN7/N7	≤10	Lowest distortion
ECC82	≤10	13 dB worse distortion than *SN7*/N7
E182CC	≤5	Good on paper, but can sound strident
6BL7	≤3	Cripplingly high C_{ag}
6BX7	≤2	Capable of driving 845; robust

The *SN7*/N7 family is by far the most linear of the previous selection, and if the *SN7GTA or* SN7GTB versions are used, $V_{a(max)}$ =450 V. The octal 6BX7 and 6BL7 were designed for use as the field scan amplifier in televisions, but field scan amplifiers are required to be non-linear, so audio performance is variable. In a distortion test of thirty 6BX7 triode sections, the author found that distortion varied by a factor of 4 between samples and that very few envelopes contained a pair of low distortion triodes.

If ultimate performance is required, it is far better to use a pair of single valves and accept that more metalwork is required. The octal 6AH4 is a single triode quite similar to the dual 6BX7, whereas the noval 6S4A is more akin to the 6BL7. Having tested 224 6S4A, the author found that those made by GE were most likely to meet their specification, but Sylvania was the worst (Table 6.2).

Table 6.2 Single Pentodes Suitable as Drivers (When Triode Connected)

Type	r_a (kΩ)	Comments
EF184	≤5	Cheap and really plentiful, μ =60
N78	≤3	Uncommon, but dirt cheap
A2134, EL84	≤2	NOS EL84 extinct, but current Slovak production fine

Even lower driving resistance can be provided by an additional direct coupled cathode follower stage, which also has the advantage of buffering the differential pair from grid current effects (see Figure 6.13).

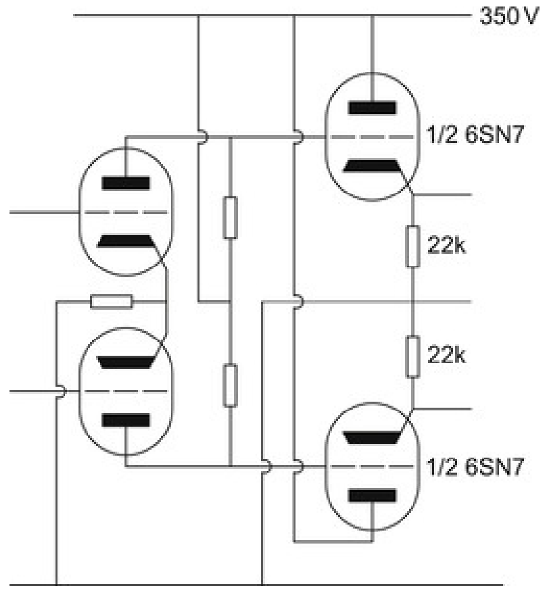


Figure 6.13 Driver stage using differential pair direct coupled to cathode followers.

The Phase Splitter

The driver stage is always preceded by the phase splitter, and traditionally the two stages have been combined–although as we shall see, this is not always a good idea. Design of the phase splitter is crucial to the success of a push–pull amplifier, so we will look at this in detail.

The phase splitter converts a single-ended signal into two signals of equal amplitude, but one has inverted polarity. There are three fundamental ways of achieving this goal (see Figure 6.14):

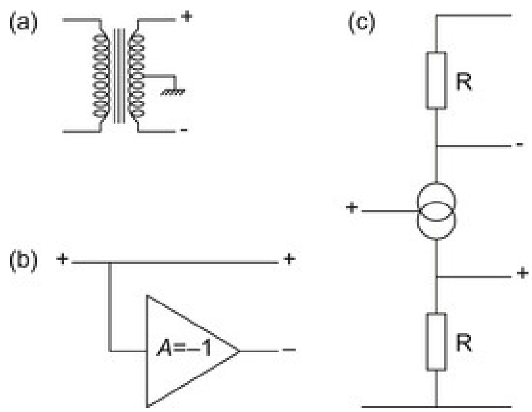


Figure 6.14 Fundamental basis of **all** phase splitters.

- We use a centre-tapped transformer in the same way that we use an output transformer to provide inverted and non-inverted signals. All of the previous considerations about transformers apply with a vengeance because of the comparatively high impedances involved, so the technique is not widely used, even though balance is near-perfect under all conditions (see Figure 6.14a).
- We have two outputs: one is the original signal, and the other is simply the input passed through an inverter (see Figure 6.14b).
- A valve controls the flow of current between two resistors, one of which is connected to ground and the other to HT. Increased current causes the instantaneous voltage dropped across each resistor to rise, so at any instant the voltage relative to ground is falling at one output, whilst the other is rising (see Figure 6.14c).

The third method is the basis of the concertina phase splitter. Although we will see shortly that the floating paraphase inverter can be a pure expression of the second technique, it is more usual for the second and third techniques to be combined in some form of cathode-coupled amplifier or differential pair.

Triode phase splitters having low resistance outputs are very sensitive to their loading and have different output resistances when both outputs are loaded compared to only one output being loaded. Phase splitters with a high r_a have an output resistance dominated by R_L , so phase splitters using pentodes and cascodes are immune to loading problems.

Loading sensitivity means that triode phase splitters should only be loaded by a stage that can be guaranteed never to draw grid current because the interaction between phase splitter loading sensitivity and drastically reduced input resistance due to grid current exacerbates distortion caused by grid current.

The Differential Pair and Its Derivatives

Most classic phase splitters were based on the differential pair, and much ingenuity was demonstrated in improving their small-signal performance.

A perfect differential pair comprises two devices having equal load resistances connected so as to allow a signal current to swing backwards and forwards between the two load resistances without any loss whatsoever. Loss of signal current from the cathode to ground impairs performance, so the tail resistance is crucial, and should ideally be infinite.

>RL Solution",5,0,2,1,105pt,105pt,0,0>The $R_k >> R_L$ Solution

A differential pair can be optimised with a pentode or cascode constant current sink in its tail. An EF184 pentode can achieve a tail resistance >10 MΩ, and even the larger pentodes, such as the EL83, can attain 1 MΩ unaided. Alternatively, we can use a cascoded semiconductor constant current sink, but we will always be limited at high frequencies by C_{kh} from the differential pair's cathode, even if the sink is perfect (see Figure 6.15).

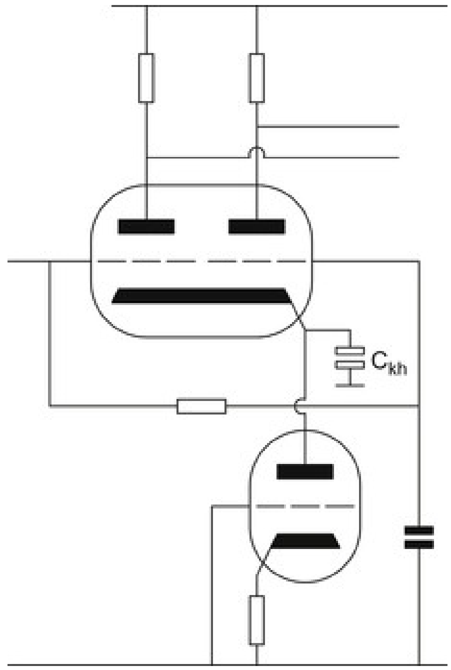


Figure 6.15 Differential pair with triode constant current sink as phase splitter.

The behaviour of the differential pair was discussed in Chapter 2 . Provided that $R_k \approx \infty$, a balanced output must be achieved if $R_{L1} = R_{L2}$. Output resistance must also be identical from both outputs, and $r_{out} = r_a \| R_L$, as before.

However, if only one output is loaded, then:

$$r_{out} \approx \frac{R_L(R_L + 2r_a)}{2(R_L + r_a)} \approx \frac{R_L}{2}$$

Loading one output heavily (driving grid current) is equivalent to leaving the other output unloaded, so we find ourselves in the unfortunate situation of driving a condition that requires low source resistance with increased resistance, which is why grid current distortion is exacerbated.

The $R_k \approx R_L$ Compensated Solution

We accept that we cannot easily achieve a high tail resistance, and do not even try. We use a resistor, typically between 22 kΩ and 82 kΩ, as a tail, calculate what the errors will be, and try to correct them. The principle is often known as the *cathode-coupled* or *Schmitt* phase splitter [6] because he was the first to analyse the errors and show how they could be corrected. Practical implementations modify the circuit slightly to accommodate external DC conditions, so the Leak TL series of amplifiers paid the price of biassing their input stage optimally by needing another low-frequency time constant to couple to the self-biased variant (Figure 6.16a), whereas the Mullard 5-20 used the elegant DC coupled Clare [7] variant (Figure 6.16b).

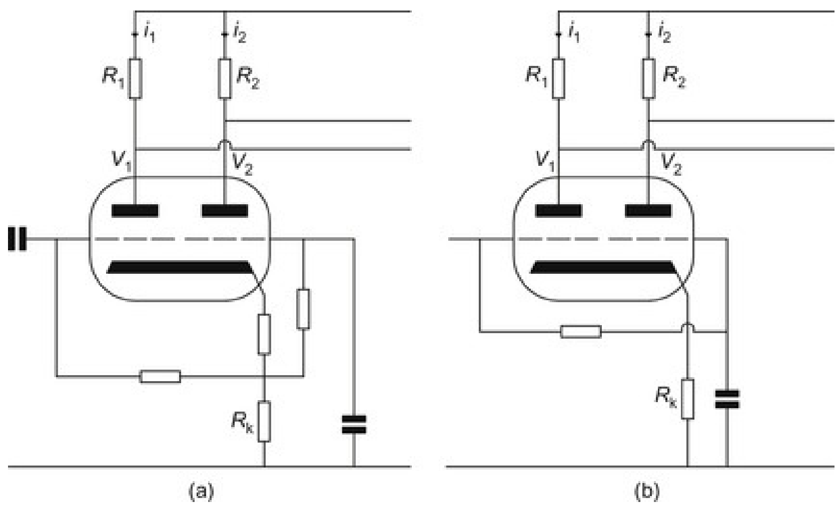


Figure 6.16 Schmitt cathode coupled phase splitters: Leak vs Clare variants.

V_2 can be considered to be a grounded grid amplifier, fed from the cathode of V_1. It is this use of the first valve as a cathode follower to feed the second valve that results in the apparent loss of gain of the second valve, since for a cathode follower, $A_v < 1$. By inspection, we see that $2\ v_{gk}$ is required to drive the stage, so the gain of the compound stage to each output is half what we would expect from an individual valve.

If the outputs are in balance, then $v_1 = v_2$, so:

$$i_1R_1 = i_2R_2$$

The gain of V_2 is A_2, so

$$v_{gk} = \frac{v_2}{A_2}$$

The signal current flowing in the cathode resistor is the out-of-balance output signal current:

$$v_{gk} = (i_1 - i_2)R_k$$

The signal at the output of V_2 must be:

$$(i_1 - i_2)R_k \cdot A_2 = i_2R_2$$

Expanding and collecting I_2 terms:

$$i_1 \cdot R_k \cdot A_2 = i_2(R_2 + R_k \cdot A_2)$$

Substituting $i_1R_1 = i_2R_2$, and simplifying:

$$\frac{R_2}{R_1} = \frac{R_2}{R_k \cdot A_2} + 1$$

This shows that unless the gain or the tail resistance of the stage is infinite, the ratio of the anode loads should be adjusted to maintain balance. Note that A_2 is the individual, unloaded, gain of V_2, and not the gain of the entire stage.

As an example, the Leak TL12+ phase splitter/driver was investigated. This uses an ECC81 and gives a gain of 42 for V_2 (R_2 =100 kΩ, μ =53, r_a =26.5 kΩ), R_1 should therefore be 91 kΩ, and this is exactly what Leak used (see Figure 6.17).

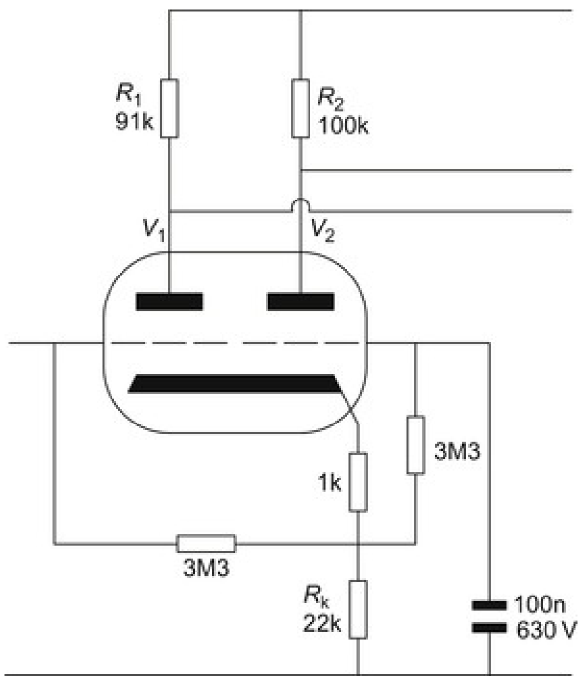


Figure 6.17 Cathode coupled phase splitter as used in Leak TL12+.

The output resistance of each half of the stage is slightly different because it is in parallel with a slightly different anode load, but curing this in order to preserve High Frequency balance upsets the voltage balance at low frequencies. The nearest approximation that we could achieve is to include the grid-leak resistances as part of the anode loads when calculating the necessary changes. The following grid-leak resistors are 470 kΩ, so R_{L2} =100 kΩ‖470 kΩ=82.46 kΩ, and the gain of V_2 falls to 40. The required total load for V_1 (including the 470 kΩ grid leak) is thus 75.7 kΩ, and R_{L2} =90.2 kΩ.

Low Frequency balance is determined by the time constant of the grid decoupling capacitor and its series resistor, since it cannot hold the grid of V_2 to

AC ground at very low frequencies.

Rather than tinker with resistor values whose calculation is critically dependent on valve parameters, the author would rather add a constant current sink to the cathode to force the stage into balance.

The R $_k$ << R $_L$ High Feedback Solution

We make no attempt at providing a large value of tail resistor, and rely on feedback to maintain balance. This circuit was devised by *Carpenter* [8], but is also known as the *floating paraphase* or *see-saw* phase splitter. Typically, the design uses a high- μ valve, such as the ECC83, from whose data sheet this circuit was taken (see Figure 6.18).

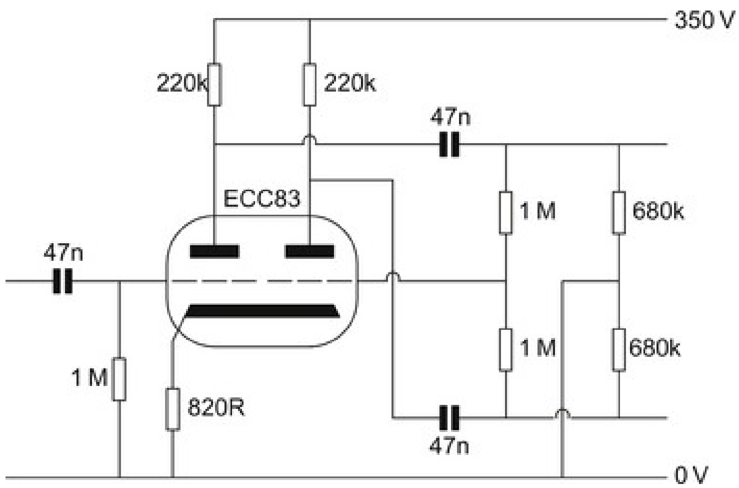


Figure 6.18 Carpenter floating paraphase or see-saw phase splitter.

If we redraw the circuit (making it identical to Figure 6.6 of Carpenter's patent), we see that V_2 is simply a unity gain inverter, whose gain is defined by resistors R_1 and R_2 (see Figure 6.19).

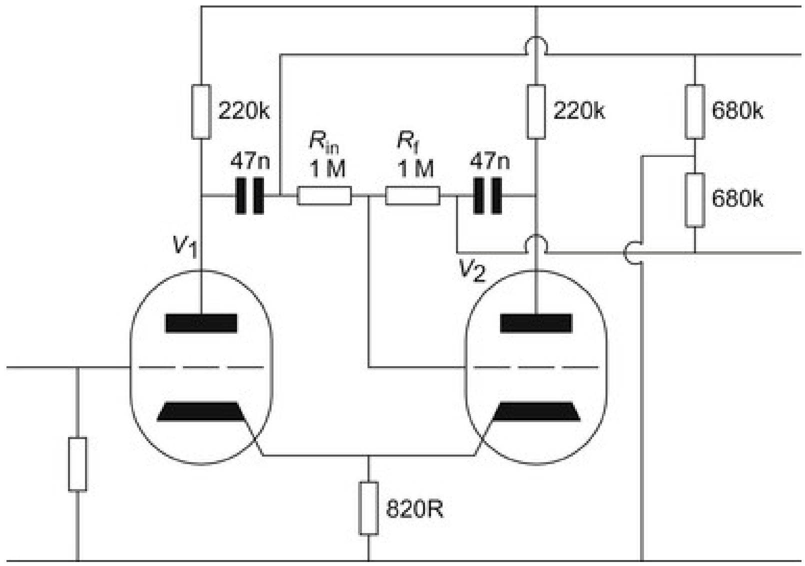


Figure 6.19 Carpenter floating paraphase phase splitter redrawn to reveal invertor.

Since the open-loop gain of V_2 is not infinite, these values must be adjusted to give a gain of –1. Unfortunately, the calculation is further complicated by the fact that R_2 affects the loading, and open-loop gain, of the stage. V_2 also requires a build-out resistor to equalise its output resistance, which has been significantly reduced by negative feedback. Once these corrections have been made, the balance of this phase splitter is good, since the operation of V_2 is stabilised by negative feedback.

The circuit is analysed by first drawing a DC loadline to correspond to the 220 kΩ anode load. The Mullard operating point is at V_a =163 V. The 1 MΩ feedback resistor is in parallel with this at AC, so we draw an AC loadline through the operating point corresponding to 180 kΩ. From this we find that the AC gain of the valve is 67.

We need to find the value of β that will give final gain of 1, using:

$$\beta = \frac{1}{A} - \frac{1}{A_0}$$

We find that β =0.985. The easiest way to achieve this is to increase the value of the feedback resistor:

$$R'_f = \frac{R_f}{\beta}$$

This gives a value of 1,015 kΩ, so we would add 15 kΩ in series. So far, we have only discovered a 1.5% error, which is trivial, but if we consider the output resistances we find a much larger error. The output resistance of V_1 is r_a in parallel with R_L, and is ≈53 kΩ, but the output resistance of V_2 has been reduced by a factor of $(1+\beta A_0)$, from 53 kΩ to ≈790 Ω. 52.2 kΩ of build-out resistance is therefore required, but the nearest standard value of 51 kΩ would be fine. These outputs are then each loaded by 680 kΩ, and if corrections are not made, the output from V_2 is ≈6% high.

In practice, these corrections were never made, which perhaps accounts for the poor reported performance of the stage. It might be thought that the cathode coupling would improve balance, but V_2 has such heavy feedback that if inadvertently set to non-unity gain (wrong ratio of gain-defining resistors) it can easily overcome any self-balancing action generated at the cathode. The justification for cathode coupling has little to do with improving balance and much more to do with eliminating a pair of undesirable cathode bypass capacitors.

The Concertina Phase Splitter

The phase splitters based on the differential pair were all able to provide overall gain, but once $R_k < \infty$, output balance became partially dependent on the matching of μ between the valves.

Feedback causes the concertina phase splitter to have slightly less than unity gain to each output, and its output balance is almost totally determined by passive components, so valve characteristics hardly enter the picture. Conceptually, operation is very simple. Modulation of grid voltage causes a signal current to flow in the valve, the anode and cathode loads are equal and they have the same current flowing through them, so the signals generated across them are equal, implying perfect balance (see Figure 6.20).

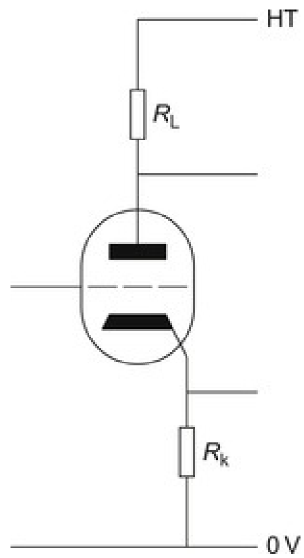


Figure 6.20 Concertina phase splitter.

Gain of the Concertina

The gain of the concertina to its anode may be found using the standard triode gain equation, but noting that all undecoupled resistances down to ground via the anode resistance are multiplied by a factor of (μ +1), then:

$$r'_a = (\mu + 1)R_k + r_a$$

$$A = \frac{\mu R_L}{r_a + (\mu + 1)R_k + R_L}$$

But for the concertina, $R_k = R_L$, so:

$$A = \frac{\mu R_L}{R_L(\mu + 2) + r_a} \approx 1$$

Because of this low value of gain to the anode, Miller capacitance is also low, and the stage has wide bandwidth. Since there is no grid current, anode current is identical to cathode current and applying Ohm's law to the equal anode and cathode load resistances shows that gain to the cathode is equal to gain to the anode.

Output Resistance with Both Terminals Equally Loaded (Class A1 Loading)

The concertina is a special case ($R_k = R_a$) of an unbypassed common cathode amplifier with outputs taken from both anode and cathode. The general feedback equation is:

$$A = \frac{A_0}{(1 + \beta A_0)}$$

The denominator of the feedback equation is the factor by which resistances are changed and is known as the *feedback factor* . Since we know the gain of the concertina and the gain of a simple triode amplifier, we can substitute them into the feedback equation to solve for the feedback factor:

$$\frac{\mu R_L}{R_L(\mu + 2) + r_a} = \frac{(\mu R_L / R_L + r_a)}{\text{feedback factor}}$$

Cross-multiplying to find the feedback factor:

$$\text{Feedback factor} = \frac{R_L(\mu + 2) + r_a}{R_L + r_a}$$

The anode output resistance of a common cathode triode amplifier with no feedback is:

$$r_{out} = \frac{R_L r_a}{R_L + r_a}$$

The feedback works to reduce anode output resistance, so this value must be divided by the feedback factor (practically, we multiply by the inverse of the feedback factor):

$$r'_{out} = \frac{R_L r_a}{R_L + r_a} \cdot \frac{R_L + r_a}{R_L(\mu + 2) + r_a}$$

The ($R_L + r_a$) terms cancel, leaving:

$$r'_{out} = \frac{R_L r_a}{R_L(\mu + 2) + r_a}$$

Initially, it seems most surprising that series feedback ($R_k = R_a$, after all) should *reduce* output resistance from the anode so that $r_{out} \approx 1/g_m$, but this can be understood by considering an external capacitive load on each output. In the same way that $R_k = R_a$ defines a gain of 1 at low frequencies, so $X_{C(k)} = X_{C(a)}$ defines a gain of 1 at high frequencies, and changing this ratio of capacitances certainly would change the gain, or frequency response at high frequencies, since it would change the feedback ratio β .

Because $Z_k = Z_a$, the frequency response at each output is forced to be the same, so the output resistances must also be equal, and $r_{out(k)} = r_{out(a)}$.

Concertina Output Resistance, Only One Output Loaded (Class B Loading)

Looking into the cathode, we see R_k down to ground, in parallel with r_k the anode path to ground:

$$r_k = \frac{R_a + r_a}{\mu + 1}$$

Substituting:

$$r_{out\,(cathode)} = \frac{R_k(R_a + r_a/\mu + 1)}{R_k + (R_a + r_a/\mu + 1)}$$

Simplifying, and noting that $R_a = R_k = R_L$:

$$r_{out\,(cathode)} = \frac{R_L(R_L + r_a)}{R_L(\mu + 2) + r_a}$$

Alternatively,

$$r_{out\,(cathode)} = \frac{R_L + r_a}{(\mu + 2) + r_a/R_L}$$

In a practical application, the (μ +2) term is usually significantly larger than the r_a / R_L term, so to a reasonable approximation:

$$r_{out\,(cathode)} \geq \frac{R_L + r_a}{\mu + 2}$$

Although the cathode output resistance can be calculated fully, the approximation is usually good enough, and generally gives a value of about 1 kΩ.

Looking into the anode, we see R_a to HT, which is AC ground, in parallel with the cathode path to ground:

$$r'_a = R_L(\mu + 1) + r_a$$

Substituting

$$r_{out\,(anode)} = \frac{R_L[R_L(\mu + 1) + r_a]}{R_L(\mu + 1) + r_a + R_L}$$

Tidying terms

$$r_{out\,(anode)} = \frac{R_L^2(\mu + 1) + R_L \cdot r_a}{R_L(\mu + 2) + r_a}$$

If we inspect this equation closely, we see that the terms involving μ are the only significant terms, and that if μ is reasonably large, then (μ +1)≈(μ +2), so that $r_{out} \approx R_L$. Thus, the concertina suffers a similar weakness to phase splitters based on the differential pair in that one output driving grid current is equivalent to

leaving the other unloaded. However, if the cathode output drives grid current, its output resistance remains low, whereas if the anode output drives grid current, its output resistance rises significantly. This imbalance implies that concertina distortion due to driving grid current will contain additional even harmonic components compared to phase splitters based on the differential pair.
It is usual to direct couple to the anode of the input stage, and let that determine the DC conditions of the concertina, resulting in the saving of a coupling capacitor and a low-frequency time constant. Although the concertina is frequently criticised for its unity gain to individual outputs, this is a differential gain of 2, so the combination of input stage and concertina gives double the gain of the same 2 valves used as a phase splitter based on the differential pair.

Phase Splitters and Class B Output Stage Miller Capacitance

If a phase splitter drives a Class B output stage, only one output valve is switched on at any given time, suggesting that there could be no Miller capacitance from the switched-off valve, thereby unbalancing the phase splitter. However, the output transformer acts as an autotransformer, so the valve's anode still swings and its C_{ag} still has a large voltage across it, causing the charging current that manifests itself as Miller capacitance *despite* that valve being switched off, so it is not necessary to buffer the phase splitter.
Obviously, the autotransformer argument does not apply to OTLs, but because OTLs have a cathode follower output stage, its input capacitance would be expected to be low. However, because the cathode follower drives such a low impedance load, its gain is very low, so it cannot bootstrap C_{gk}, resulting in a higher input capacitance than would otherwise be expected. Since OTL output stages are almost inevitably Class AB, on one half-cycle the load is driven by a cathode follower and on the other by an anode follower (common cathode amplifier). The two different amplifiers have entirely different input capacitances ($C_{ga} \neq C_{gk}$), suggesting that buffering the phase splitter could be worthwhile.

The Input Stage

The input stage is where global negative feedback is applied, so it must provide an inverting and a non-inverting input, both with low noise. The triode differential pair is an obvious candidate for this stage, but the common cathode triode or pentode can also be used, in which case global negative feedback is habitually applied to its cathode (see Figure 6.21).

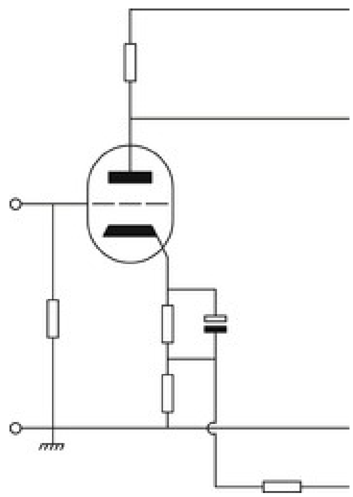


Figure 6.21 Application of global feedback at the input stage.

Design of the input stage is fairly trivial, but can be slightly complicated by the usual practice of direct coupling to the phase splitter, which restricts choice of anode operating conditions.

Stability

When we looked at RC networks in Chapter 1 , we saw that a single RC network tended towards a maximum phase shift of 90°. To make an oscillator, we need 180° of phase shift, so a single stage amplifier with one RC network causing an Low Frequency or High Frequency cut-off *cannot* oscillate. If we cascade two such stages, we can approach 180° of phase shift, and if we feed this back into the input, it will ring, but not oscillate. If we have three such stages, it is a racing certainty that we can make the cascade oscillate when we apply feedback, and this is the basis of the phase shift oscillator.

To achieve oscillation, we need more than phase shift. Just because our feedback signal's phase has been shifted by 180° will not necessarily generate oscillation. We also need sufficient *loop gain* . The basis of oscillation is that it is self-sustaining; the gain of the amplifier must be sufficiently high to overcome the losses in the feedback loop before oscillation can occur. Loop gain is thus defined as the gain of the amplifier multiplied by the loss of the feedback loop.

If we have a phase shift of 180° *and* loop gain ≥1, the circuit *will* oscillate, and this is known as the Barkhausen criterion.

Now that we have this criterion, we can see how to avoid designing oscillators. We have two weapons at our disposal:

- We can reduce the number of stages, such that phase shift never reaches 180°. We rarely achieve this ideal, because the output transformer plus output stage plus driver stage harbours so many phase shifts, but the principle of minimising the number of stages within a feedback loop is still valid.
- We attack the second condition of the oscillator statement, and reduce loop gain to ≤1 at the troublesome frequencies. This is the basis of all the methods that you will see for stabilising amplifiers, and it is a powerful weapon capable of oppressing anything. Whether the resulting amplifier is of any use may be more debatable.

Slugging the Dominant Pole

This rather vibrant and mystifying description is actually very simple.

A pole in electronic jargon is simply another way of saying 'high frequency cut off'. What we aim to do is to make the amplifier *look* as if it only has one High Frequency cut off, which, with a maximum phase shift of 90°, is unconditionally stable. We look for the RC network with the lowest High Frequency cut-off frequency, *i.e.* the dominant one, and we *slug* it with yet more capacitance to make it even lower.

Suppose that as a worst case, we cascaded four identical amplifiers, each with an High Frequency cut-off frequency of 300 kHz, and a gain of 10. At 300 kHz, each amplifier contributes a phase shift of 45°, making a total shift of 180°. The gain of each amplifier is 3 dB down at 300 kHz, so the gain of each amplifier at that frequency is $10/\sqrt{2} = 7.071$, so the gain of the total amplifier must be:

$$A_{\text{total}} = \left(\frac{10}{\sqrt{2}}\right)^4$$
$$A_{\text{total}} = 2{,}500$$

In a typical amplifier we might want to reduce this gain from 2,500 to 125, which would be 26 dB of negative feedback, and would reduce distortion to a twentieth of its original value. In order to do this, the feedback loop would have a loss of 0.0076. If we now check for stability, 0.0076×2,500=19. The amplifier has a loop gain ≥1 *and* a phase shift of 180°, so it will oscillate.

We need to reduce the open-loop gain at 300 kHz by a factor of 19, or 25.5 dB, to achieve stability (this has little effect on the frequency response of the final amplifier). Remembering that 6 dB/octave is equivalent to 20 dB/decade, reducing *one* cut-off from 300 kHz to 30 kHz will give us 20 dB reduction, and halving from 30 kHz to 15 kHz will give us another 6 dB, making 26 dB in total. This procedure may be formalised by the following statement:

The loop gain may be as large as the ratio of the two most dominant time constants.

To apply this rule, we simply choose how much feedback we want, calculate the loop gain, and adjust the dominant time constant until the ratio between it and the adjacent time constant is equal to the loop gain.

The amplifier is now stable, but only just, and it *will* ring. We should distance the dominant time constant still further to increase stability and remove ringing, certainly by a factor of 2, and preferably a little more. It should be realised that over-zealous stability compensation reduces feedback, and compromises distortion reduction.

Most practical amplifiers, having exhausted the first two possible methods of achieving stability described, resort to manoeuvring the amplitude response independently of phase response using *step networks* , usually one inside the feedback loop to manoeuvre the open-loop response and another on the feedback loop to manoeuvre the closed-loop response. Such networks are adjusted on test whilst observing a square wave and typically require simultaneous adjustment of four components (not as hard as you might think).

Realistically, we ought to consider that methods of designing for stability, in priority order, are:

- Reduce the number of time constants or stages within the loop.
- Slug the dominant pole.
- Fudge the phase/amplitude response using step networks.

There are some stability problems that are peculiar to valve amplifiers, and they have well-known symptoms and cures.

Low Frequency Instability, or Motorboating

This is an oscillation at about 1 or 2 Hz, and is invariably caused by unintentional feedback via the power supply, due to the rising impedance of filter capacitors at low frequencies. In effect, the entire amplifier becomes a relaxation oscillator [9] . The traditional cure was to insert an low frequency step network, or to reduce the value of the coupling capacitors, *in the signal path* , so as to reduce the loop gain. This solution molests the second condition of the Barkhausen criterion, but only treats the symptoms.

The *real* solution is to attack the first condition by *removing* the filter capacitors and their associated RC time constants by fitting HT regulators. This generally kills the problem stone dead. It is this improvement in stability that is the reason for the superior bass in designs using regulated supplies, since it removes previously unidentified low frequency ringing. It should be noted that this problem does not necessarily need a global feedback loop to make it felt, and 'zero feedback' pre-amplifiers are not immune.

It is not unusual to discover that not only is the amplifier motorboating, but that it also has bursts of high frequency oscillation, known as *squegging* . If possible, it is best to cure the high frequency problem first, since it indicates marginal stability when the amplifier is under maximum stress and may be concealed once the low frequency instability has been cured.

It *is* possible to build a power amplifier with only two low frequency time constants – one in the output transformer and one in the driver circuitry – but it is so tempting to allow more than one in the driver. We saw earlier that we can have as much loop gain as the ratio of the two nearest time constants, so if we have three time constants that are not widely separated, we must place the middle time constant at the geometric mean of the outer two. Unfortunately, bandwidth considerations force the middle time constant to be the output transformer which, having an iron core, has inductance that varies with level, so its time constant is variable, implying that it is always moving towards one or other of the outer two time constants, reducing stability.

We can now see that the temptation to allow more than one time constant in the driver must be resisted at all costs if Low Frequency stability is to be achieved, so we must now consider whether τ_{driver} should be larger or smaller than $\tau_{transformer}$:

- $\tau_{driver} < \tau_{transformer}$: This can be achieved by fitting a small coupling capacitor, but this forces our global negative feedback to correct falling open-loop low-frequency response rather than reducing low-frequency distortion generated in the output transformer.

- $\tau_{driver} > \tau_{transformer}$: We know that the output transformer inductance falls as saturation approaches or, to put it another way, it has its largest time constant at low level – which is where it is easiest to test for stability. Thus, there are two advantages to making $\tau_{driver} > \tau_{transformer}$:

  - Stability increases as output transformer inductance falls.

  - Global negative feedback now reduces output transformer distortion rather than correcting falling open-loop response.

Unfortunately, there is a disadvantage. We cannot tolerate blocking at any price because if it occurs it will be prolonged. Thus, the $\tau_{driver} > \tau_{transformer}$ solution requires that τ_{driver} be implemented before a stage that cannot be overloaded. This is not a problem, but it *must* be consciously considered.

Parasitic Oscillation and Control Grid-Stoppers

The solution is almost encompassed by the description. Parasitics are the unwanted stray capacitances and inductances that result from the practical attempt to build a component or amplifier.

Miller capacitance in the valve combines with series inductance in the control grid circuit to form a resonant circuit, so valves with a high g_m (low r_k) are particularly prone to oscillation. (Inductance in the cathode circuit is not a problem because it causes negative feedback that reduces loop gain.) The best cure is to damp the resonant circuit by fitting a *grid-stopper* resistor in series with the grid *as close as possible* to the grid pin of the valve base. The physical positioning of the resistor reduces the wire's inductance (very roughly 0.75 nH/mm), whilst a given value of resistance in the grid circuit is far more effective at increasing the loss of the resonant circuit without compromising frequency response than if it is placed in the cathode circuit. Remembering that:

$$Q = \frac{1}{R}\sqrt{\frac{L}{C}}$$

Thus, Q is most dependent on series resistance, so adding 10 kΩ of series resistance to the grid circuit is a sure-fire way of oppressing parasitic oscillation. For small-signal valves that are prone to this problem (E88CC, 5842, EC8010) a thin film surface mount resistor actually touching the pin is ideal, whereas leaded carbon film resistors are more convenient for power valves. Typical values range from 100 Ω to 10 kΩ, and are usually found by experiment since individual layout is critical. If you can guarantee >100 MHz bandwidth at your probe tip, use an oscilloscope to determine the smallest acceptable value because excessive resistance increases grid current distortion, but beware that touching the typical 8 pF capacitance of a probe to a marginally stable circuit can sometimes tip the balance between stability and oscillation, so probe at the anode. If you can't probe, play safe and use a large grid-stopper because unseen RF oscillation always increases distortion.

Parasitic Oscillation of Ultra-Linear Output Stages, and g_2 Stoppers

Ultra-linear amplifiers with poor output transformers or parallel pairs of output valves sometimes need a series RC network between anode and g_2. This is because this section of the winding has resistance in series with leakage inductance – the additional network attempts to return this impedance to a pure resistance. For 43% taps, the impedance between the anode and g_2 tap is≈9% of the total anode to anode impedance, so this is a good starting point for the value of resistor, but both have to be determined empirically (adding them to both halves of the transformer simultaneously, which makes life difficult), and are often around 1 nF and 1 kΩ. Bear in mind that each capacitor has to withstand V_{HT} when the associated anode swings towards 0 V.

Parasitic Oscillation and Anode Stoppers

Now that NOS audio valves are in short supply, it is common to substitute cheaply available RF valves, but (being designed for RF) these can sometimes oscillate at high frequency or VHF. A common amateur radio solution was to add a series anode stopper inductor made from a 2 W 100 Ω carbon resistor overwound with about 10 turns of ≈0.7 mm enamelled copper wire soldered to each end of the resistor – the inductor prevents the valve seeing the capacitance that made it oscillate and the carbon resistor damps the inductor's VHigh Frequency self-resonance. Exactly the same trick is used at the output of transistor power amplifiers to protect the amplifier from excessive loudspeaker cable capacitance that could cause High Frequency instability and consequent destruction, and it was the absence of this component that caused the wholesale destruction of some well-known amplifiers when low inductance high capacitance loudspeaker cable became briefly fashionable.

High Frequency Stability and the 0 V Chassis Bond

Although we tend to think that we require a low resistance bond at the input socket between the amplifier's 0 V and the chassis to minimise hum, it also needs to be low inductance to maintain High Frequency stability. There is inevitably stray capacitance between each and every part of the circuit and the chassis. Thus, there will somewhere be two capacitors in series that connect an amplifier's output back to the input *in phase* , potentially causing oscillation. Connecting the chassis to 0 V short circuits the junction of those two capacitors to 0 V and breaks that positive feedback loop.

Stability Margin

Stability of valve amplifiers is often described by the amount of additional global negative feedback that would be required to cause oscillation. This is simply the ratio by which the dominant time constants were further distanced, over and above the necessary minimum required for stability. The designers of the Mullard 5-20 were proud to claim that 10 dB more feedback would be required to cause instability, whereas the Williamson is questionable at Low Frequency even without additional feedback.

Unlike a transistor amplifier, it is unusual to be able to apply more than 30 dB of global negative feedback to a valve amplifier, and even then design needs to be planned very carefully to ensure that the feedback reduces distortion rather than stability.

Classic Power Amplifiers

Now that we can recognise and analyse individual stages, we can investigate the design of some classic amplifiers such as the Williamson, the Mullard 5-20 and the Quad II.

The Williamson

The design of this amplifier was published in *Wireless World* [10] in 1947, and set a standard of performance that was years ahead of its time.
The input stage is the standard common cathode triode with 20 dB of global negative feedback applied from the loudspeaker output to the cathode. The phase splitter is a concertina circuit direct coupled from the input stage, and feeds a differential pair using both halves of a 6SN7 (see Figure 6.22).

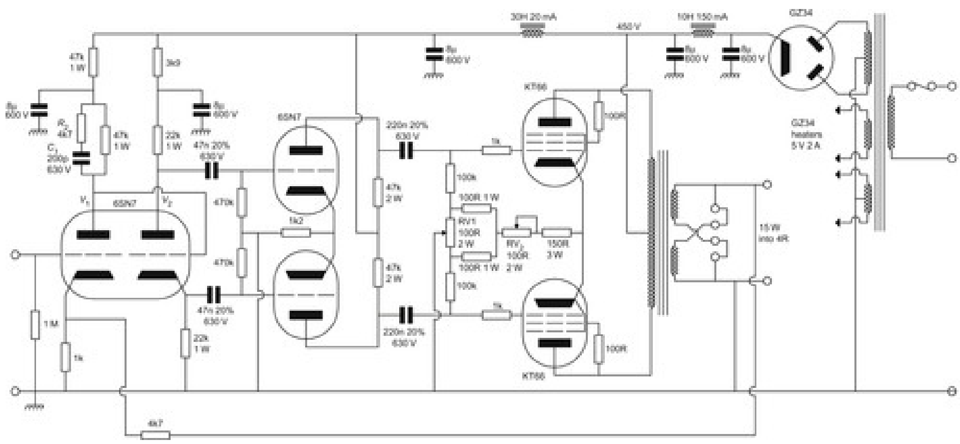


Figure 6.22 Williamson amplifier
(by permission from Electronics World).

The output stage is a push–pull pair of KT66 beam tetrodes operated as triodes that provide 15 W output in Class AB1, operating mostly in Class A. RV_1 adjusts the DC balance of the output valves in order to minimise distortion due to the transformer core, whilst RV_2 sets the quiescent current to 125 mA for the entire stage.
The linearity and headroom of each stage is excellent due to the careful positioning of operating points and choice of valves, but because this amplifier has four stages enclosed by the feedback loop, stability needs to be taken very seriously.
The input stage initially has an output resistance of ≈7.5 kΩ, but this is raised by the feedback to ≈47 kΩ. In combination with 12 pF of input capacitance from the concertina, this gives a high frequency cut-off of ≈280 kHz. However, this

has been modified by adding the step compensation components R_2 (4.7 kΩ) and C_1 (200 pF) to the anode circuit of V_1. This circuit puts a step in the amplitude response which begins to fall at ≈130 kHz, but the phase response remains virtually unchanged until 280 kHz.
The concertina drives a driver stage with an input capacitance of 60 pF, and because the output transformer for the Williamson was very carefully specified, it seems unlikely that losses in the output transformer would cause the global feedback loop to force the driver stage out of Class A operation. The concertina, thus, faces a balanced load, and has an output resistance ≈350 Ω, resulting in $f_{-3\ dB}$=7.5 MHz, which is sufficiently high to be insignificant.
The driver stage has an output resistance of ≈8.7 kΩ, together with 55 pF of input capacitance from the output stage, the cut-off is ≈330 kHz, and the output transformer is specified to have a cut-off of 60 kHz.
The number of high frequency cut-offs within the feedback loop has not been minimised, and the dominant high frequency cut-off (the output transformer) is rather close to the pair which are next most dominant. Thus, the only remaining way to achieve stability at high frequency was to adjust the phase response independent of amplitude response by means of a step network.
At low frequencies it is more useful to consider time constants than –3 dB points. The input stage is direct coupled to the concertina, so we can ignore this. The concertina feeds the driver stage with a CR of ≈22 ms, as does the driver to output stage, and the output transformer is set to 48 ms. Additionally, the author's experience has been that decoupling the HT between input stage and concertina can sometimes induce motorboating. In view of this, it is not surprising that Low Frequency stability is questionable, as was conceded in the original *Wireless World* article. In 1952, Hafler and Keroes decided that their output stage would benefit from a Williamson driver [11], so they quintupled the concertina to driver stage coupling capacitors from 50 nF to 0.25 μF to separate the low-frequency time constants.
It should be remembered that in 1947, circuits were designed using long multiplication or tables of logarithms, and if speed was needed – slide rules. Computer-aided AC analysis was not an option! Most amplifiers were designed as carefully as possible, then adjusted on test for best response – and wide-bandwidth (>1 MHz) oscilloscopes were recently developed luxuries.

The Mullard 5-20

This was a 20-W design introduced by Mullard [12] to sell the EL34 pentode. There is considerable similarity between this design, the Mullard 5-10 (10 W using EL84), and some Leak amplifiers (see Figure 6.23).

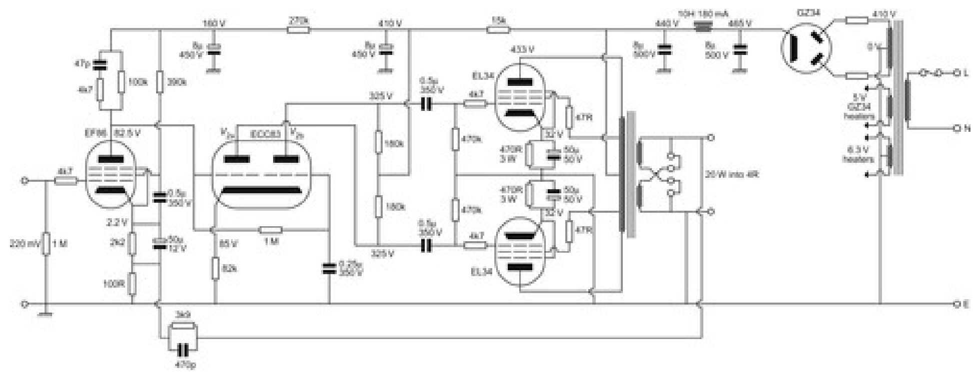


Figure 6.23 Mullard 5-20 (Mullard Ltd, originator of this design is now included in Philips Components Ltd).

The input stage is an EF86 pentode, which is responsible for the high sensitivity but poor noise performance of these amplifiers. Most of the cathode bias resistance is bypassed, since it would otherwise reduce the gain from around 120 to 33, which would be a waste of open-loop gain that could be used to correct distortion produced by the output stage. Unadorned, the pentode has an output resistance of 100 kΩ, and drives ≈50 pF of input capacitance from the phase splitter, which would give a cut-off of 32 kHz, but this is modified by the usual step network across its anode load.

A slightly unusual feature is that the g_2 decoupling capacitor is connected between g_2 and cathode, rather than g_2 and ground. In most circuits, the cathode is at (AC) ground, and so there is no reason why the g_2 decoupling capacitor should not go to ground. In this circuit, there is appreciable negative feedback to the cathode, and so the g_2 capacitor must be connected to the cathode in order to hold g_2–k (AC) volts at zero, otherwise there would be *positive* feedback to g_2.

The cathode-coupled phase splitter is combined with the driver circuit using an ECC83. When loaded by the output stage, for V_2, A_v =54, but gain to one output is half this at 27.

The anode load resistors have not been modified to give perfect balance. With the 470 kΩ grid-leak resistors of the output stage in parallel with the 180 kΩ anode loads, the *effective* anode load is 130 kΩ. Using the formula derived earlier, this means that V_{2b} should have an AC anode load 3% higher than V_{2a}, and R_L for V_{2b} would then be 187 kΩ. Mullard did actually state this [12] , but probably assumed that most constructors would not have access to sufficiently close precision resistors to use the information.

The output stage has an input capacitance of ≈30 pF, and the driver stage has an output resistance of 53 kΩ when loaded symmetrically, giving a cut-off at ≈100 kHz, which is quite poor. Loaded asymmetrically, the output resistance rises to ≈90 kΩ, which lowers the cut-off to ≈60 kHz.

Looking at the driver stage, we should investigate how capable it is of driving the output stage. 85 V will be wasted across the 82 kΩ tail resistor, but with 410 V of HT, this still leaves us with 325 V. With the component values given, this puts the operating point at 240 V on the 180 kΩ DC loadline. Drawing the AC 130 kΩ loadline through this point shows that the stage would generate ≈4% second harmonic distortion at full drive (V_{out} =18 V_{RMS}), if it were not operated as a differential pair. Mullard claimed 0.4% distortion for the entire driver circuitry.

Although distortion appears acceptable, the driver stage has only 10 dB of overload capability. When output stage gain begins to fall due to cathode feedback or insufficient primary inductance in the output transformer, or input capacitance loads the driver, the global feedback loop will try to correct this by supplying greater drive to the output stage, and the 10 dB margin will be eroded, increasing distortion.

The driver circuitry was designed to produce an amplifier of high sensitivity even after 30 dB of feedback had been applied, and this has forced other factors to be compromised. Whereas the Williamson sacrificed stability for linearity, the Mullard 5-20 achieves stability at the expense of noise and linearity.

The output stage is a pair of EL34s in Blumlein distributed load configuration, with 43% taps for minimum distortion. Unlike the Williamson, there is no provision for adjusting or balancing bias, and this might seem to be a retrograde step.

Bias adjustment implies connecting the cathodes together and using a proportion of grid bias to provide the balance adjustment. Because the biassing is firmly set by the potentiometers, there is no self-regulation of bias current, and as the valves age, balance will need to be reset. In short, providing this adjustment ensures that it has to be used regularly.

By contrast, the Mullard 5-20 has separate cathode bias resistors and relies on automatic bias to hold the anode currents at their correct, and therefore equal, currents. In practice, this works well, although it does not quite achieve the low transformer core distortion of a freshly balanced adjustable system.

This system does have a disadvantage in that the individual cathode bias resistors apply series negative feedback to the output valves, raising their output resistance. The output transformer could be redesigned to maintain the match to the load, but this is undesirable as it would require a higher primary impedance, which makes a high quality design more difficult to achieve. Because of this, the cathode bias resistors must be bypassed by capacitors, and this is where the problems really begin.

The capacitor is a short circuit to AC, and so prevents feedback, but its reactance rises at very low frequencies, so it is no longer a short circuit, and allows feedback. Because the output stage is matched to the load, the unwanted feedback causes a sharp rise in distortion and reduction of output power due to the mismatch. The obvious solution is to fit a large enough capacitor to ensure that the Low Frequency cut-off for this combination is below all frequencies of interest, perhaps 1 Hz. Remembering that the resistance that the capacitor sees is R_k in parallel with r_k, we can easily calculate the value required.

For a pentode, $r_k = 1/g_m$; a typical output pentode has $g_m = 10$ mA/V at its working point, so $r_k \approx 100$ Ω, which is in parallel with a bias resistor of ≈300 Ω, giving a total resistance of 75 Ω. For 1 Hz, we therefore need 2,000 μF of capacitance.

2,000 μF 50 V capacitors were simply not available at the time, and they weren’t fitted. They are readily available now, but there’s a more subtle reason for using a smaller value. When the output stage is driven into Class B, each cathode tries to move more positively than negatively. It can't turn off any further, but it can certainly turn *on* harder. The cathode capacitor integrates these changes into a gently rising DC bias voltage, which biasses the valve further into Class B, and the problem continues. The effect of this is that a momentary overload can cause distortion of following signals that would normally have been within the capabilities of the amplifier – this is a form of blocking. As the cathode bias capacitor becomes larger, the recovery time from overload lengthens. Theoretically, we never overload amplifiers, and this is not a problem, but occasional overload is inevitable, and its effects should be considered.

One way to deal with this problem is to reduce the cathode bias resistor to ≤1 Ω, so that it no longer causes noticeable feedback, and measure the current through it using an operational amplifier. This then feeds an asymmetric clipper so that when the valve strays into Class B and clips one half-cycle, the clipper removes an equal amount from the other half-cycle before feeding the processed signal to an integrator. The integrator can have an RC time constant of almost any value we choose, and 10 s is not unusual. The output of the integrator is a smoothed DC voltage proportional to anode current, which can be compared to a fixed reference, and the difference between the two levels drives an amplifier whose output sets the negative *grid* bias for the output valve.

If the anode current of one valve is set as a reference, then the other valve, or valves, can share this reference, which then forces anode currents into balance. The increased complexity of this scheme is (partly) offset by its improved performance and reduction in HT voltage required, since the cathode bias scheme wastes HT (see Figure 6.24).

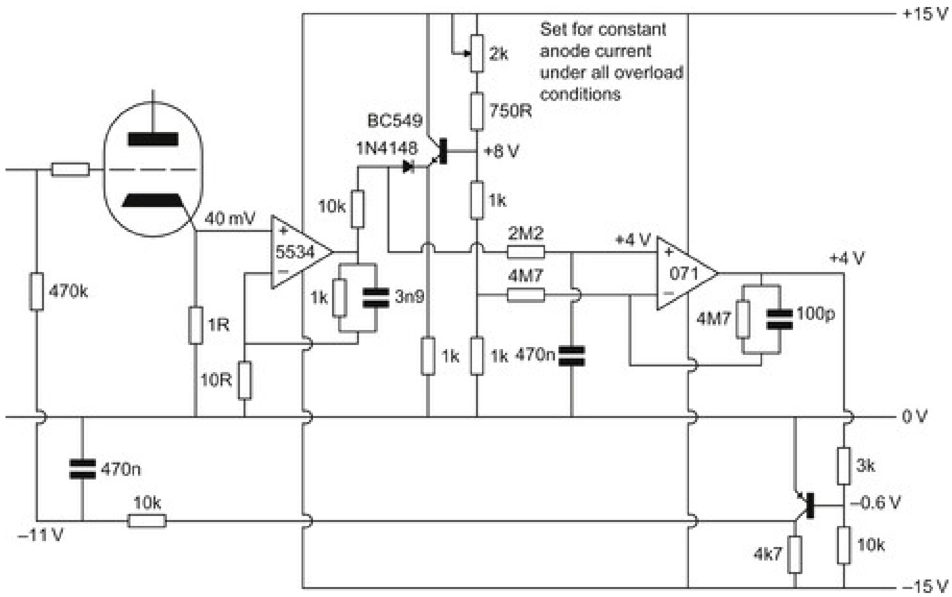


Figure 6.24 Principle of output bias servo.

This circuit was designed to sense a 40 mA anode current by developing 40 mV across the 1 Ω resistor; the rest of the circuit is based on this 40 mV signal, so if a different current is to be sensed, the sense resistor should be changed to suit. The 5534 has a gain of 100, and amplifies the mean DC level to 4 V, with AC peaks rising to 8 V. Any peak above 8 V is clipped by the diode/transistor clamp, since the other half-cycle will already have been clipped by the valve. The clipped signal is integrated by the 2.2 MΩ resistor in combination with the 470 nF capacitor, giving $\tau = 1$ s. The 071 compares this smoothed DC with a reference derived from the potential divider chain, and uses this to control the bias transistor. The clamp reference voltage set by the 2 kΩ variable resistor should be adjusted to achieve constant anode current under all conditions of overload. Although this circuit was designed to provide –11 V bias, this can easily be changed by returning the bias transistor's collector load to a more negative supply as necessary; no other changes are required.

The Quad II

The Quad II is an unusual design, which at first sight does not look too promising, but works because the design is synergetic.

In this design, not only has the phase splitter been combined with the driver stage, but it has also been combined with the input stage. In order to achieve the necessary gain, pentodes have been used. Output resistance is therefore high, as is input noise. To make matters worse, a variant of the see-saw phase splitter has been used. The output stage has local feedback, requiring increased drive voltage (see Figure 6.25).

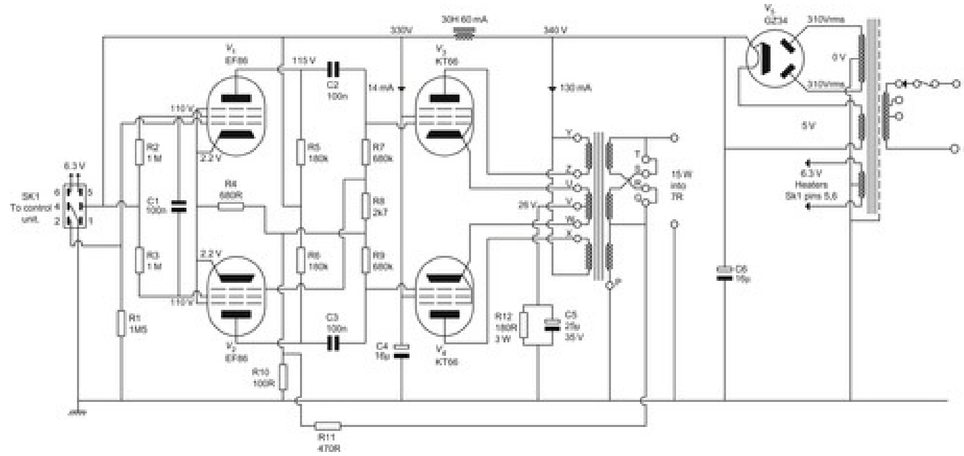


Figure 6.25 Quad II
(by permission from Quad Electroacoustics Ltd).

The output stage is a pair of KT66 beam tetrodes with anode and cathode loads split in the ratio 9.375:1. The cathode connection, therefore, provides little drive to the loudspeaker and may be considered to be series feedback from the output transformer. However, the cathode current in the output transformer is the sum of the anode and g_2 currents, and it was found that this summation reduced third harmonic distortion by a further 8 dB over that due to the negative feedback [13] .

The effect of this transformer-coupled feedback on output resistance is the opposite to what might be intuitively expected [14] . If we simply leave a cathode resistor unbypassed, then this generates series feedback which increases r_a , yet transformer-coupled feedback *reduces* r_a . This can be understood if we apply a short circuit as a load. Clearly, the output stage is unable to drive any voltage into this load, but conversely there is no feedback signal applied to the cathodes. The grids are therefore driven by the full input signal, rather than the input signal minus the feedback, so the output stage is driven harder as it attempts to maintain its voltage into a short circuit. This action is directly equivalent to reducing output resistance, and the new value of output resistance can be found using the normal feedback equation.

The transformer primaries are equivalent to 3 kΩ anode to anode. With tetrodes, this low value of anode load results in a reduction of third harmonic distortion, and an increase in second harmonic, which is then cancelled by push–pull action in the output transformer (assuming that the output valves are perfectly matched).

The automatic bias is shared, so there is no provision for balancing anode current, and we can expect an increase in distortion at low frequencies due to saturation of the (rather small) transformer core. Curiously, the cathode resistor was only rated at 3 W, yet it dissipates 3.8 W. If your Quad II distorts, a burnt-out cathode bias resistor may well be the cause.

Even with pentodes, there is not a great deal of gain from the driver circuitry, and input sensitivity is low: 1.4 V for full output. This is an excellent choice of input sensitivity for a power amplifier, as not only does it guarantee impeccable noise performance (even from a pentode), but also it means that the input is far less susceptible to hum and noise from input cables or heater circuitry. The Quad II was only beaten in signal-to-noise performance by the Williamson, which was quieter because it had a triode input stage.

Despite being a variant of the see-saw phase splitter, the phase splitter/input stage does not rely on feedback for balance, and its operation is quite elegant. The output valves *must* each have a grid-leak resistor, so instead of applying additional loading to the driver valves, a tapping is taken from one of these to provide the input for V_2. Provided this tapping has an attenuation equal to the gain of V_2, the output of the phase splitter must be balanced. Component variation means this will not always be true, so the cathodes of the two valves are tied together to improve balance.

Pentode stages have output resistance ≈ R_L. Since R_L for the Quad input/phase splitter/driver is 180 kΩ, this would appear to be very poor at driving the ≈30 pF input capacitance of the output stage, resulting in a cut-off of ≈30 kHz. However, apart from the output transformer, this is the only High Frequency cut-off in the circuit, and it is not a problem because it is the dominant pole. Each output valve requires a swing of ≈80 $V_{pk–pk}$, which is easily provided, because pentodes can approach 0 V more closely than triodes, and also because LC filtering was used on the HT line, rather than RC filtering, thus increasing the available HT. The LC filtered HT supply also feeds g_2 of the output valves, which has the valuable advantage of reducing hum, since the anode current of a tetrode or pentode is far more dependent on g_2 voltage than anode voltage.

Pentodes need to have g_2 decoupled to ground. Instead of each EF86 having a capacitor to ground, a single capacitor is connected between g_2 of the two valves. This has three advantages:

- If we had two individual capacitors, they would effectively be in series, with a centre tap to ground. Since each valve is connected to an equal but opposite signal, the centre tap would be at ground potential even if it were disconnected from ground. Therefore, we could cheerfully disconnect the centre tap from ground, leaving two capacitors in series that can be replaced by a single capacitor of half the value.
- Since this one capacitor is connected between two points of equal potential, it doesn't need the full voltage rating to ground. However, it is as well to consider the effect of fault conditions when determining the voltage rating, so this is not a great advantage.
- Connecting g_2 of each valve together at AC helps maintain balance in the same way as commoning the cathodes.

Although substituting one stage that combines the functions of input, phase splitter and driver does not achieve the linearity of purpose designed stages, it achieves better linearity than the Mullard circuit because less gain is demanded from it.

With only a simple driver circuit and output stage within the global feedback loop, the elegant Quad II has no stability problems, but it has by far the smallest output transformer of all three amplifiers. A small output transformer always means compromised primary inductance, which means that much of the output stage's available current is wasted driving the output transformer's reactance rather than the external load. However, remember that the Quad II was designed to drive an electrostatic loudspeaker (open circuit at low frequencies), so this weakness was not apparent in its intended use.

New Designs

We have investigated individual stages, we have looked at functional blocks, and we have seen how classic designs were configured. Rather than merely observing, it is now time to put that knowledge to use, and design an amplifier.

In early editions of this book, it was suggested that an old amplifier could be cannibalised for its transformers and chassis. Sadly, this approach can no longer be justified because classic amplifiers are now likely to be fifty years old, and the amount of work expended in re-using components that might fail within ten years is prohibitive. It is now cheaper and easier to make an amplifier with completely new components.

Single-Ended Madness

There are three reasons why a single-ended design has been included:

- The author feels that the single-ended genre should be given a fair trial and hanging. Or not, as the case may be.
- The author was given an NOS 6528, and realised that not only could it form the basis of a stereo amplifier, but also he already had a suitable mains transformer, chokes, rectifiers, and HT capacitors. (The fact that a pair of custom-designed output transformers would be needed was not allowed to intrude upon this logic.)
- A single-ended amplifier is electrically simpler than push–pull, so it can make a good first project.

However, be warned. For a given output power, single-ended amplifiers are significantly bigger, heavier, and more expensive than push–pull.

The Scrapbox Challenge Single-Ended Amplifier

Unfortunately for the transformer manufacturers, much of the ironmongery for this amplifier came from the author's salvage stock – hence the amplifier's name. Almost none of the components are critical, although alternatives will be offered as the design argument progresses.

Choice of Output Valve

Power amplifier design starts with required output power, which then leads to the choice of output valve(s). Thankfully, loudspeakers are gently becoming more sensitive as their designers appreciate the advantages of a carefully chosen cone material, so even ≤10 W suffices perfectly well without having to resort to expensive high-efficiency designs such as horns. Having decided on ≤10 W, the next choice is *which* 300B to use. The world probably has enough 300B designs, so the 6528 made for an interesting alternative.

The 6528 (distributed by Tung-Sol/Chatham, Cetron, and Raytheon) is a double triode intended for use as a series-pass valve in regulated power supplies. The glass envelope and base resembles a GEC KT88, and it is internally akin to a 6080, but the detailed construction and consequent specifications are positively heroic (Table 6.3).

Table 6.3 Comparison of 6080 and 6528 Dual Triode Series-Pass Regulator Valves

	6080	**6528**
μ	2	9
g_m	7 mA/V	37 mA/V
r_a	280 Ω	245 Ω
$P_{a(max)}$	13 W	30 W
$I_{k(max)}$	125 mA	300 mA
$V_{a(max)}$	250 V	400 V
I_h	2.5 A	5 A
C_{ag}	8.6 pF	23.8 pF
C_{gk}	5.5 pF	17.8 pF

The primary attraction of the 6528 is its astonishingly low r_a, which suggests that we could use a low-impedance output transformer – enabling better transformer design. We now need to decide what the primary impedance should be, so we will soon plot loadlines on the anode characteristics, but we must first clarify the class of output stage.

Choice of Output Class

Actually, there is very little room for manoeuvre. A single-ended amplifier can only be Class A. Class A2 implies grid current and requires power driving circuitry with extremely low output resistance, so this low-power amplifier will be Class A1, where no grid current is permitted, allowing a simple voltage driver stage to suffice.

Choosing the DC Operating Point by Considering Output Power and Distortion

A power valve consumes expensive heater power, so it does not make sense to operate it at any point other than its maximum anode dissipation (30 W). This means that our loadline must be a tangent to the $P_{a(max)}$ curve. We also know that for maximum output, we should not clip one half of the waveform before the other, so our DC operating point must allow equal and opposite grid swings. Using the Tung-Sol/Chatham curves, sweeping a transparent ruler along the $P_{a(max)}$ curve resulted in a 2 kΩ loadline with an operating point at $V_a = 255$ V, which required $V_{gk} \approx -27$ V and $I_a = 120$ mA (see Figure 6.26).

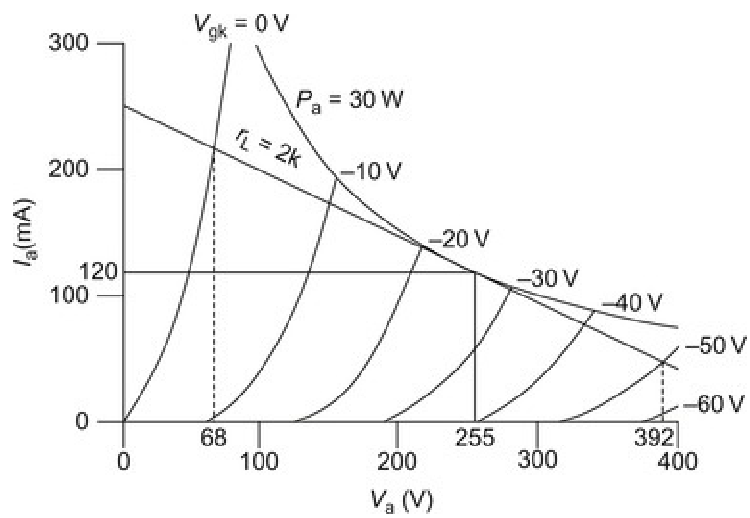


Figure 6.26 Operating conditions of 6528 output valve.

From the operating point, we can swing quite linearly to almost $V_{gk} = 0$ V before drawing grid current, and to a roughly equal and opposite swing of $V_{gk} = -50$ V before cut-off begins to cramp the grid curves. Once we know how many grid volts can be swung, we can check the corresponding anode swing. At $V_{gk} = -0$ V, $V_a = 68$ V, and at $V_{gk} = -50$ V, $V_a = 392$ V, so the peak-to-peak anode swing is 324 V. If we assume that we swing an undistorted sine wave, then 324 V $_{pk-pk}$ =115 V $_{RMS}$.

The loadline passes from 500 V (I_a =0) to 250 mA (V_a =0), which corresponds to 2 kΩ, and this is much higher than the conventional choice of 2 r_a . If we know the anode load and the voltage swing across it, we can calculate the power developed in the load to see if the proposed loadline is acceptable:

$$P = \frac{V^2}{R} = \frac{115^2}{2{,}000} = 6.6\ \mathrm{W}$$

We can estimate the percentage of second harmonic distortion using:

$$\%D_{\mathrm{second\,harmonic}} \approx \frac{V_{\mathrm{quiescent}} - (V_{\mathrm{max}} + V_{\mathrm{min}})/2}{V_{\mathrm{max}} - V_{\mathrm{min}}} \times 100\%$$

At the chosen operating point, V_{max} =392 V, V_{min} =68 V and $V_{quiescent}$ =255 V, resulting in 7.7% second harmonic distortion at full output.

This performance is typical for the genre, and other loadlines predicted significantly less audio power or worse distortion. (There is no point in being too critical about the anode load, since real loudspeakers are nothing like pure, constant resistances anyway.)

Specifying the Output Transformer

We are now able to specify the output transformer:

Type: Single-ended

I_{DC} =120 mA

P_{max} ≈6.6 W

Primary impedance=2 kΩ.

The transformer designer will immediately want to know the secondary load impedance, and because loudspeakers are *not* pure 8 Ω resistances, it is usually better to design for a 4-Ω load. Since a good output transformer has multiple secondary sections, transformer manufacturers commonly offer four sections that can be configured for 1 Ω, 4 Ω (preferred setting for practical loudspeakers), 8 Ω and 16 Ω (ideal if you could find a genuine 16-Ω loudspeaker).

The next important question is the lowest frequency for which maximum power is required. This is an expensive question. As an example of classic commercial practice, the Leak Stereo 20 and TL12+ amplifiers could only produce their specified power down to 50 Hz. The Sowter 9512 output transformer was specified for 8 W at 25 Hz in the unlikely event that practical power might exceed predicted power.

Biassing the Valve

We could apply −25 V directly to the grid of the valve to set the required 120 mA of anode current, but a small drop in grid bias voltage would cause $P_{a(max)}$ to be exceeded instantly. This is the reason why valve manufacturers do not recommend grid bias with high mutual conductance valves – 35 mA/V is very high by valve standards.

We *must* use cathode bias. We need to drop 27 V across a resistor that passes 120 mA, so by Ohm's law:

$$R = \frac{V}{I} = \frac{27}{0.12} = 225\ \Omega$$

If the resistor has a voltage across it, and current passing through it, then it must dissipate power:

$$P = I \times V = 0.12 \times 27 = 3.24\ \text{W}$$

A 5 W power resistor is required as an absolute minimum. We could use an MPC-5 thick film resistor, which is non-inductive, but these get very hot when dissipating >2 W in still air, or we could screw a WH15 aluminium-clad wirewound resistor to the chassis, which would be cool. After much havering, the prototype choice was a 200 Ω MPC-5 positioned in the air flow below the valve and connected in series with a 100 Ω wirewound variable resistor, thus allowing a precise current to be set.

The Cathode Bypass Capacitor

The cathode resistor must be bypassed with a capacitor to avoid unwanted feedback that would raise the valve's r_a, and require a new loadline. As in small-signal calculations, we look to see the AC resistances to ground from the cathode. Looking into the valve, the capacitor sees r_k:

$$r_k = \frac{r_a + R_L}{\mu + 1} = \frac{350 + 2000}{9 + 1} = 235\ \Omega$$

But this is in parallel with the 225-Ω cathode resistor R_k, so:

$$R_k // r_k = 115\ \Omega$$

In a small-signal stage, we would want the bypass to work down to 1 Hz, but this is not necessarily the case in a power stage. By definition, power stages swing large voltages and produce distortion. In a single-ended triode stage operated below clipping, the distortion is primarily second harmonic, but this includes a DC component that the bypass capacitor integrates, shifting the bias away from the intended DC operating point. Once the large signal has gone, the bypass capacitor gently recovers to the designed bias. Recovery time is determined by the time constant of the capacitor in conjunction with $R_k // r_k$. If we set $f_{-3\ \mathrm{dB}} = 1$ Hz, this implies a time constant $\tau = 159$ ms. It takes 5 τ for a CR combination to recover fully from a disturbance, and 0.8 s might be considered to be too long in musical terms, so we might set f −3 dB=10 Hz, which means that the output stage would recover from bias shift in only 80 ms:

$$C = \frac{1}{2 \times \pi \times 10 \times 115} \geq 150\ \mu\mathrm{F}$$

The author didn't have any 150 µF capacitors in stock, or even 220 µF, but he did have some low ESR 1,000 µF 35 V, so he used these and accepted an overload recovery time of 0.5 s. His reasoning was that overload should be uncommon, but that a small capacitor would increase anode resistance and increase bass distortion in the output transformer, and this defect would be there all the time. Far more designs are done this way than you might think – we calculate carefully (possibly on the back of an envelope), then justify our engineering compromise according to component availability at the time. It is quite possible that the optimum value of this capacitor could best be set by ear.

Finding the Required HT Voltage

The output transformer drops some HT across its primary winding resistance, so $R_{\mathrm{DC(primary)}}$ needs to be known. The transformer manufacturer can usually predict this value, but it is useful to have a measurement. For the transformer used, $R_{\mathrm{DC(primary)}} = 152\ \Omega$, so Ohm's law dictates that the voltage drop due to 120 mA of anode current is:

$$V = IR = 0.12 \times 152 = 18.24\ \mathrm{V}$$

Our design requires $V_a = 255$ V, with cathode bias of 27 V, and we drop 18 V in the output transformer, so we need an HT of 300 V at the top of the output transformer. Finding this HT voltage is significant because it determines the maximum HT available to the driver stage (unless we are prepared to add a subsidiary HT supply).

HT Smoothing

Push–pull amplifiers cancel HT hum in their output transformer, but single-ended amplifiers cannot cancel, so their HT supplies must be much quieter. Worse, single-ended amplifiers demand a changing current (from $0 \leq I_{DC} \leq 2I_{DC}$) from their supply, so the low output resistance of a choke input supply is almost obligatory.

HT Rectification

The heroic specifications of the 6528 were not achieved without an Achilles' heel. To avoid cathode stripping, the data sheet specifically warns that the cathode requires 30 s to warm up before HT may be applied. This seems to be a perfect application for a valve rectifier. We only need 120 mA plus a little for the driver, perhaps 10 mA, and the HT voltage is only 300 V, so an EZ81 would be ideal.

In practice, typical valve rectifiers start conducting ≈11 s after power is applied, so a further delay is required, which can be provided by a thermal delay relay. Thermal delay relays look just like valves, and consist of a heater actuating a bimetallic strip in an evacuated glass envelope. The bimetallic strip is composed of two bonded metals having different thermal coefficients of expansion, so the strip bends when heated, and forms the moving half of a switch contact. Because the relay is almost evacuated, switch contact arcing is almost eliminated, and thermal losses are insignificant, so the heat required to make the strip actuate the switch contacts is determined substantially by its specific heat capacity and thermal mass. If necessary, delay can be increased by a factor of up to 3:1 over the relay's rated delay by reducing its heater voltage.

If the delay relay contacts are placed in the path of the rectifier's heater supply, then the delay of the relay is added to the delay of the rectifier, and the HT rises gently over a period of ≈5 s at the appointed time. Alternatively, many delay relays can safely switch AC mains, but there must be negligible voltage between the heater and the moving contact, implying switching the neutral, which always makes the author uneasy because it leaves an apparently dead amplifier with live mains on many terminals. At the time, the author didn't have any data for the Amperite 6NO45T delay relay he found in his scrapbox, but from the part code he deduced that it required a 6.3 V supply, that the contacts were Normally Open (pretty obvious when you can see the contacts through the glass), and that it would give a 45 s delay time. The device was an all-glass envelope on a B9A button base, so it was easy to see which pin was which, and test the deduction. On test at 6.3 V, the heater drew 300 mA, and the switch contacts closed after 41 s.

When power is applied, the mains transformer delivers heater and HT voltage simultaneously to the valve rectifier, but its heater is cold, so the cathode suffers ion bombardment. Although placing the delay relay in the rectifier's heater circuit ensures that the HT to the audio circuitry rises gently from zero, it does mean that the rectifier suffers an extra 45 s of ion bombardment each time the amplifier is switched on. This is an engineering compromise – the EZ81 is a cheap sacrifice to appease the far more expensive 6528.

The HT Transformer

We need 300 V of HT at the top of the output transformer, and have elected to use valve rectification combined with a choke input supply. The choke drops voltage across its R_{DC}, so this must be determined. The author's scrapbox (more of a room, really) yielded a pair of 15 H, 250 mA Parmeko chokes that looked hopeful and whose R_{DC} =136 Ω, so 130 mA passing through one of these chokes would drop 17 V, and this would be added to the 300 V, to give 317 V.

The rectifier manufacturer's choke regulation curves were used to determine the required transformer voltage. Interpolation of Mullard EZ81 curves predicted ≈375 V_{RMS} for our required 317 V_{DC}. Nowadays, we would simply drop the numbers into PSUD2.

Further rummaging in the scrapbox uncovered a large C-core transformer with a pair of 375–0–375 V at 250 mA windings and numerous 6.3 V heater windings, so this seemed ideal, allowing dual mono construction on one chassis.

HT Choke Suitability

We finally have sufficient information to test whether or not the posited 15 H 250 mA HT choke is satisfactory. Using equations from Chapter 5 and assuming 50 Hz mains:

$$I_{\text{AC (positive peak)}} = \frac{V_{\text{in (RMS)}}}{1155\,L} = \frac{375}{1155 \times 15} = 22\ \text{mA}$$

$$I_{\text{total peak current}} = I_{\text{DC}} + I_{\text{AC (positive peak)}} = 130\ \text{mA} + 22\ \text{mA} = 152\ \text{mA}$$

The choke is rated at 250 mA, so it should easily support this current.

The minimum load current required is:

$$I_{\text{min (mA)}} = \frac{V_{\text{in (RMS)}}}{L_{\text{(H)}}} = \frac{375}{15} = 25\ \text{mA}$$

The output stage draws 120 mA, so we are safely above this lower limit.

We can estimate the hum due to HT once we know the proposed value of smoothing capacitor. The author had some 120 μF 400 V polypropylene capacitors in stock, so:

$$V_{\text{hum (RMS)}} \frac{3V_{\text{in (RMS)}}}{L_{(\text{H})}C_{(\mu\text{F})}} = \frac{3 \times 375}{15 \times 120} = 0.625\ \text{V} = 625\ \text{mV}$$

The anode load and r_a form a potential divider, so the ripple voltage seen at the anode is:

$$V_{\text{hum (anode)}} = \frac{r_a V_{\text{ripple}}}{r_a + R_L} = \frac{400 \times 625\ \text{mV}}{400 + 2{,}000} = 104\ \text{mV}$$

An output transformer responds to the voltage *across* it, so it sees 625 mV–104 mV=521 mV of hum. At full output, the output stage swings 115 V_{RMS}, so 521 mV corresponds to a 47 dB signal/hum ratio, which is inadequate, so a further stage of filtering is needed.

A second stage of LC filtering having a loss at 100 Hz of only 32 dB would improve the signal/hum ratio to almost 80 dB. 32 dB corresponds to a voltage ratio of 40, so the AC potential divider formed by the second LC filter would need $X_L / X_C \approx 40$. If another 120 μF capacitor were available, then even a 1 H 130 mA choke would be adequate.

However, the author *didn't* have another suitable pair of chokes, and had already realised that the amplifier was going to be large and heavy (even by valve standards). Adding yet more mass was not at all attractive.

The HT Regulator Option

LC filters might be good at reducing hum, but their output impedance is still quite high (many tens of ohms). This is particularly significant for a single-ended amplifier because the output valve is unable to distinguish between the reflected load of the loudspeaker through the output transformer and the output resistance of the supply (see Figure 6.27).

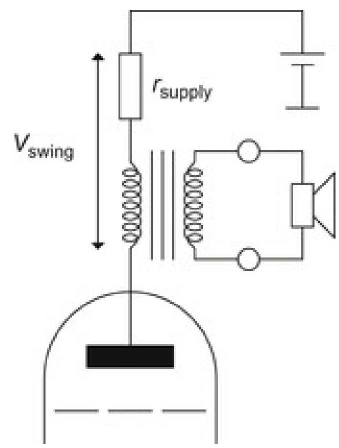


Figure 6.27 The effect of non-zero supply resistance on a power amplifier.

The swing of the output valve is developed across both these components, yet we can only couple the swing developed across the output transformer. We lose power, and our output resistance rises. An HT regulator allows optimum bass performance from a single-ended amplifier.

Each channel of the amplifier requires 300 V at 130 mA, and you could simply use the entire HT supply of Figure 5.45 without any modifications. But we do not need a great deal of ripple rejection, so an adaptation of the simple two-transistor regulator is a possible alternative (see Figure 6.28).

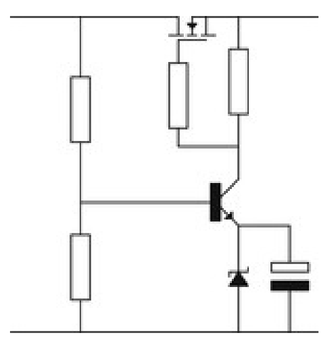


Figure 6.28 Two transistor HT regulator.

The two-transistor regulator has the advantage that it does not need to drop many volts, which reduces heat dissipation. We will assume that the regulator must drop ≥10 V, and that this will occur when the mains has dropped by 6% (as it is allowed to do). Thus, the nominal HT voltage required at the input to the regulator is:

$$V_{HT\ (nominal)} = \frac{300\ V + 10\ V}{1 - 0.06} = 330\ V$$

Checking the EZ81 data sheet, this would require a mains transformer with 412–

0–412 V HT windings.

High-voltage bipolar power transistors have rather low h_{FE}, they are slow, and they are expensive, so a high-voltage MOSFET power transistor can often be a better choice for the series-pass element.

When noise is not critical, it makes sense to make the Zener reference voltage as high as possible because this reduces dissipation in the error transistor, and also allows increased loop gain, which gives more feedback to reduce output resistance. 220 V is therefore a good Zener choice for a regulator that must provide an output of 285 V. Although 220 V Zeners are available (and the author had some in stock), three cascaded 72 V Zeners are better. The reason for this is that the very high-voltage Zeners are rather noisy because they must be operated at a low current to reduce device dissipation ($P = IV$). Using three cascaded Zeners allowed a Zener current of 4 mA, which reduces noise. To reduce noise further, the Zeners are bypassed by a 22 μF 350 V capacitor.

The gate of the MOSFET will be at $V_{out} + V_{gs}$ =300 V+4 V=304 V (despite huge device variation, 4 V is a reasonable rough assumption for V_{gs} of a power MOSFET). Since the collector of the error transistor is connected to the gate of the MOSFET, and the emitter is tied to the 216 V (3×72 V) Zener reference, V_{CE} =304 V – 216 V=88 V. We want the error transistor to pass 4 mA into the Zener reference, so I_c =4 mA, and the power dissipated in the transistor is 352 mW. This is a significant calculation because it confirms that our choices of V_{CE} and I_c enable us to use a small-signal transistor.

When working, the error transistor only has V_{CE} =88 V, but at the instant of start-up, the 22 μF Zener bypass capacitor clamps the emitter of the error amplifier to 0 V, so it must be able to survive V_{CE} =330 V. The requirements for the error transistor are now clear, and the 400 V, 625 mW MPSA44 is ideal.

High-voltage transistors have low h_{FE}, and the MPSA44 is no exception. When tested under the expected operating conditions, h_{FE} ≈100. I_c =4 mA, so $I_b = I_c / h_{FE}$ =40 μA. Even if we pass 1 mA through the sampling potential divider chain, we cannot treat it as a pure potential divider because the 40 μA base current disturbs the result.

We set the sampling divider chain current by considering the power dissipation of the lower resistor first. If we use a 0.6 W component, and allow it to dissipate 0.2 W, it should remain sensibly cool. The resistor is connected to the base of the MPSA42, which is 0.7 V higher in voltage than the emitter, so the resistor has 217 V across it. Using $P = V^2 / R$, its resistance must be 217^2/0.2=235 kΩ, so we will use the nearest preferred value of 240 kΩ, which will dissipate

196 mW. The current through the resistor is 217 V/240 kΩ=904 µA. Because the error transistor steals 40 µA for its base, the upper potential divider resistor passes 904 µA+40 µA=944 µA of current. The voltage across this resistor is 300 V – 217 V=83 V, so its resistance must be 83 V/944 µA=87.9 kΩ, and the standard value of 91 k will be fine.

There is no point in adding a speed-up capacitor across the upper resistor because the low attenuation of the divider chain (2.8 dB) means that it could only marginally improve ripple rejection, yet the required value would slow the response of the regulator to Low Frequency transient current demands (see Chapter 5).

The least critical circuit value to be calculated is the error transistor collector load resistance. We know that the input to the regulator is 330 V, and the collector voltage is 304 V, so this resistor has 26 V across it, passes I_c =4 mA, and its value must be 26 V/4 mA=6.5 kΩ, so the standard value of 6k2 will be just fine.

Back-of-an-envelope calculations suggested that the output resistance of this regulator would be <5 mΩ, and that it would reject hum by >50 dB. In this instance, these figures are more than adequate, and offer better performance than another choke.

Estimating Amplifier Output Resistance

Checking AC conditions, r_a can be approximated by drawing a tangent to the −25 V grid line adjacent to the operating point, and this gives a value of ≈400 Ω. This is significant because we can use it to calculate the output resistance of the amplifier. The output transformer matches the assumed 8 Ω load of the loudspeaker to the 2 kΩ required by the valve, giving an impedance ratio of 250:1. (The author matched to 8 Ω rather than 4 Ω because he knew the amplifier would be used with 12 Ω Rogers LS3/5a.) Conversely, source resistance is divided by this ratio, and we should include output transformer primary resistance (Sowter 9512, R_p ≈150 Ω), giving 2.2 Ω, which is in series with output transformer secondary resistance plus wiring resistance (0.95 Ω), leading to a final estimated output resistance of ≈3 Ω. An output resistance of half load resistance is typical of single-ended amplifiers and is too high for most loudspeakers to operate as their designer intended. Later, we will see that not only did adding cathode feedback halve distortion, it also lowered measured output resistance to a quarter of load resistance, making it slightly more acceptable.

What are the Driver Stage Requirements?

In the output stage, the maximum undistorted grid swing from the operating point is limited by the onset of grid current and a symmetrical swing in the opposite direction. Grid current occurs at 0 V, and by symmetry the maximum opposite swing must be 2 V_{gk}, so the required peak-to-peak grid swing for *any* Class A amplifier is *always* twice the grid–cathode bias. In our case, this means that we need 54 V_{pk-pk} or 19 V_{RMS}.

We know that the 6528 must swing ≈115 V_{RMS} at its anode and ≈19 V_{RMS} at its grid, so it amplifies by a factor of $A_v \approx 6$, allowing us to find the Miller capacitance, which is $C_{ag} \cdot (A_v + 1) = 23.8$ pF(6+1)≈167 pF. This is in parallel with C_{gk} (17.8 pF), so the total input capacitance including strays is ≈200 pF.

We will investigate detailed arguments for the required high-frequency response of an amplifier in Chapter 7 , but if we make the sweeping assumption that $f_{-3\,dB} > 150$ kHz, this requires a source resistance of:

$$r_s = \frac{1}{2 \times \pi \times 150{,}000 \times 200 \times 10^{-12}} \geq 5{,}300\ \Omega$$

The 6528 has a very high g_m, and will oscillate at RF given half a chance. It therefore needs a grid-stopper resistor to prevent oscillation, and the manufacturer's recommended minimum value of 1 kΩ bites into our required source resistance, reducing it to 4.3 kΩ.

This is a low output resistance for a valve driver stage, and severely limits our design choice. In a practical common cathode driver stage, output resistance is roughly equal to the valve manufacturer's claimed value of r_a, so we are looking for a valve with a very low r_a. Frame-grid valves can achieve this value of r_a, but they tend to produce more third harmonic distortion than valves with helical grids, so a driver stage using conventional valves with the output taken from a cathode would be preferable.

Driver Stage Topology

There are various options for the driver stage:

- A carefully designed common cathode stage (probably with active load) could be DC coupled to a cathode follower. This could give stunningly low distortion and $r_s < 4.3$ kΩ.
- A μ-follower could give low distortion and $r_s < 4.3$ kΩ.

- An SRPP could give r_s <4.3 kΩ, and a higher voltage swing than a μ -follower but with higher distortion.

At the time of construction, the author didn't have any DN2540N5s for simple active loads, so the μ -follower was the obvious choice, but testing showed that 290 V wasn't quite enough HT to enable the required 18 V_{RMS} swing. Further, we know that the output stage is a capacitive load, for which the SRPP is very well suited, and it can swing more signal than the μ -follower for a given HT voltage, so higher second harmonic distortion due to the low R_L / r_a ratio at the lower valve is its only drawback.

Choice of Valve for the Driver Stage

Ideally, we would like a driver valve that produces primarily second harmonic distortion with insignificant higher harmonics, because that might conceivably allow some distortion cancellation with the output stage. Frame-grid valves are now almost eliminated (although the E88CC is easily one of the best), so the obvious choice is the *SN7/*N7 family. The author makes no apologies for arriving at the same choice as dozens of other single-ended designs. If sound engineering arguments dictate that round wheels are best, then that is what we will use.

However, an SRPP has its upper cathode at 0.5 V_{HT} , so it requires an elevated heater supply if heater/cathode insulation is not to be strained. In a stereo amplifier, one *SN7/*N7 could be shared by the upper valves, and another for the lower. Alternatively, the 6J5GT is half of a 6SN7, so a pair of these could be used in a monoblock amplifier. The author chose to use 6J5GTs because he had previously bought lots of them. Additionally, if individual valves are used, no unsightly metalwork is needed to modify the design later on; perhaps a high- μ input valve (6SQ7) as a common cathode stage with LED bias and constant current sink load DC coupled to a 6J5GT cathode follower if global negative feedback was required.

On test, the 6J5GT/6J5GT SRPP stage easily swung 21 V_{RMS} at 1 kHz, with second harmonic at −40 dB and third at −54 dB. However, loading considerations meant that this measurement was taken purely by the oscilloscope/spectrum analyser, so the reliable measurement dynamic range was limited to only ≈55 dB, and higher harmonics could not be seen. Nevertheless, this distortion was felt to be acceptable compared to the predicted distortion of the output stage.

Determining the Driver Stage Operating Point

Determining the Driver Stage Operating Point

The title 'SRPP' (Shunt Regulated Push–Pull) implies that SRPP stages should use identical upper and lower valves. In practice, this seems not to be critical, and the author has not measured any advantage when trying different valves, but it is certainly easier to design with identical valves. The valves are in series (so they pass the same current), and identical, so the anode of the lower valve must be at half the HT voltage. Therefore, we design by considering the lower valve to be a common cathode stage with $V_a = 0.5\,V_{HT}$.

The 6J5GT ideally likes $I_a \geq 8$ mA for constant r_a, or to consider it another way, $I_a \geq 8$ mA is likely to give lowest distortion. We will set $I_a = 8$ mA, requiring $V_{gk} \approx 3.4$ V.

As a corollary, driving 200 pF of shunt capacitance at 20 kHz with 19 V_{RMS} requires ≈0.48 mA_{RMS}, or ≈1.3 $mA_{pk\text{–}pk}$ of signal current. An SRPP passing 8 mA should be comfortably able to provide this signal current without adding slewing distortion.

Setting Driver Stage Bias

The upper valve of an SRPP *must* be resistor biassed, *without* a bypass capacitor, otherwise there would be no signal to drive the valve, but the lower valve has a little more freedom.

The conventional cathode bias choice for the lower valve would be a 430 Ω resistor bypassed by an appropriately sized capacitor. However, when we designed the output stage, we considered the effects on bias after recovery from distortion. Since each half of the SRPP operates with only half of the available HT voltage (limiting signal swing), recovery after distortion or overload is important, so it would be better to use fixed bias in the lower valve. Fixed bias could be provided by grid bias or by LED cathode bias. Grid bias is expensive, but LED cathode bias can increase distortion if the anode load is not large. Fortunately, the author's measurements found that even in this stage, at these signal voltages, the additional distortion produced by LED bias was insignificant, and it allows instantaneous recovery from overload.

Is the Output Resistance and Gain of the Proposed Driver Stage Adequate?

The output resistance of an SRPP driver stage would intuitively be expected to be sufficiently low, but this can be checked using:

$$r_{\text{out}} = \frac{r_{a_2}(R + r_{a_1})}{r_{a_2}(\mu_1 + 1) + r_{a_1} + R[\mu_2(\mu_1 + 1) + 1]}$$

For this design, using measured values of μ =21 and r_a =7.1 kΩ, the equation predicts r_{out} ≈2.3 kΩ, which comfortably allows for a 1 kΩ cathode stopper resistor to reduce the likelihood of RF oscillation in the SRPP.

The gain of the SRPP stage is ≈14, and the output stage requires ≈19 V_{RMS}, so the input stage requires ≈1.4 V_{RMS} to drive the amplifier to full output – which is convenient because it allows ≈3 dB of gain from a standard 2 V_{RMS} CD player to allow for poorly conformed recordings.

But What About Global Feedback?

It is de rigueur for single-ended valve amplifiers not to use global negative feedback. The argument generally presented for rejecting global feedback is not that feedback has to be applied carefully to maintain stability, but that single-ended amplifiers produce distortion that is primarily second harmonic, innocuous and proportional to level, and that adding feedback would translate distortion harmonics up in frequency to where they are more noticeable. This argument is based on the following facts:

- Classical tests of distortion showed that second harmonic distortion was inaudible on sine waves when it was <5%.
- Distortion in a single-ended amplifier is proportional to level.
- Baxandall showed that feedback implied that an amplifier generating second harmonic distortion would distort the distortion, leading to the production of higher harmonics that were not present before the feedback was applied.
- Shorter showed that higher order distortion harmonics are more objectionable and that a reasonable weighting was $n^2/4$.

We will return to this vexed question later, but for the moment merely note that adding distortion to music raises the noise floor.

However, nobody appears to object to adding a small amount of local feedback at the output stage by including the secondary of the output transformer in the cathode circuit.

Summing Up

Now that detailed design is complete, it is worth reviewing the consequences of the design choices to see whether the whole design looks worthwhile (see Figure 6.29).

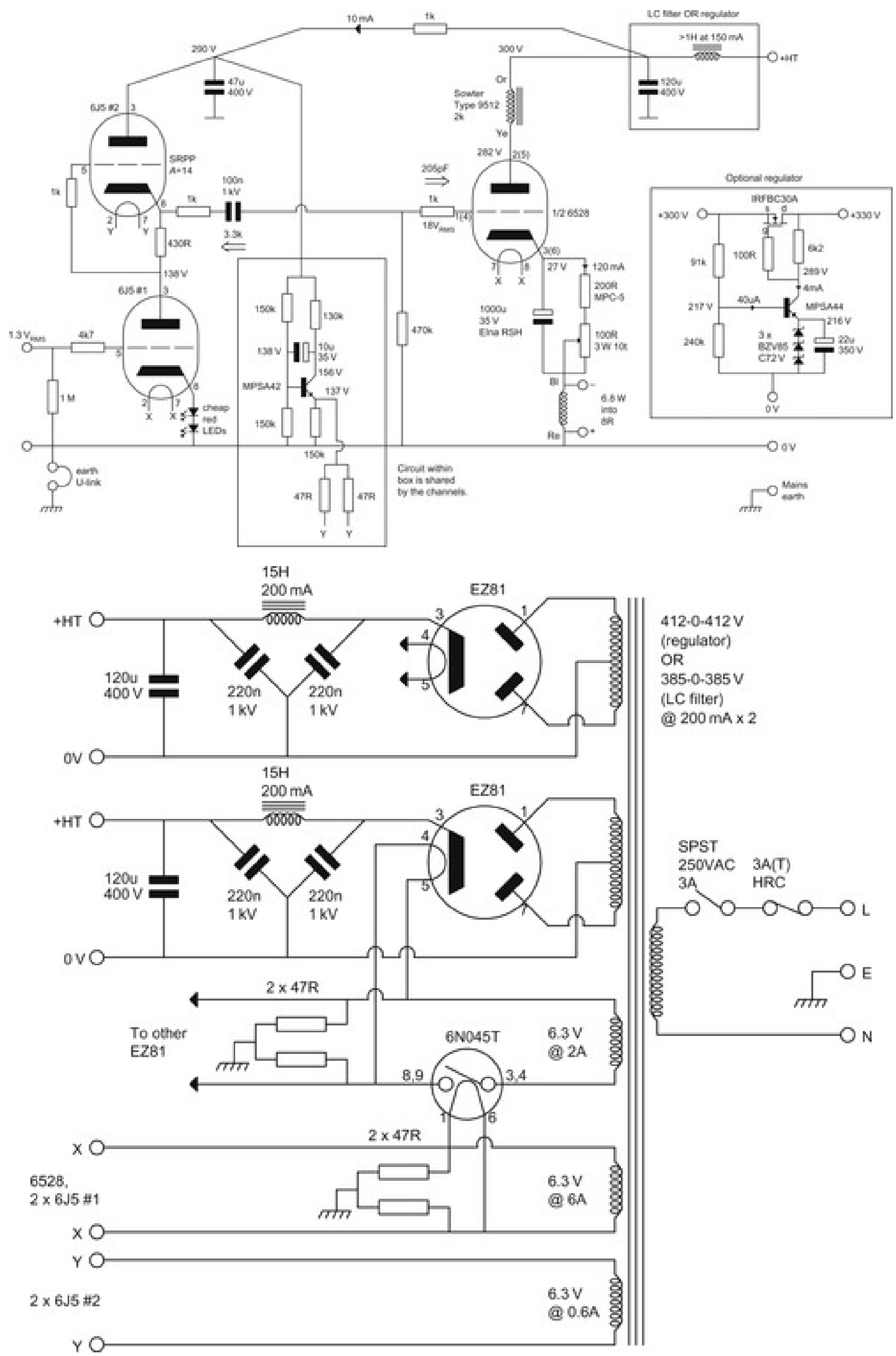


Figure 6.29 Practical 'Scrapbox Challenge' power amplifier.

- Predicted output power ≈6 W with ≈8% distortion.
- Input sensitivity for full power ≈1.4 V_{RMS}.

The amplifier was built to test the predictions and weighs 64 lb. Put another way, it weighs ≥10 lb per stereo watt. Compared to a push–pull design, it is heavy and expensive to achieve a limited objective, but precisely the same allegation would be levelled by a semiconductor designer at *any* valve amplifier. Valve amplifiers are the steam engines of the electronic world – and they arouse similar passion.

Teething Problems

Racing car engines achieve their outstanding performance by operating every part *just* under its limit – so small errors are catastrophic. The 6528 can be considered to be a racing car engine in that it is rated at 30 W per anode, and consumes 31.5 W of heater power, so it has to lose 91.5 W of heat from an envelope the size of a KT88 (P_{total} =52 W). Initial testing therefore concentrated on ensuring that the 6528 would not expire before the chequered flag.

The 6528 valve base was mounted on a wire finger-guard intended for an 80 mm fan, and an 80 mm low noise (claimed 12 dBA) fan blows gently from below the valve base. As a consequence, even though the valve is operating at maximum P_a, its measured envelope temperature is just within limits and the chassis is stone cold.

At the first test, the AC heater voltage at the valve base of the 6528 measured by a reliable true RMS meter was 6.5 V instead of 6.3 V, so the mains transformer primary tapping was changed from 240 V to 250 V, which reduced the heater voltage to 6.296 V – which *is* close enough. (Because the AC mains waveform generally contains ≈5% distortion with significant harmonics up to 1 kHz, AC heater voltage measurements should always be made by a true RMS meter having good accuracy up to 1 kHz.)

Additionally, one of the 6J5GTs had to be rejected because of poor heater/cathode insulation (despite measuring >25 $M\Omega_{(hot)}$ on an AVO VCM163 valve tester). Even though the heaters are decoupled to ground, the fault caused 5 mV of hum and rectifier switching spikes to be fed to the output stage.

Listening Tests

Because of the importance of the trial, the Scrapbox Challenge amplifier was carefully auditioned over a considerable time through the author's pair of Rogers LS3/5as. Although the LS3/5a is a very nice little loudspeaker, it is not ideally suited to weedy amplifiers. Unfortunately, they are the only passive crossover loudspeakers with any pretensions to quality that the author owns.

The amplifier started barely tolerable, but improved greatly over the first four hours of listening, and the fact that a bottle of Veuve Clicquot was symbolically opened seconds after the music began is quite irrelevant.

Designer's Observations

The amplifier's wiring was completed over a weekend followed by two national holidays. Murphy's law thus dictated that missing parts would only be discovered late on the Friday evening. The author intended to use his variant of choke snubbers for the HT supply, but the cupboard was inexplicably bare of 220 nF polypropylene film/foil, so he was initially forced to use a traditional 10 nF film/foil+10 kΩ snubber (although this defect was corrected a week later). With hindsight, this was fortuitous, because the poorer snubbing revealed that the cheap EZ81 rectifier switches on and off surprisingly cleanly and provoked very little ringing (see Figure 6.30).

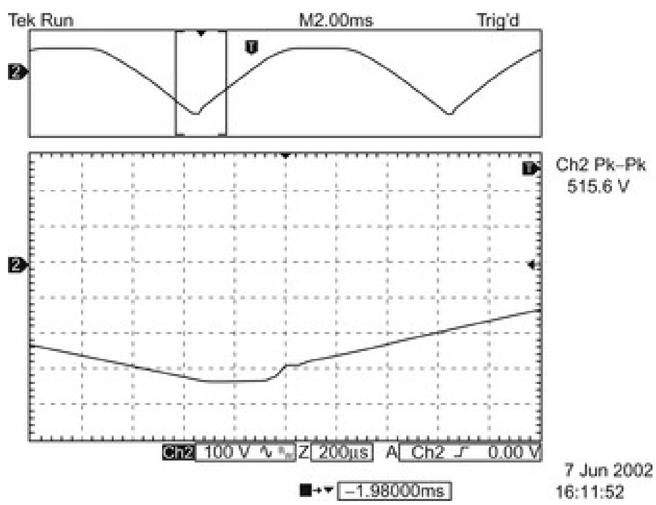


Figure 6.30 Voltage waveform at the output of the EZ81 rectifier with traditional snubber. (The lower, expanded trace shows rectifier behaviour as it switches on and off.)

Building the Scrapbox Challenge amplifier initially reinforced the author's deepening suspicion that the magnetic cores of chokes and transformers can deteriorate with age. The ripple predicted in 2002 after the first stage of HT filtering was 56 mV $_{RMS}$, but the measured value was 7 dB higher at 124 mV $_{RMS}$, and this was thought to be due to the chokes (one of which had previous form – buzzing in an early EL84 amplifier). Fortunately, working through VA3 uncovered an equation error (corrected in this edition), correcting the ripple prediction to 104 mV $_{RMS}$, implying a mere 1.5 dB discrepancy. Sadly, the oversize C-core mains transformer throbbed even before HT current was drawn, and required a 5 A fuse in the mains plug just to survive the inrush current as the amplifier was switched on.

Moral : Forty-year-old electro magnetic components could be the weakest link. A measured failing *might* not be due to your design.

Listening closely at switch-on (before the delay relay activated the HT rectifiers) revealed that some hum was being induced directly into the output transformers from the mains transformer. If you decide to build this amplifier, it would be best to build it on two chassis, one for the audio circuitry and one for the power supply – allowing the noisy power supply to be distanced from the sensitive amplifier.

On test, the amplifier delivered 6 W at 1 kHz with 3.2% THD+N. The amplifier was tested with and without output stage cathode feedback at 1 dB below full power (Table 6.4).

Table 6.4 Scrapbox Challenge Distortion

	2nd	3rd	4th	5th	6th
Without cathode feedback (dB)	−25.2	−57.8	−55.6	−60.6	−58
With cathode feedback (dB)	−31.7	−64.3	−57.1	−74.3	–
Improvement (dB)	6.5	6.5	1.5	13.7	–

The feedback reduced amplifier gain by 3 dB, yet the table shows a significantly greater improvement on all harmonics except the fourth – making it a very worthwhile trade.

Conclusions

With the right programme (yes, acoustic jazz), this amplifier is extremely easy on the ear. It *isn't* accurate – the inevitable high output resistance (2.1 Ω on 8 Ω setting) causes under-damped bass, it falls apart on choral music and Led Zeppelin, and it is very heavy (10 lb/W). Nevertheless, when the author was foolish enough to lend it to a friend having high-efficiency open baffle loudspeakers, it was 3 years before he was able to reclaim it. On the available evidence, the case has not quite been proven, so the hanging has been postponed, even though the amplifier's distortion figures are risible compared to a competently designed transistor amplifier.

What this amplifier needs for an acquittal is some global negative feedback. If it was to be the author's primary amplifier for listening to music (rather than a test mule), the SRPP would be replaced by a 6SQ7 (μ =100) biassed by an insipid green LED (1.8 V at 300 μA) and loaded by a DN2540N5 CCS active load direct coupled to a 6J5GT cathode follower so that the extra 18 dB of gain could be spent on shunt applied global negative feedback to reduce distortion by a factor of seven.

Taking the distortion at 6 W as 3.2%, dominated by second harmonic, this figure was entered into the Baxandall equations to estimate the amplitude of higher distortion harmonics generated by 14 dB of global negative feedback (Table 6.5).

Table 6.5 Predicted Harmonic Generation Resulting from 14 dB Global Negative Feedback

	2nd	3rd	4th	5th
Before feedback (%)	3.2	–	–	–
After 14 dB feedback (dB)	−45	−70	−95	−118

Note that second harmonic falls 14 dB to –45 dB (0.6%) exactly as expected, but that additional higher harmonics have been generated. The amplifier already produces fourth at −57.1 dB, and fifth at −74.3 dB, so the addition of new harmonics at ≈40 dB lower amplitude (≈1%) is entirely negligible. However, the feedback-generated third harmonic at −70 dB is comparable with the amplifier's existing third harmonic at −64.3 dB, so 14 dB feedback can be expected to degrade third harmonic. Without explicitly knowing the phase relationship between the two third harmonic signals, we can only apply a power summation, and doing so suggests that the third harmonic amplitude will increase by 1 dB to −63.3 dB, or in the worst case scenario when the two signals are perfectly in phase by 3.6 dB to 60.7 dB. Shorter's $n^2/4$ weighting suggests that the subjective effect of third harmonic distortion is 3.25 dB worse than its raw amplitude would suggest, so we might be really ungenerous and consider the new third harmonic distortion to be at −57 dB. However, this still means that distortion is dominated by second harmonic at −45 dB.

To summarise, there is no logical argument whatsoever for rejecting the application of a 14 dB global negative feedback, and if more open-loop gain could be found without compromising High Frequency phase response, yet more distortion-reducing global negative feedback would be even better.

Obtaining more than Single Digit Output Power

Bearing in mind that whilst increasing volume by 3 dB is noticeable, it is not significant, merely doubling output power is not enough. The traditional method of significantly increasing power was to use a push–pull pair of Mullard EL34 or GEC KT88. Once push–pull has been chosen, pure Class A is no longer enforced, and Class AB can be used; using these techniques, we can obtain 50 W from a pair of Mullard EL34 or GEC KT66, and 100 W from a pair of GEC KT88 [15] . After this, we resort to transmitter valves at a much higher cost per watt.

Transmitter valves have a number of disadvantages:

- They are invariably disproportionately expensive.
- They tend to need higher impedance anode loads – making the design of a good output transformer more difficult.
- They have savage drive requirements – often needing a power valve as driver.
- They use high HT voltages – so the smoothing capacitors are expensive, and the HT supply is a major safety hazard.
- Smoothing at high HT voltages tends to enforce high L/C ratios, making it increasingly difficult to prevent subsonic ringing in the power supply.

However, there are ways of avoiding these problems.

Sex, Lies and Output Power

In the late 1960s and early 1970s, some quite unpleasant audio amplifiers were made using *transistors* . Compared to the valve behemoths, these transistor amplifiers were very small and light, but they didn't actually sound any better (in fact, most sounded a good deal worse), so something was needed to make them sell. The one thing that even early transistor amplifiers could do was to provide plenty of power into a resistive load, and thus the power rating war started.

To make a truly powerful amplifier, a large power supply is needed, but this is expensive. Now (classical) music generally only has short duration peaks, and nobody listened to anything else (or at least, nobody whose opinions were taken seriously at the time), so amplifiers were designed that could manage higher output powers, but only for a very short time. This allowed power ratings to be increased further, and the ‘music power’ rating was born. We measure the maximum output power at 10% distortion, or the onset of *clipping* (the point at which a sine wave begins to have its peaks clipped off), with bursts of 1 kHz driving one channel *only* into a resistive load. By this means, it is perfectly possible to convert a 20 W amplifier with a poor power supply into a 50 W model, and if we now double the output to account for two channels, we have a 100 W amplifier.

At least four fallacies were used in the previous argument, but they were as nothing compared to the outrageous power claims made for many computer loudspeaker systems. One example having a woofer box the size of a large loaf of bread and a pair of small loaf satellites claimed a power rating of 800 W PMPO and all for £23! (PMPO=Peak Music Power Output, or in this case, Purely Mythical Power Output.)

Loudspeaker Efficiency and Power Compression

We *can* make more efficient loudspeakers. This is the best solution, since inefficient loudspeakers invariably suffer from power compression, an effect whereby the resistance of the voice coil rises due to temperature, and reduces sensitivity until the coil has cooled down.

Unfortunately, making an efficient loudspeaker that is uncoloured is difficult, and the lazy way to reduce colouration is to add a coat of visco-elastic damping material to the moving diaphragm. Unfortunately, the damping material invariably adds significant mass, reducing efficiency. The worst examples of this approach occurred during the 1970s when the plastic Bextrene was preferred as a cone material over paper because of its ease of moulding and comparative consistency, but damping the Bextrene quack to acceptable levels required a heavy coating of aqueous polyvinyl acetate, resulting in staggeringly inefficient loudspeakers.

However, the most criminally effective way of discarding loudspeaker efficiency is to make loudspeakers small. Let’s be clear about this: ‘small’ in a loudspeaker context means ‘smaller than a domestic washing machine.’ Anything smaller suffers not just because acceptable bass can only be obtained by deliberately adding mass to the cone (drastically reducing efficiency) but also because the box’s small baffle area typically requires an extra 2 dB of baffle step compensation that subtracts directly from efficiency. 2 dB might not sound much but it’s the difference between 10 W and 16 W, and that costs.

Active Crossovers and Zobel Networks

We can drive the loudspeakers more effectively. If each drive unit is driven by a dedicated amplifier preceded by an active crossover, many benefits result [16] .

Producing loud bass means moving a large volume of air, so we either have a large area cone (which requires a large box), or we use a small cone with large excursion, which wastes power because with the notable exception of ATC (who use underhung voice coils with a long gap), this forces the voice coil to be many times the length of the gap. Class D switching amplifiers now offer 200 W from an amplifier module smaller than a pair of boxed EL34s, and although they are improving all the time, their fundamental mode of operation means that their weaknesses must show up at high frequencies, so whilst they are ideal for those low efficiency bass boxes only a very few are acceptable full range.

Conversely, valve power at low frequencies requires an output transformer with a large expensive core, so valves excel at midrange and treble. The enforced low mass of midrange and treble drivers and the recent availability of powerful neodymium magnets makes high efficiency easily attainable, ideally matching a small valve amplifier.

Taking the two preceding arguments together, it now makes a great deal of sense to design an active loudspeaker system using a 200 W switching amplifier driving a professional ≥15″ driver chosen for its efficiency and tame cone break-up, and a 20 W valve amplifier above 250 Hz driving a high efficiency 4–6″ driver augmented by a ribbon tweeter above 5 kHz via a passive crossover.

Almost all moving-coil loudspeakers have significant self-inductance, so their dedicated amplifier sees a rising impedance which can compromise High Frequency stability. Additionally, beam tetrodes and pentodes produce higher amplitude higher harmonics in their distortion spectrum as load resistance rises – so correcting voice coil inductance to the optimum impedance load would be worthwhile. Fortunately, a simple moving-coil loudspeaker is easily corrected by adding a Zobel network directly across its terminals (see Figure 6.31).

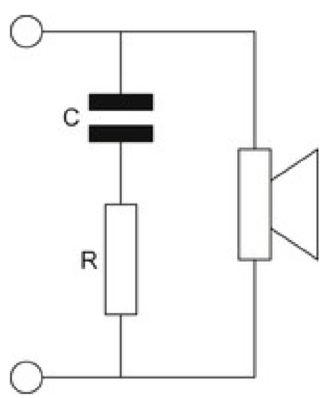


Figure 6.31 Zobel network for cancelling voice coil inductance.

In theory, the Zobel resistor is equal to the DC resistance of the loudspeaker, and the capacitor value is found using:

$$C_{\text{Zobel}} = \frac{L_{\text{voice coil}}}{R_{\text{DC}}^2}$$

In practice, because the loudspeaker can be considered to be a transformer with its voice coil loosely coupled to the shorted turn of the pole pieces which have hysteresis losses, the simple model of pure inductance in series with resistance is somewhat inaccurate, and the required Zobel resistance is typically 1.2 R_{DC}. Given that the voice coil inductance may not be known, the best way to determine Zobel values is to make one with resistance initially set to 1.2 R_{DC}, adjust the capacitor value to give minimum change in meter reading as frequency is swept across the audio band, then fine-tune resistance (see Figure 6.32).

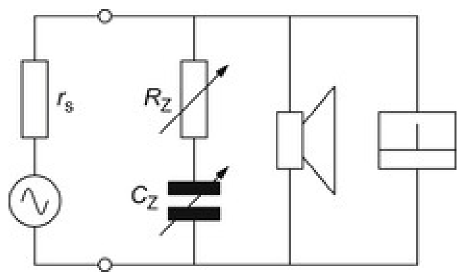


Figure 6.32 Test circuit for determining Zobel values.

Fortunately, values are not critical and the author simply selects E24 resistor values from his stock rather than finely adjusting a precision variable resistor. Likewise, the required capacitor value can easily be made up if a few values between 0.47 μF and 4.7 μF are available. Having done so, the improvement is quite remarkable (see Figure 6.33).

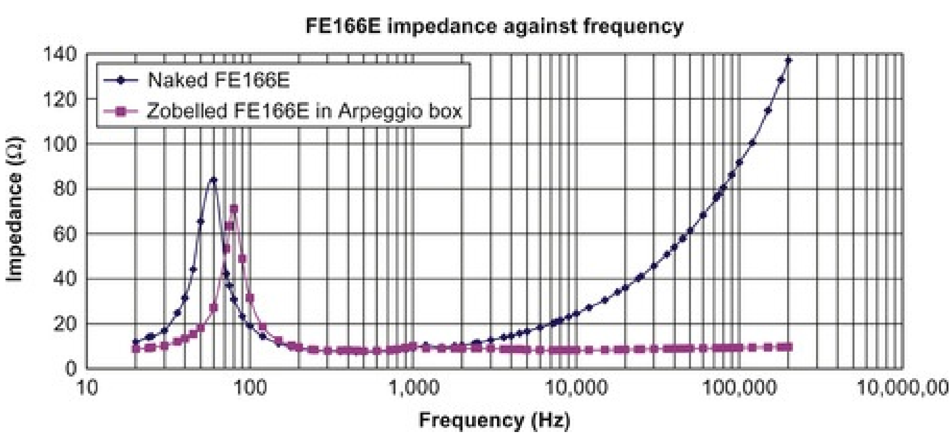


Figure 6.33 Impedance against frequency with and without Zobel network.

Parallel Output Valves and Transformer Design

This is a cracking solution, and gives many advantages. If we use multiple pairs of parallel output valves, we can keep the HT voltage within reasonably safe bounds, perhaps even at 320 V, if we are prepared to use many pairs of valves. With each additional pair of valves, the transformer primary impedance falls, as does the turns ratio, making it easier to design a good quality component. Statistically, total anode current per side becomes better balanced as we increase the number of valves, and deliberate selection will improve this still further.

Driving Higher Power Output Stages

Whether they are composed of paralleled devices or not, higher powered output stages always require more of the driver circuitry. When we investigated the Williamson amplifier, we found that it had a dedicated driver stage, but the large number of stages made stability a problem. Clearly, a better approach is needed. As before, listing the requirements helps solve the problem:

- We need low output resistance to drive the increased input capacitance of the output valves, and a cathode follower may be needed.
- We need to provide a large output voltage with low distortion; this invariably demands some form of differential pair.
- Wide bandwidth and high gain are also desirable, because we would like to have only one set of coupling capacitors to ensure Low Frequency stability, and the cascode might be ideal, although a carefully designed cascade of DC-coupled differential pairs could be even better.

We will first investigate a cascode differential pair with direct coupled cathode followers, sometimes known as the Hedge [17] circuit, after its designer (although the original Hedge circuit did not include cathode followers) (see Figure 6.34).

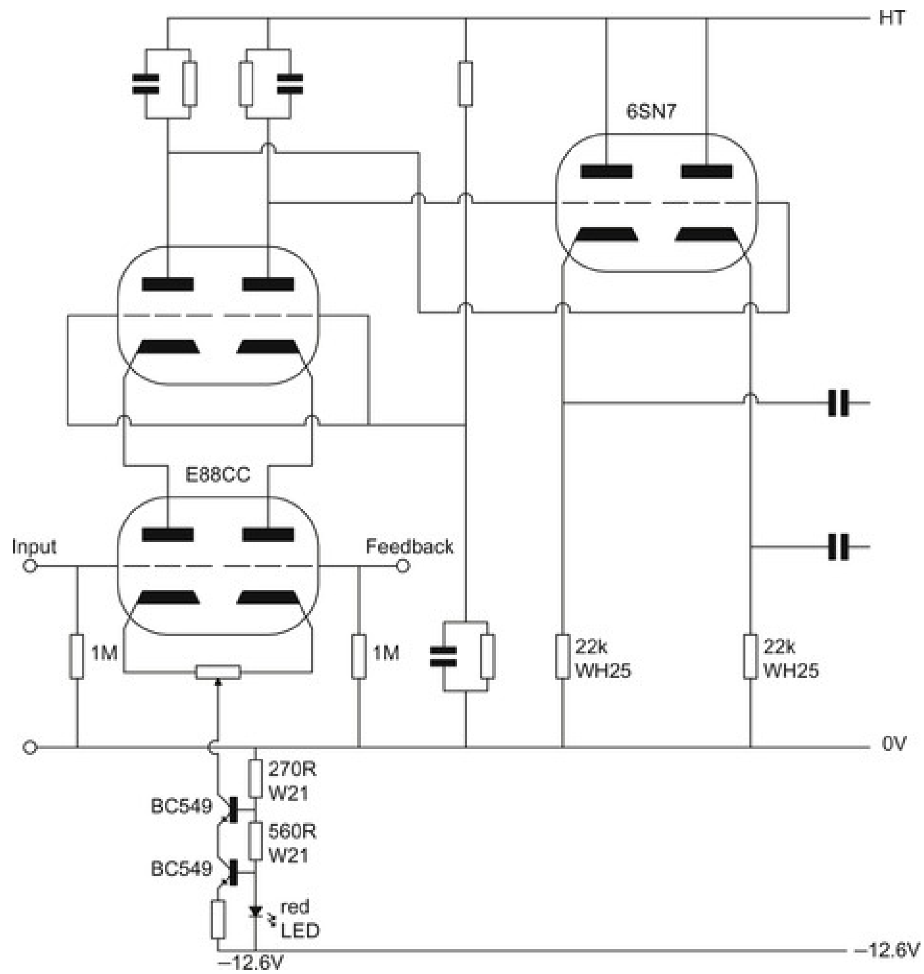


Figure 6.34 Hedge cascode differential pair plus direct coupled cathode followers.

Design of the individual parts of this circuit was covered in Chapter 2 , so we need not go into great detail on this circuit other than to make a few observations.

A single differential pair is not the ideal phase splitter, so we must take extra care over this to obtain a good result. The anode load resistors should be aged, matched and generously rated to avoid drift. The constant current sink should be made to have as high an output resistance as possible, and stray capacitance to ground from the cathode should be minimised to maintain a high impedance at high frequency. Matching the valves would be useful if possible.

Each pair of valves requires a separate heater supply. Sad, but true. The cathode followers need ≈200 V superimposed on their heaters, the upper pair of the cascode need ≈100 V, and the lower pair 0 V. Flirting with this rule will generate problems related to heater cathode insulation breakdown/leakage, and emission from the heater to the cathode will be summed with the intentional cathode current. You have been warned!

As was mentioned before, the only really satisfactory valve for use as the lower valve in a cascode is the E88CC; any other type wastes HT. The cathode voltage on the lower valves is usually quite low, $\approx 2.5\ \text{V}$, and because phase splitters inevitably have half the input signal voltage on the cathode, the tail of the sink needs to be taken to a subsidiary negative supply.

Feedback from the output can be applied to a grid, which makes the calculations of feedback network much easier, or the stage could accept a balanced input.

The Crystal Palace Amplifier

As with the Scrapbox Challenge amplifier, the design of any power amplifier begins at the output. Once all of the soft-start and general safety considerations have been taken into account, the cost of an amplifier supply is proportional to the square of its HT supply voltage. Thus, reducing the HT voltage releases money that can be spent elsewhere to achieve a better overall set of compromises. Possible contenders for the output stage are shown in Table 6.6 .

Table 6.6 Comparison of Possible Output Valves

	845	813	4 × EL34	13E1
$P_{a(max)}$	100 W	100 W	100 W	90 W
$P_{g2(max)}$	–	22 W	32 W	10 W
$I_{k(max)}$	120 mA	180 mA	600 mA	800 mA
$V_{a(max)}$	1,250 V	2,250 V	800 V	800 V
$V_{g2(max)}$	–	1,100 V	500 V	300 V
μ	5.3 [a]	8.5 [a]	10.5 [a]	4.5 [a]
g_m	3.4 mA/V [a]	4 mA/V [b]	46 mA/V [a]	35 mA/V [b]
r_a	1.6 kΩ [b]	2.1 kΩ [b]	230 Ω [a]	130 Ω [a]
V_h	10 V	10 V	6.3 V	26 V
P_h	32.5 W	50 W	37.8 W	33.8 W
C_{ag}	12.1 pF [a]	17 pF [c]	44 pF [c]	40 pF [c]
C_{Miller}	76 pF	162 pF	500 pF	220 pF

Note that the values in this table apply to NOS valves, and may not be applicable to recently manufactured valves.

[a] Manufacturer's claimed value.

[b] Value derived from manufacturer's data sheet.

[c] Value measured by author.

Of these valves, the 845 is a true triode, the 813 is a triode-strapped beam tetrode, the quartet of EL34s is a triode-strapped pentode, and the triode-strapped beam tetrode 13E1 actually contains a duet of paralleled valves.

Even though the AEI data sheet specifies $P_{a(max)} = 95$ W for a triode-strapped 13E1, all the options offer $P_{a(max)} \approx 100$ W, so they could all achieve approximately the same output power. NOS 845 are extremely expensive, but modern production is available. NOS 813 are readily available, but require the same expensive HT as the 845 (≈1,000 V). Sadly, the 13E1 is often more expensive than a quartet of EL34. Nevertheless, when the author saw a 13E1, it was lust at first sight. You will probably be more rational, and opt for a quartet of EL34.

13E1 Conditions

Push–pull output stages can be analysed using *composite* curves [18] . Composite curves are obtained by placing a second set of curves back to back below the first. Fictitious lines are then drawn between opposite true anode curves, and these are deemed to be the composite anode curves (see Figure 6.35).

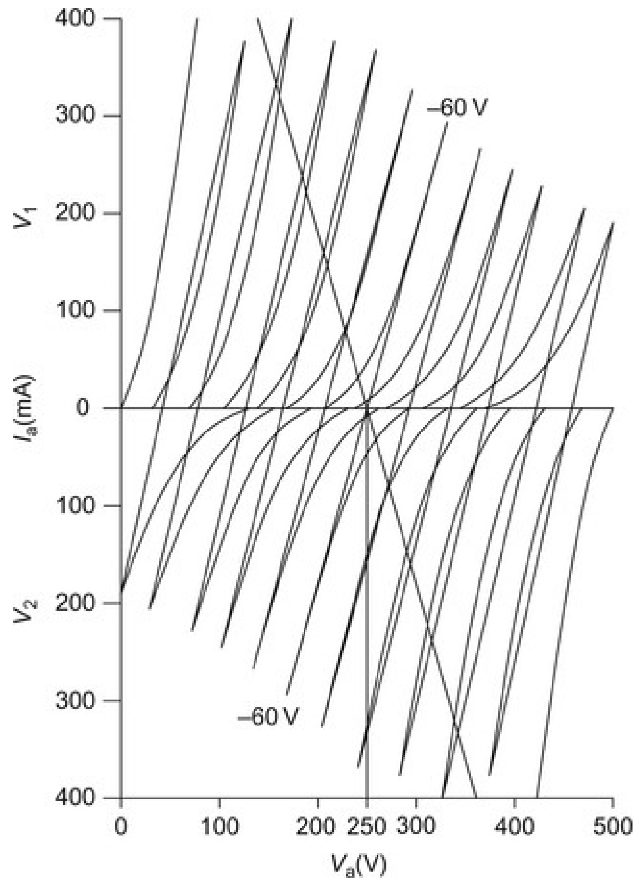


Figure 6.35 Composite anode curves for push–pull stage.

The operating point is where the composite line from V_{gk} =−60 V of V $_1$ to V_{gk} =−60 V of V $_2$ passes through I_a =0 at V_a =250 V. For maximum power output, R_L =2 r_a , and this loadline can be drawn by mirroring the composite V_{gk} =−60 V line about a vertical line passing through the operating point. In this particular instance, R_L =277 Ω, and the predicted power output is 42 W. Note that this particular operating point implies Class AB operation, since V_a =250 V, V_{gk} =−60 V implies I_a =49 mA. Quite apart from any reservations we might have about Class AB, the extremely steep loadline produced by this method greatly increases odd harmonic distortion. Although composite curves are a useful theoretical concept, and demonstrate the difference between Class A and Class B

very clearly, they imply ideal valves and are quite fiddly to produce and adjust, even on a computer.
We will find that it is much easier to analyse one half of the output stage, treating it as a single-ended stage. Once we have discovered the optimum single-ended loadline, we simply convert that into the push–pull requirement. In theory, we lose some accuracy by not using composite curves, but precise loadlines in power stages are futile because loudspeakers are not pure resistances, so the saving in drawing effort is well worthwhile.
Because this is an output stage, and we want to extract maximum power, we must operate the valve at $P_{a(max)}$ =95 W. In general, when shuffling loadlines and operating points for a given valve, we will find that output power is proportional to anode voltage, whilst distortion is inversely proportional. But cost rises with the square of anode voltage, so we will set V_a =400 V. Since $P_{a(max)}$ =95 W and V_a =400 V, we can use $P = IV$ to calculate that I_a =237.5 mA, and plot this point (see Figure 6.36).

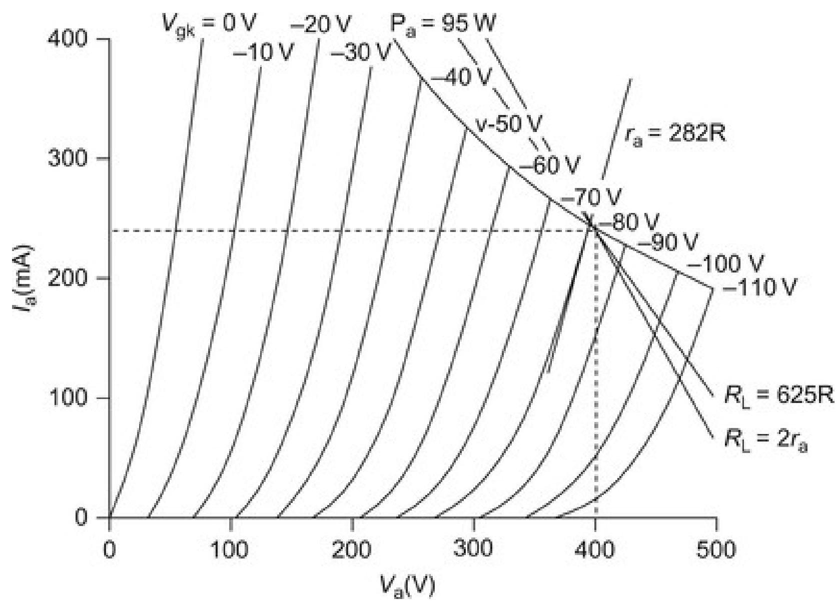


Figure 6.36 Setting the 13E1 operating points.

There is an anode curve near to our operating point, so we can find r_a . In this instance, we find that r_a =282 Ω. (We should not worry that this is significantly poorer than the manufacturer's claimed value of 130 Ω – they typically measure at V_{gk} =0, $I_a = I_{a(max)}$). Traditionally, we set R_L =2 r_a for maximum power, so the author tried this loadline. After extrapolating the curves (plausibly making them up), the 564 Ω loadline offered 14 W from the valve. Since P_a =95 W, this didn't seem too promising, but a gentler loadline of 625 Ω predicted 20 W with lower distortion at the same dissipation, so a push–pull pair should provide 40 W.

In a push–pull Class B amplifier, each output valve must see a load of 625 Ω, so each half of the transformer is wound with the correct number of turns to reflect this load. However, push–pull transformers invariably specify the load from anode to anode. Since reflected impedances change by the square of the turns ratio, doubling the number of turns quadruples the impedance. Thus, our output transformer would measure 4×625 Ω=2.5 kΩ from anode to anode. However, in Class A, both valves contribute to load current at all times, so the current is doubled, requiring a halving of load resistance from anode to anode, resulting in a theoretical optimum load of 1.25 kΩ from anode to anode for the Class A push–pull pair of 13E1s. Measurement showed that maximum power was obtained when the dummy load was set to 5 Ω (which corresponded to 1,124 Ω anode to anode once wiring resistances such as transformer windings were taken into account) (see Figure 6.37).

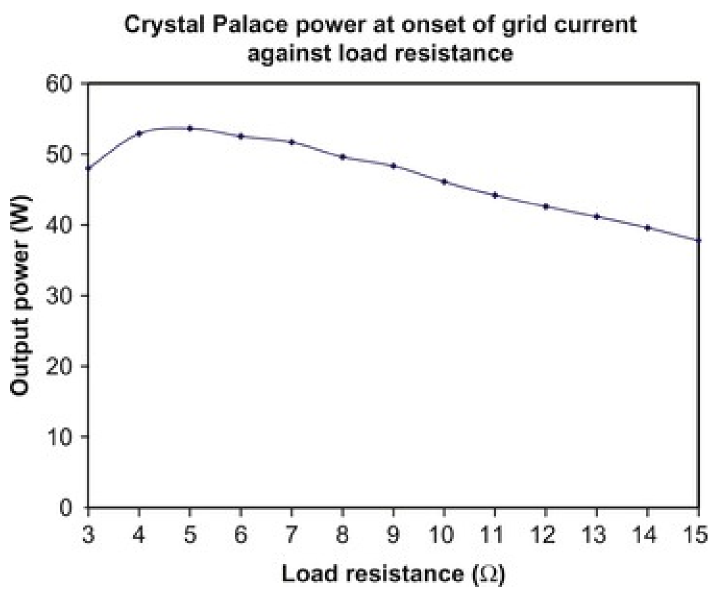


Figure 6.37 Measured Crystal Palace output power against load resistance.

As can be seen, power falls rapidly below the optimum load resistance but much slowly above, and this is why it is better to be pessimistic about expected load resistance. The author's Crystal Palace was conceived and built to drive Jordan JX92S having an average impedance above 200 Hz of 5.1 Ω once corrected by a Zobel network, so it is perfectly matched, but you might not wish to sail quite so close to the wind.

Eagle-eyed readers will note that $V_{g2(max)} = 300$ V, but that the final design sets $V_{g2} = 400$ V. Strapping pentodes and tetrodes as triodes and then exceeding their $V_{g2(max)}$ rating has been done before, most notably by Langford-Smith, using a pair of triode-strapped 807s at 400 V ($V_{g2(max)} = 300$ V) to replace KT66 in a Williamson amplifier. More significantly, Philips [19] gave performance data for a triode-strapped QE 05/40 beam tetrode audio amplifier operating at 400 V despite the fact that they quoted $V_{g2(max)} = 250$ V for the same valve used as a beam tetrode in an audio amplifier.
Nevertheless, the author still has misgivings about the long-term effects of exceeding $V_{g2(max)}$ ratings, so the emphasis in this design is on designing driver circuitry of irreproachable performance that could drive any of the valves listed in the table.

Driver Requirements

The stipulation 'irreproachable performance' is very vague, and needs to be converted into engineering requirements that can lead to engineering solutions:

(1) Minimal measured distortion

(2) Distortion to be composed of low order harmonics

(3) Push–pull output with good balance

(4) Large undistorted voltage swing

(5) Sufficient gain to enable global negative feedback if required

(6) Low DC output resistance to avoid problems with DC grid current

(7) Low AC output resistance to drive load capacitance

(8) Tolerance of output stage conduction angle changes from 360° to 0°

(9) Instantaneous recovery even after gross overload.

Finding a Topology that Satisfies the Driver Requirements

(1) Minimal Measured Distortion

This requirement implies nearly horizontal loadlines. A horizontal loadline implies an active load, but large resistive loads requiring a higher HT voltage are also a possibility.

(2) Distortion to be Composed of Low Order Harmonics

This requirement implies triodes rather than pentodes. Taking requirements (1) and (2) into account simultaneously suggests that triodes from the *SN7/*N7 family would be ideal.

(3) Push–pull Output with Good Balance

This requirement is best solved by two cascaded differential pairs with constant current sink tails. Since triodes produce predominantly second harmonic distortion, which the differential pair cancels, this satisfies requirement (1), but reinforces the preference for the *SN7/*N7 family because valves producing low third harmonic distortion are now needed due to odd harmonics summing constructively in a differential pair.

(4) Large Undistorted Voltage Swing

One of the strengths of the differential pair is its linearity when swinging large voltages. Nevertheless, the more HT voltage available the better, so this requirement implies that the driver stage should have HT>400 V. Since the output stage is likely to use ≈400 V, this implies that the driver stage needs a dedicated HT supply.

(5) Sufficient Gain to Enable Global Negative Feedback if Required

This requirement can probably be satisfied by two cascaded *SN7/*N7 differential pairs. If necessary, the gain could be doubled by using a high- μ valve dual triode such as 6SL7, 7F7, ECC83 or ECC808 in the input differential pair, but this would probably increase distortion because the high- μ valves tend to need 400 V HT, which might not be available.

(6) Low DC Output Resistance to Avoid Problems with DC Grid Current

Most of the larger power valves pass significant grid current even when the grid is negative, which is why manufacturers' data sheets recommend such low maximum values for their grid-leak resistors. Yet a small grid-leak resistor is an unnecessarily harsh load for the preceding stage.
Satisfying this requirement demands that the drivers be DC coupled to the output valve grids. Output stage HT is used most efficiently if the cathodes of the output valves are at 0 V because this means that $V_{HT} \approx V_a$. Therefore, V_a of the driver stage must be negative to bias the output valves correctly. The anodes of the driver differential pair can only be at a negative voltage if the tail of the differential pair is returned to a substantial negative HT supply, perhaps –300 V. If the driver stage uses the output stage supply for its positive HT, it now has a rail-to-rail HT of 700 V, which easily satisfies requirements (4) and (1).

(7) Low AC Output Resistance to Drive Load Capacitance

Although the *SN7*/N7 family produces low distortion, r_a is not particularly low, and fails this requirement. Valves such as the 6BX7 and 6BL7 have lower r_a, but their distortion tends to be very variable, and their Miller capacitance is punitive. Adding cathode followers to the outputs of the differential pair divorces the responsibilities for low distortion and low output resistance, allowing the differential pair to be optimised for linearity and swing, and the cathode followers for current driving ability.

(8) Tolerance of Output Stage Conduction Angle Changes from 360° to 0°

There is more to meeting this requirement than first appears. When we investigated phase splitters, we found that all phase splitters were sensitive to their loading, requiring Class A loads. The requirement for cathode followers has now been reinforced, since their buffering action allows the two-stage phase splitter to operate undisturbed by arbitrary output stage conduction angles.

The author considers that attempting to drive an output stage cleanly into Class AB2 is not worth the candle, so the cathode followers will be biassed only so that they can drive the output stage Miller capacitance cleanly, and no explicit attempt to drive grid current will be made. To maximise output swing, the cathode followers are likely to be operated with their cathodes at half the rail-to-rail HT voltage, so V_a =350 V. If we do not attempt to drive grid current, I_a =7 mA is adequate, resulting in P_a =2.5 W, which is just within range of the *SN7*/N7 family (see Appendix).

(9) Instantaneous Recovery Even After Gross Overload

This requirement means that the amplifier must not suffer from blocking. Therefore, the coupling capacitors must be positioned so that they couple to a stage that cannot be overloaded. By definition, the output stage can be overloaded, but we have already specified that it must be DC coupled. There is no advantage in placing the coupling capacitors between the second differential pair and cathode followers, because the anodes of the differential pair need to be at roughly the same voltage as the grids of the cathode followers in order to take advantage of the rail-to-rail HT voltage. The correct position for the coupling capacitors is *between* the two differential pairs.

An early iteration of the driver circuitry achieved 0.03% THD+N just below the point where output stage grid current imposed catastrophic loading. The author is simultaneously embarrassed and proud to report that measuring its distortion whilst it cleanly swung a differential voltage of +47 dBu (177 V_{RMS}) into a 100 kΩ load briefly gave a figure of 0.11% THD+N before his MJS401D audio test set briefly flashed ‘LEVEL HIGH’ and died.

Moral : Don’t get carried away and abuse your test equipment. In a more careful grid current test on the author’s ‘Plug and Pray’ test mule, identical driver circuitry only exceeded 0.2% distortion when a 2A3 grid going up to +57 V met its anode coming down to +57 V and conducted hard, clipping the driver.

Summarising the outcomes of the requirements, we need a cascade of differential pairs separated by coupling capacitors, using valves from the *SN7/*N7 family, and powered by a split-rail HT supply. The output of the driver differential pair will be DC coupled to cathode followers, which will be DC coupled to the output stage grids (see Figure 6.38).

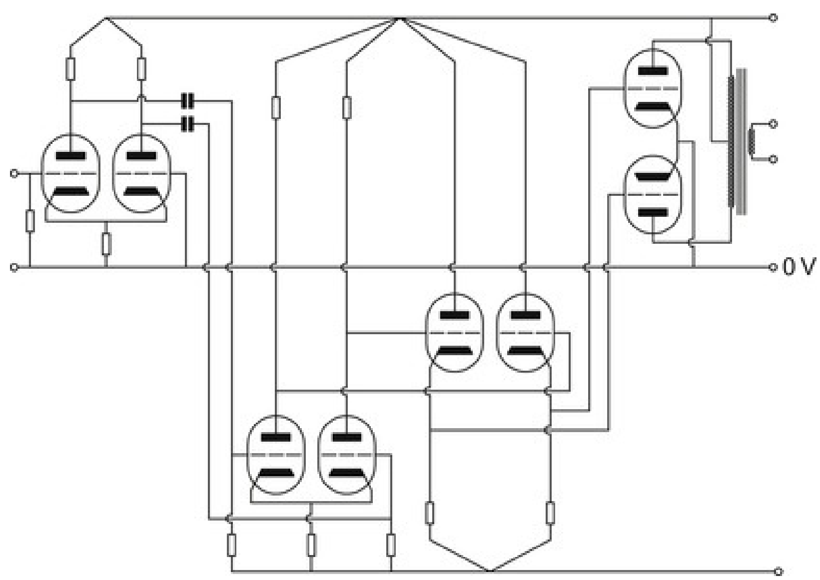


Figure 6.38 Conceptual diagram to show topology and position of coupling capacitors.

Circuit Topology: Power Supplies and Their Effect on Constant Current Sinks

The *SN7/*N7 family of valves produces primarily second harmonic distortion, which can be cancelled in a differential pair *if* no signal current is lost in the tail. Thus, we need active tails for both differential pairs, but because the grids are capacitor coupled from the previous stage, their grids are returned to the same supply as the constant current sink, V_k must be low, so these must be semiconductor constant current sinks.

The second differential pair is likely to have ≥500 V of HT voltage, so V_{gk} is likely to be ≈−10 V, which allows sufficient voltage for a cascode constant current sink to operate without an additional supply, especially when we consider that the input signal is already differential, so we do not need to cope with audio at the cathodes.

Sadly, the first stage is likely to have quite a low HT voltage, reducing V_{gk}, and if the input signal is unbalanced, half its amplitude must be on the cathodes, further reducing voltage available to the CCS. Thus, a cascode CCS would need a subsidiary supply, but the 334Z IC constant current sink can cope without. Just. However, the 334Z has an absolute maximum current rating of 10 mA, whereas we can design a cascode CCS to pass any current we like. Thus, our choice of valve operating points has already been restricted.

$V_{a(max)}$ and the Positive HT Supply

We have elected to direct couple the cathode followers to the grids of the output valves, so their cathodes will be at ≈−82 V, depending on individual output valves. If the anodes of the cathode followers are connected to the output stage HT, V_{ak} =482 V, which we cannot allow. (Even for the GTA or GTB versions of the *SN7, $V_{a(max)}$ =450 V.) However, this problem is not as bad as it sounds, because we do not need the cathode follower to swing 482 V_{pk}, so we can lower our positive supply to 160 V, which reduces V_{ak} to ≈250 V, allowing any valve from the *SN7*/N7 family to be used.

Next, we must consider the DC biassing of the output stage. The high g_m of our chosen output stage (whether 13E1, or a quartet of EL34s) means that the output stage current is extremely sensitive to changes in V_{gk}, and 30 mA/V is a very high mutual conductance in valve terms, so we cannot permit the grid bias voltage to drift. Because the driver circuitry is DC coupled from the output valves to the anodes of the second differential pair, a change in its V_a could potentially damage the output valves since (by definition) they operate at ≈ $P_{a(max)}$.

AC considerations dictate that the differential pairs require constant current sink tails. Once the driver differential pair has anode loads made of (constant value) resistors, Ohm’s law ensures that unchanging V_a can be achieved by regulating the 160 V positive HT supply. Our design has now evolved to the point where it *requires* an HT regulator to work safely. Drawn in full on a circuit diagram, regulators are intimidating, but only a valve regulator is more expensive than a decent HT capacitor and series resistor.

Symmetry and the Negative HT Supply

Because cathode followers operate under 100% negative feedback, they contribute very little distortion compared to the second differential pair, but because they drop ≈8 V across V_{gk}, they modify the anode voltage of the second differential pair from −82 V to −90 V. Since we require maximum linear swing from the second differential pair, the negative HT should be symmetrically opposite its anode voltage, requiring a negative HT of −90 V to 260 V=−350 V. The negative HT voltage is not at all critical and does not even need to be regulated because variations simply change V_{ak} without affecting I_a.

Although the precise voltage of the negative HT is not critical, it is essential that this supply is reliable. Failure would drive the output valve grids positive and the resulting anode dissipation would quickly destroy them. Thus, not needing a regulator for the negative HT has the bonus of improving reliability. Nevertheless, the output stage includes an HT fuse for protection in the event of negative HT failure.

The Second Differential Pair and Output Stage Current

Now that we have some firm HT voltages, we can begin detailed audio design, working back from the cathode followers towards the input stage.

The *SN7*/N7 family offers optimum linearity when $I_a \geq 8$ mA. The voltage across R_L for the cathode follower is −82 V – (−350 V)=268 V, so Ohm's law dictates that a 33 kΩ 6 W wirewound resistor dissipating 2.2 W would achieve I_a=8.1 mA.

Furthermore, the 13E1 presents 220 pF of Miller capacitance which must be driven cleanly. At 20 kHz, the reactance of 220 pF is 36 kΩ. At full power, the output stage demands 58 V_{RMS}, and maintaining this voltage across the reactance of the 220 pF capacitance demands 1.6 mA_{RMS} or 2.3 mA_{pk}. The capacitive load forces anode current to swing vertically ±2.3 mA on the loadline, which requires g_m to be constant. Fortunately, at I_a=8.2 mA, g_m *is* reasonably constant.

V_a=160 V – (–82 V)=242 V, and we already knew I_a, so we can determine V_{gk} at these conditions. Referring to the curves, V_a=242 V and I_a=8.1 mA roughly intersects the V_{gk}=−8 V curve. Knowing this voltage is important because it means that we now know that the grid voltage is −82 V≈8 V≈90 V. The grids are DC coupled from the anodes of the second differential pair, so their required anode voltage is also −90 V.

For the second differential pair, the voltage across R $_L$ is 160 V – (–90 V)=250 V. Each triode in the second differential pair can pass <8 mA because most of its distortion will be cancelled by push–pull action (this is not true for the cathode followers when the output stage enters Class AB). 50 kΩ MPC-5 anode load resistors are convenient, so the current required to set V_a to −90 V is 250 V/50 kΩ=5 mA, so the total tail current is twice this at 10 mA.

The tail current is highly significant. Increased tail current causes increased voltage drop across the anode load resistors of the second differential pair, causing their absolute voltage to become more negative. The cathode followers faithfully follow this negative change, and so the grids of the output valves become more negative, reducing their anode current. Thus, making the tail current adjustable allows us to set output stage current.

The output valves may not be perfectly matched, so interposing a variable resistor between the cathodes of the second differential pair allows us to adjust the balance of the stage, and therefore output stage current balance (see Figure 6.39).

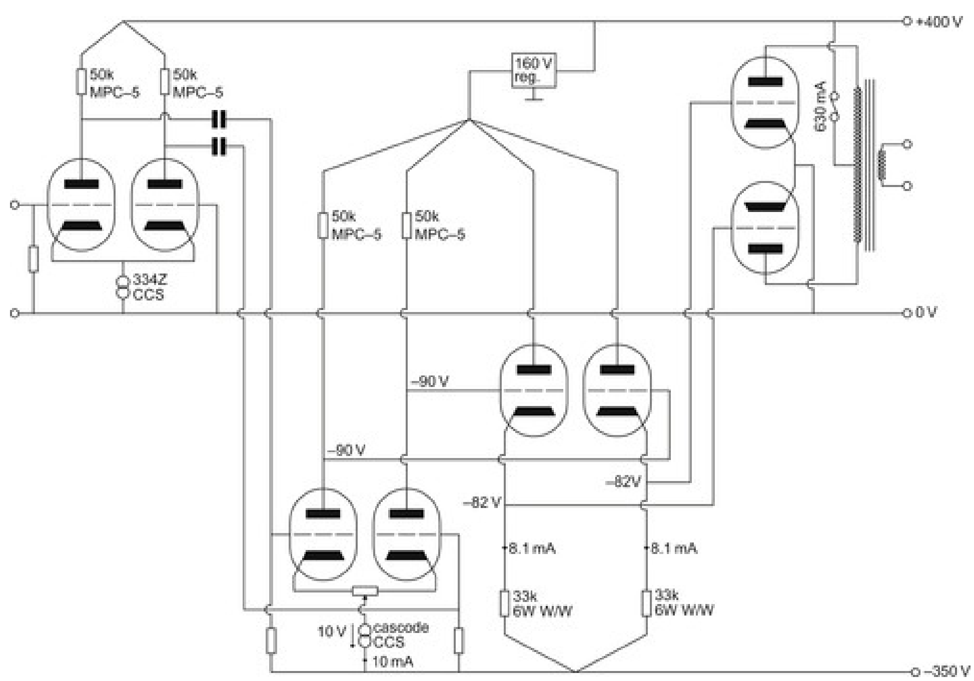


Figure 6.39 Setting the DC conditions for the amplifier.

Why Not Have Tighter Stabilisation?

Since we saw earlier that the output stage was sensitive to changes in V_{gk} , and that this is set by tail current, it seems intuitively obvious to stabilise tail current as tightly as possible. However, we should consider the effect of an increase in mains voltage in more detail. When mains voltage rises, the negative HT rail becomes slightly more negative, and more current flows through the resistor chain supplying the reference LED, so the voltage across its slope resistance rises slightly. But V_{be} for the transistor is unchanged, so the voltage across the current programming resistor rises identically, and tail current increases. Increased tail current *reduces* output stage current, but the rise in mains voltage caused the (unregulated) HT to the output stage to rise, which would have *increased* current, so the two effects tend to oppose one another, which is desirable. Thus, we discover that tightly stabilising the tail current and negative HT would be counter-productive because it would also require stabilisation of the high-current HT supply to the output stage.

As mains voltage varies, it changes not only the HT voltage but also the heater voltage. The DC conditions of the differential pairs are forced to be correct by the constant current sources and HT regulation, and the cathode followers have plenty of feedback, but the output valves are sensitive to heater voltage. Fortunately, because the 13E1 heaters can be configured to operate from 26 V, each pair of valves requires only 2.6 A, which can be regulated reasonably efficiently.

The First Differential Pair, Its HT Supply, and Linearity

By comparison with the second differential pair, the considerations involved in the design of the differential pair of the input stage are trivial. The stage only has to furnish ≈3.3 V_{RMS} from each output, so distortion really isn't a problem. Nevertheless, the *SN7*/N7 family requires $V_a \geq 150$ V for reasonable linearity even at very small voltage swings, so a >300 V HT would be ideal.

The second differential pair required a negative HT of −350 V, and if a traditional centre-tapped rectifier/transformer combination was used to provide this voltage, it could also provide the positive HT. As a further measure to protect the output valves, we could use a valve rectifier for the positive HT. If power was briefly interrupted to the amplifier, the valve rectifier would ensure that at the instant of power returning, the output valves would initially be biassed off, but would gently be turned on as the rectifier warmed. Unfortunately, valve rectifiers drop more volts, so the positive HT is likely to be ≈300 V.

Although DC conditions do not require the HT for the first differential pair to be regulated, it is probably the best way of achieving a sufficiently low HT ripple. To ensure that the regulator does not drop out of regulation when mains voltage drops, we could set its output to +270 V.

The first differential pair now has quite a low HT voltage, and the only way to maintain linearity and voltage swing is to reduce I_a. Reducing I_a allows V_a to swing closer to 0 V, which increases maximum output swing, and distortion is generally inversely proportional to maximum output swing. Slithering a transparent ruler over the anode characteristic curves resulted in an operating point of V_a =125 V and I_a =2.9 mA for each valve, using a 50 kΩ load resistor.

Valve Matching

2.9 mA per triode is well below the ideal 8 mA, but the differential pair cancels the predominant second harmonic distortion, and the voltage swing is very low, so this is tolerable, but the required distortion cancellation would benefit from matched valves in the differential pair.

Although the Loctal 7N7/14N7 valves tend to be quite well matched, individual triodes such as the 6J5GT allow even better matching. If we decided to standardise on single triodes, we would need twelve 6J5GTs for a stereo amplifier. The advantage of needing so many valves of a given type is that the probability of finding matched pairs increases substantially with the number of valves, so buying the required dozen valves offers a far better chance of finding matched pairs than buying two. It was the decision to use a dozen 6J5GTs of complementary appearance to the 13E1 output valves that prompted the name 'Crystal Palace' for this amplifier.

In general, when two apparently identical valves have the same anode voltages in a differential pair, their gain is likely to be matched under those operating conditions. Thus, we can test 6J5GTs in the amplifier by inserting a valve deemed to be the reference valve in one side of the differential pair, and sequentially test all the other valves against this reference. The valves with closest anode voltages are in pairs.

The Essential Twiddly Bits

We have made the broad brushstrokes, and chosen our stage topology, valve type, anode currents and load resistances. It is now time to get down to the nitty-gritty and ensure that those conditions are met. For this we need to:

- set the DC conditions of each stage by designing their constant current sinks;
- consider thermal stability of the constant current sinks;
- consider RF stability by including grid-stopper resistors and bypassing the

HT supplies correctly;

• design the HT regulators.

The Cascode Constant Current Sink and Stabilisation Against Mains Variation

We know that $V_{gk} \approx -10$ V for the second differential pair, so we should be able to design a cascode that operates at this voltage, thus avoiding a subsidiary supply.

For reasons that will become apparent in a moment, we need a low reference voltage, so we will use an infra red LED, which we will bias from the 0 V rail via a large resistor. Because the resistor has ≈350 V across it, it can only pass quite a small current to keep within its power rating. If we choose a 150 kΩ 3 W resistor, it will pass 2.3 mA whilst dissipating 0.83 W. 2.3 mA would not normally be considered to be an ideal reference current for an IR LED because $r_{internal}$ rises significantly at lower currents (16.4 Ω at 2.3 mA, as opposed to 5.4 Ω at 10 mA). However, because the slope resistance helps compensate for mains voltage variations, this isn't a problem.

To consider the effect of mains variation, we will assume a 1% rise in mains voltage.

In the output stage, we want to hold I_a constant despite changing V_a, and find the V_{gk} that would oppose this change. This is effectively the definition of μ.

Since the output stage HT=400 V, 1% rise implies +4 V. For the 13E1, $\mu \approx 3.9$, so V_{gk} must fall by ≈1 V to combat the anode change.

1 V fall at an anode of the second differential pair would be caused by an increase in individual anode current of

$$I = \frac{V}{R} = \frac{1}{50\,k\Omega} = 20\ \mu\text{A}$$

But there are two valves, so the tail current must increase by twice this, that is 40 μA.

For our infrared LED $V_{ref.} = 1.10$ V at 2.33 mA, so the current programming resistor in the emitter circuit of the cascode must be

$$R = \frac{V}{I} = \frac{1.10 - 0.7}{10\ \text{mA}} = 40\ \Omega$$

A change of 40 μA in the 40 Ω programming resistor would be caused by a change in voltage of $V = IR = 40\ \mu\text{A} \times 40\ \Omega = 1.6$ mV.

Assuming constant base–emitter voltage, $V_{ref.}$ must rise by 1.6 mV to combat the change in output stage current.

The 150 kΩ resistor passes 1% more current due to the 1% rise in mains voltage. It normally passes 2.33r mA, so the increase in current is 23.33r μA.

We now know the change in current and change in voltage across an unknown resistance, so we can find its value:

$$R = \frac{V}{I} = \frac{1.6\ \text{mV}}{23.33\ \mu A} \geq 68.6\ \Omega$$

The IR LED passes 2.33 mA, and contributes r_{slope} =16.4 Ω, so we need 68.6 – 16.4 =52 Ω.

But 52 Ω drops 121 mV at 2.33 mA, so $V_{ref.}$ rises to 1.10 V+0.121 V=1.22 V.

Therefore, the voltage across the programming resistor becomes 521 mV, and since it must pass 10 mA, its required resistance changes from 40 Ω to 52 Ω (see Figure 6.40).

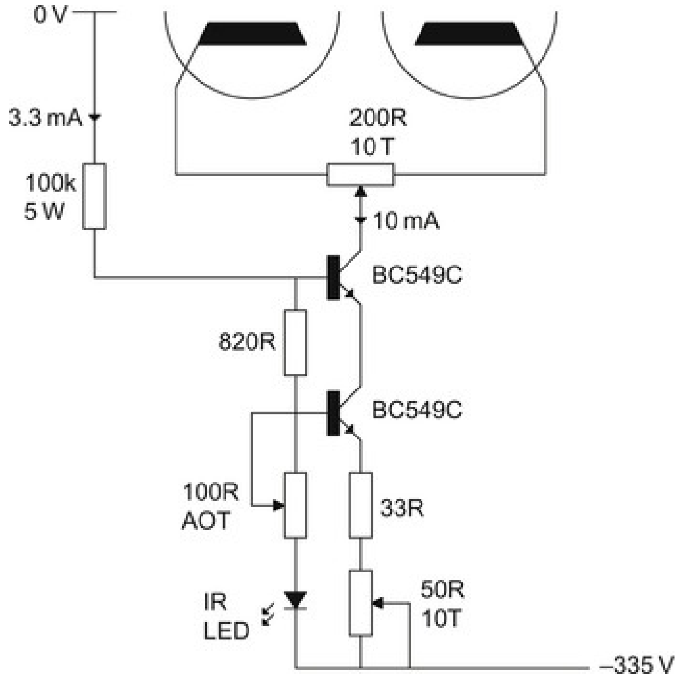


Figure 6.40 Biassing the second differential pair

As can be seen, the values of the programming and compensating resistors interact, so we will need to use variable resistors and adjust them on test as follows:

(1) Set the compensating resistor to its maximum value, set the programming

resistor to its minimum value

(2) Adjust the programming resistor for correct output stage current

(3) Using a variac, raise mains voltage by 5%

(4) Adjust the compensating resistor to restore correct output stage current

(5) Restore correct mains voltage.

Repeat steps (2)–(5) until the output stage current variation with mains voltage is minimised.
The transistors do not need to withstand large voltages, so BC549C is ideal.
We know that we need to be able to adjust the relative voltages on the grids of the output valves to equalise individual anode currents, so a variable resistor between the cathodes of the second differential pair allows this variation. If we use a 200 Ω variable resistor and notionally move the wiper to one extreme end, the resistor will pass the current of only one 6J5GT, which is ≈5 mA, so it will drop ≈1 V. The gain of the differential pair is ≈18, so there will be a grid-to-grid change of ≈18 V at the output valves. Because the wiper could be moved to the opposite end of the resistor, we could achieve the same voltage change but in the opposite direction. Thus, each valve can effectively have its V_{gk} varied ±18 V, which is quite sufficient to achieve anode current balance. (As originally built, the author used a 100 Ω variable resistor to balance output stage anode current but this did not always provide sufficient variation.)
The collector of the lower transistor hardly has to change voltage, so V_{CE}=2 V is perfectly satisfactory for this transistor at 10 mA. Because the collector of the lower transistor is connected to the emitter of the upper transistor, V_{CE} is equal to the voltage between the two emitters. Because we drop 0.7 V across the base–emitter junction of both transistors, V_{CE} for the lower transistor is equal to the voltage between the two bases. When we put a resistor between the bases, we know that it passes the 2.33 mA sourced via the 150 kΩ resistor, so its required value is 2 V/2.33 mA≈820 Ω.

The 334Z Constant Current Sink and Thermal Stability

Referring to the data sheet [20] , the current programming resistor for the 334Z can be calculated using:

$$R = \frac{227\ \mu\text{V} \times T}{I_{\text{set}}}$$

where T is the absolute temperature.

If we assume an ambient temperature of 300 K (27 °C), this simplifies to

$$R = \frac{68.1}{I_{\text{set (mA)}}}$$

Thus, to set 5.8 mA, we need ≈12 Ω resistor. However, it was worth seeing the first equation because it reminds us that all electronics drifts with temperature. The usual reason for drift is the temperature dependence of V_{be} for silicon transistors, but this can usually be compensated by adding a silicon diode in the reference chain. The fundamental assumption is that the diode is at the same temperature as the junction producing the error, so the compensating diode should be glued to the offending device with epoxy adhesive, and the entire mass insulated from convection currents by a small expanded polystyrene shroud.
Sure enough, the data sheet gives a circuit that compensates for thermal drift, and simply requires that the additional resistor is 10 times the programming resistor (see Figure 6.41).

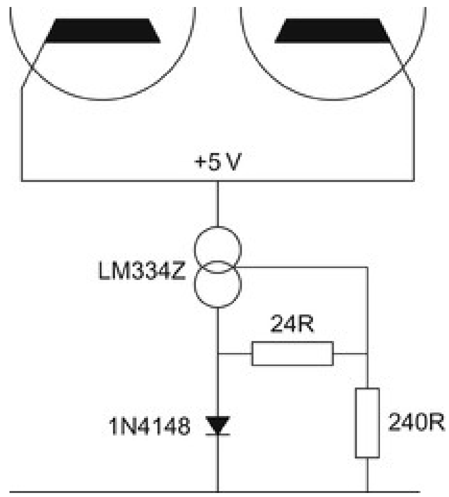


Figure 6.41 Compensating the 334Z against temperature variations.

Having compensated the (uncritical) 334Z, we should consider compensating the cascode constant current sink because this *is* critical. The traditional method of compensating the cascode adds a silicon diode in series with the reference diode to compensate for the changing V_{be} of the lower transistor. The assumption is made that the reference diode has zero drift with temperature, and this is very nearly true if a 6.2 V Zener is used, but we have chosen to use an LED. Because the forward drop of an LED falls with increasing temperature, it already tends to compensate the transistor, so no extra components are required.

High Frequency Stability

High-g_m valves such as the 13E1 are prone to parasitic oscillation, but this can be prevented by 1 kΩ grid-stopper resistors. Because cathode followers operate with 100% negative feedback and feed a capacitive load, there is a danger of oscillation, so they need 10 kΩ grid-stoppers. Differential pairs rarely need grid-stoppers, but shorting a grid directly to ground is inviting RF oscillation (no damping whatsoever), so 10 kΩ grid-stoppers were fitted to the first differential pair.

Adding 10 kΩ grid-stoppers to the second differential pair would reduce the $f_{-3\ \mathrm{dB}}$ point due to Miller capacitance and source resistance from an acceptable 130 kHz to an unacceptable 60 kHz, so grid-stoppers were omitted.

Another possible cause of High Frequency oscillation is non-zero power supply impedance. To counter this problem, the output of the +160 V regulator should be a star point, and the −350 V HT should also feed a star point. A 470 nF capacitor can then be connected between these star points to ensure stability of the second differential pair and its associated cathode followers. Similarly, a 470 nF capacitor should be connected from the centre tap of the output transformer to the junction of the 1 Ω current sense resistors in the output stage, and another 470 nF capacitor from the star point of the 270 V regulator to the bottom of the 1N4148 diode in the 334Z constant current sink. Finally, we need a THINGY to set V_{hk} appropriately. THINGY design has been covered previously, so we now have our final audio design (see Figure 6.42).

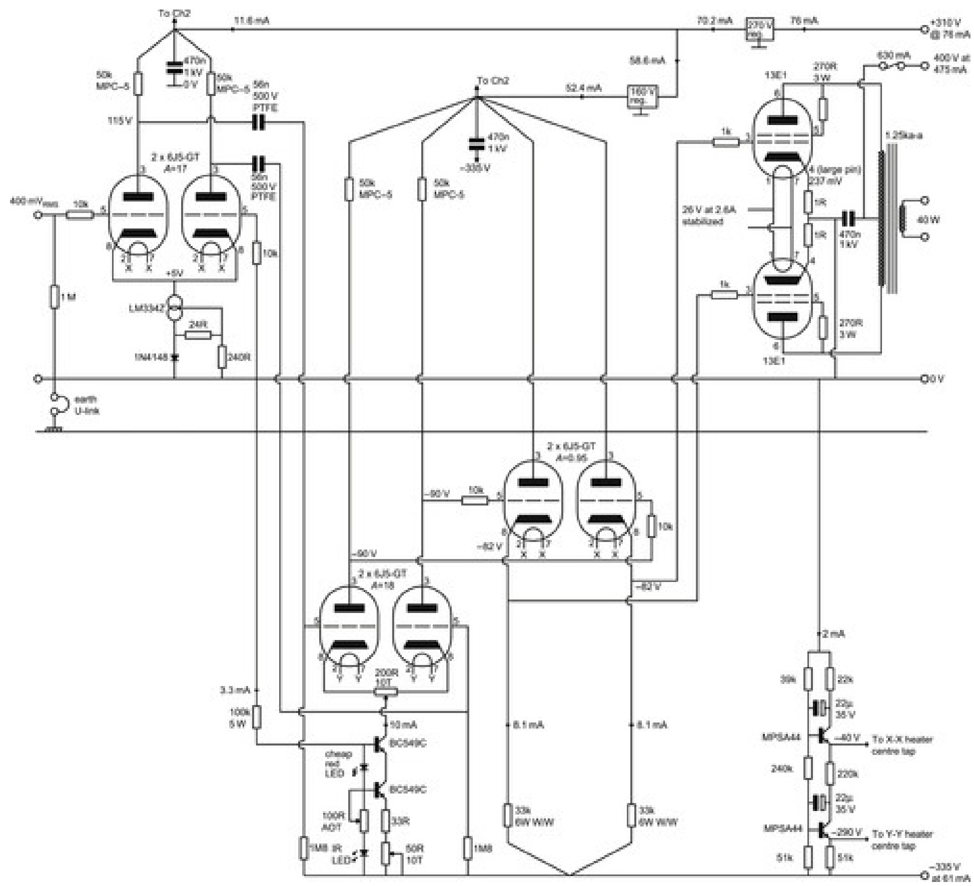
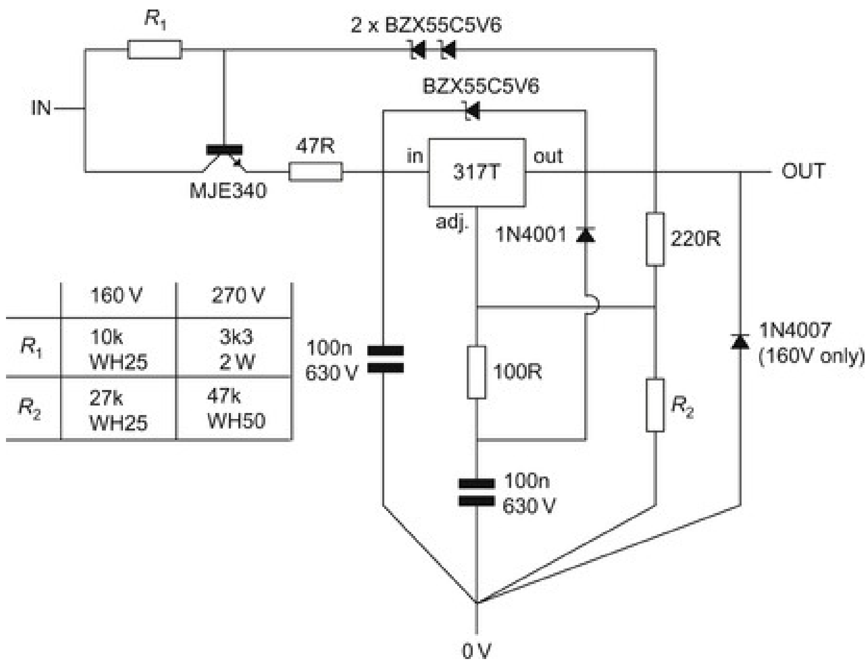


Figure 6.42 Final audio circuit of 'Crystal Palace' power amplifier.

HT Regulators

The second differential pair requires a 160 V regulator free from DC drift, and a variation of the Maida regulator was used. The first differential pair is not critical, but there is no real reason to use an alternative, so another Maida regulator is used. Supplying the 160 V regulator via the 270 V regulator ensures that the 270 V regulator passes sufficient current to operate correctly. With two component changes, the same design is used for both the 270 V and 160 V supplies (see Figure 6.43).



	160 V	270 V
R_1	10k WH25	3k3 2 W
R_2	27k WH25	47k WH50

Figure 6.43 Regulator design.

Stereo versus Mass

The author began the metalwork and layout of the Crystal Palace amplifier as a stereo amplifier *before* buying accurate scales. At over 90 lb, this is not an amplifier for the faint-hearted.

The reason for building the amplifier as a stereo amplifier is that the totally balanced audio topology renders the amplifier insensitive to power supply noise. There is therefore no need to have separate left and right power supplies, and a considerable reduction in support circuitry can be achieved. Last but not least, the author had many of the parts to achieve a stereo amplifier, but a pair of mono chassis would have doubled the metalwork and required the purchase of two large HT chokes. You might have a different opinion about the benefits of a stereo chassis.

Power Supply Design

Having decided on a stereo chassis, we need an HT supply capable of supplying 1 A at ≈400 V. A quick check with a spreadsheet revealed that a choke input HT supply would need a 2 H 1.5 A choke and a 455 V_{RMS} mains transformer. The author took one look at the size of his 1 H 1 A choke, and decided that an even larger choke was not acceptable. A capacitor input HT supply was therefore necessary. 1,200 V soft recovery diodes are used for the bridge rectifier. In order to protect the 13E1 output valves, the output stage's HT transformer can be switched by a thermal delay relay.

The 13E1 heaters require 26 V at 2.6 A per channel, so a 2 × 25 V_{RMS} 300 VA toroid was chosen because a 160 VA transformer would have been marginal, and the 300 VA transformer cost no more than a 250 VA, yet was slim enough to fit inside the 2″ chassis. The regulator arrangements are standard, but the reservoir capacitors are deliberately small to reduce regulator dissipation.

Traditional centre-tapped HT transformers are intended for use with valve rectifiers, so only one-half of the winding is in use at any instant. If we derive a positive *and* a negative supply, both windings are in use simultaneously, so we must be careful not to exceed the VA rating. The simplest way to ensure this is to say that the sum of the positive and negative currents must be less than the winding rating. Thus, if we need 78 mA for the positive supply and 61 mA for the negative, the total current is 139 mA, so a 150 mA 275 V–0–275 V winding is fine. Fortuitously, the salvaged HT transformer also had the bonus of a pair of 6.3 V 4 A centre-tapped windings suitable for the driver valves, EZ81 rectifier and delay relay (see Figure 6.44).

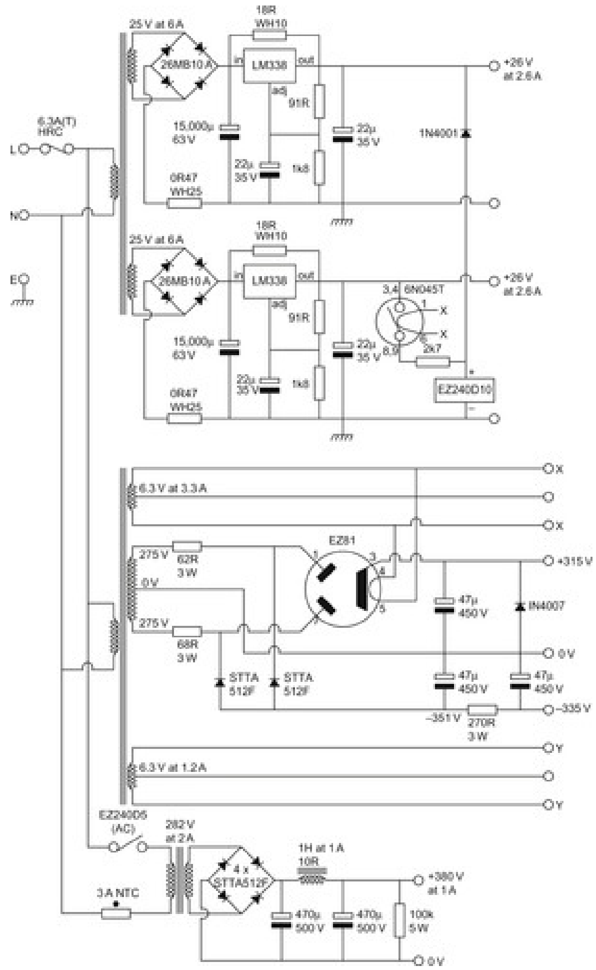


Figure 6.44 Main PSU design.

Designer's Observations

In one significant respect, this amplifier is even worse than the Scrapbox Challenge – it is far too heavy. The author aims to avoid repeating this mistake. The other problem is sample variation between 13E1s. Given that it was designed to be a power supply regulator valve, close tolerances were not necessary, so we can't really complain. Those two problems aside, the author is very pleased with this amplifier, and especially with the driver philosophy.

Exceeding V_{g2}

Over the years, the author's misgivings about exceeding V_{g2} ratings have proven well-founded. Although some valves were happy at 180 mA, others needed to be backed off to 150 mA, and one or two just got hot under the collar, no matter what current they passed.

We can easily drop V_{g2} to 200 V $_{DC}$ by interposing a large Zener diode between it and the output transformer, but although this would work for small signals, full power would cause g_2 to switch off. Thus, we would also need to attenuate the audio signal fed to g_2 and this would mean changing from triode operation to beam tetrode Blumlein configuration using 43% output transformer taps.
Realistically, the author has to concede that although the 13E1 is very pretty and permits a low primary impedance output transformer (ensuring excellent output transformer performance), it could be bettered. Keeping that excellent output transformer requires high g_m in the output valves, so we are back to the choices faced at the start of this design, and multiple pairs of EL34s *are* the best solution.

GM70

If multiple pairs of output valves are deemed unacceptable but high anode voltages (and impedances) *are* acceptable, then the Cold War GM70 125 W triode operated at V_a =1,250 V, I_a =100 mA becomes a possibility. Under these conditions, V_{gk} =−125 V, so the 90 V $_{RMS}$ driver swing required for class A is a little more than is comfortable from 6J5GT but well within the capability of 6S4A ($V_{a(max\ DC)}$ =550 V). Thus, 6S4A could be substituted for the cathode followers (10 mA each) and driver (6 mA each). Grid current is always a problem with transmitter valves, but a 6S4A cathode follower ought to be able to achieve a 500 Ω output resistance, so if 0.1% THD was tolerated due to grid current, then this would be 100 mV $_{RMS}$ developed across the source resistance of 500 Ω, equating to a non-linear grid current of 200 µA – which is considerably more than the 4 µA specified on the GM70 data sheet, suggesting that grid current should not be a problem.

Measuring I_k

It didn't take many burnt knuckles before the author realised that plugging DVMs into 4 mm sockets directly adjacent to hot output valves was a daft idea, so he splashed out on the luxury of a postage stamp DVM per valve. Since their FSD is 200 mV, the 1 Ω cathode resistor converted them directly into 200 mA meters. They only needed 1 mA apiece so each obtained its supply via a resistor and Zener from the nearby 26 V 13E1 heater supply. Although not cheap, this was a very worthwhile modification.

Global Neaative Feedback

As presented, the amplifier does not have global negative feedback. The traditional way to apply feedback to a differential pair is to lift the unused input from ground and apply feedback to this point (transistor power amplifiers use this technique). The author tried this configuration for a while but it is not ideal.

The sensitivity of the amplifier before feedback is 400 mV$_{RMS}$. If we applied 20 dB of feedback, sensitivity would fall to 4 V$_{RMS}$, but there would still be 400 mV$_{RMS}$ between the two grids of the input differential pair, so the feedback signal must be 3.6 V$_{RMS}$ and it must be of the same polarity. More significantly, we could observe that the grids are moving up and down in voltage together at ≈4 V$_{RMS}$, with very little voltage (400 mV$_{RMS}$) between them. More formally, we would state that our feedback has imposed a large common-mode signal on the grids. If the grids have a large common-mode signal, then the cathodes must be following this signal, and this is where the problem starts. The LM334Z only has 3.5 V$_{DC}$ across it, and drops out at 1.2 V$_{DC}$, so it has insufficient voltage compliance to cope with a superimposed ≈4 V$_{RMS}$ audio signal. It is this sort of nasty little hidden problem that makes the author so keen to use cascode constant sinks supplied from an auxiliary negative supply – such a supply gives the CCS enough voltage compliance to cope with a 4 V$_{RMS}$ signal and have a substantially unchanging output capacitance ($C_{cb} \propto 1/\sqrt{V_{cb}}$).

Fortunately, there is a way around the problem, and it fitted a new requirement rather neatly. Rather than applying series feedback, we can apply shunt feedback to make an inverting amplifier (see Figure 6.45).

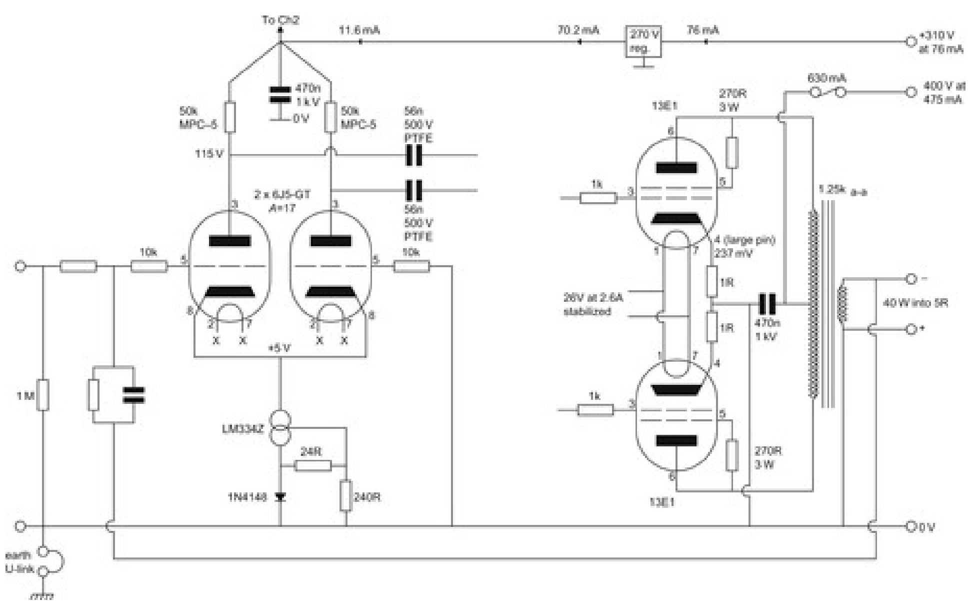


Figure 6.45 Applying global negative feedback to the Crystal Palace.

The effect is to make the input grid a virtual earth, so we can now apply as much feedback as the output transformer will permit without disturbing the LM334Z. There is a slight disadvantage in that the amplifier now inverts, but swapping the colour of the loudspeaker terminals and reconnecting loudspeaker wires is hardly difficult. However, we can now take a further step and apply two loops of feedback, one from each loudspeaker terminal to each grid, to give the amplifier a balanced input. This would be particularly handy if the amplifier needed to be driven from a digital crossover with a balanced output (see Figure 6.46).

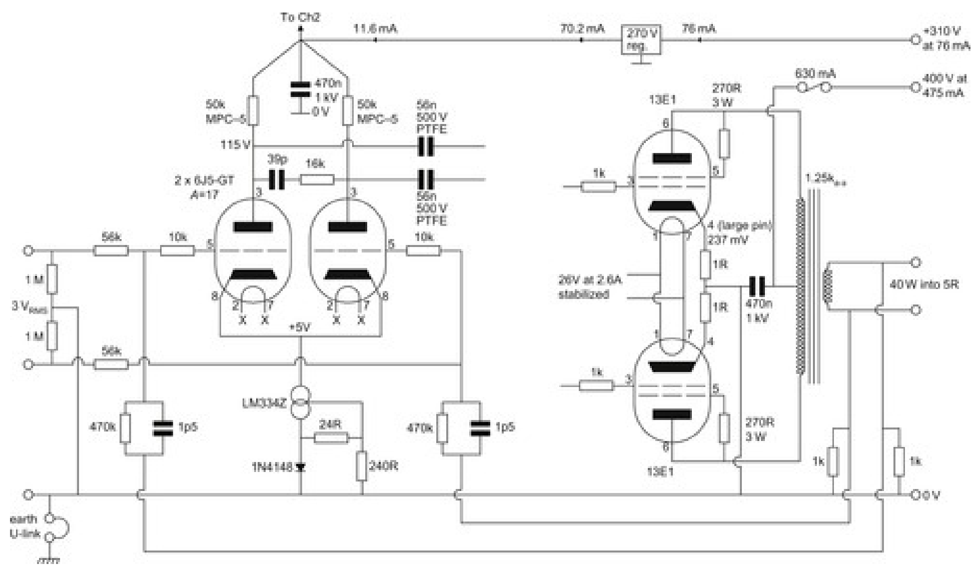


Figure 6.46 Applying balanced global negative feedback to the Crystal Palace.

Note that very small PTFE trimmer capacitors are required across the feedback resistors, and it is the practical minimum value of these capacitors that determines the maximum value of feedback resistors which, in turn, determines the series input resistors and input resistance of the amplifier. As usual, it has been necessary to add a step network at the anodes of the first stage, but because it is a differential pair, a single network between the anodes is possible, leaving 0 V across the 39 pF capacitor. The author applied 18 dB of balanced feedback but was initially mystified by the results. Bear in mind that the amplifier output no longer has one leg connected to ground, so both legs had to be monitored on the oscilloscope, yet it simply wasn't possible to achieve identical (and perfect) 10 kHz square waves on each leg. A little thought suggested that this might be because of imbalances in the output transformer, and that since the load responds to the difference in signals, this is what should be monitored (see Figure 6.47).

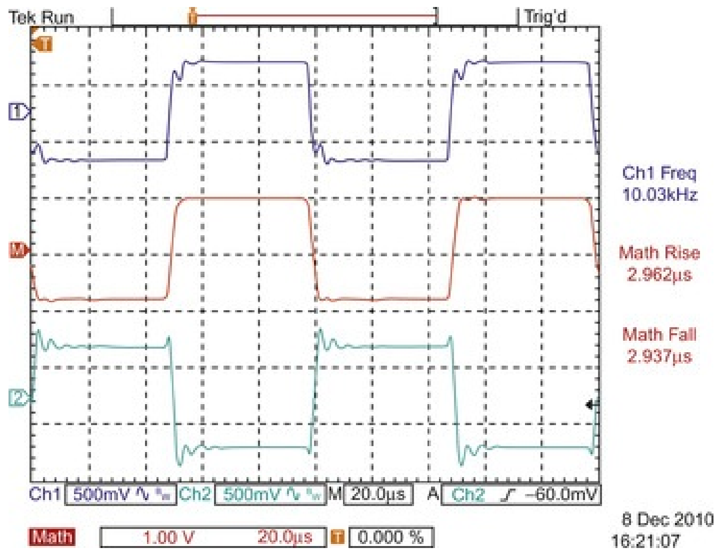


Figure 6.47 Crystal Palace output waveforms (Math=Ch1−Ch2=the voltage across the load).

As can be seen from the oscilloscope traces, the difference between the two channels (Math, centre trace) *is* a clean 10 kHz square wave even though the two legs do not match. Note that 1 kΩ resistors from each output leg to ground are necessary when balanced feedback is used in order to reference the output to ground, otherwise inter-winding capacitances within the transformer can cause instability. It would be better if the transformer secondary could be centre-tapped to ground, but the author needed a secondary configuration that didn't permit that.

Conclusions

What the world needs is a smaller, lighter Crystal Palace.

The Bulwer-Lytton Scalable Parallel Push–Pull Amplifier

> *“ It was a dark and stormy night; the rain fell in torrents – except at occasional intervals, when it was checked by a violent gust of wind which swept up the streets (for it is in London that our scene lies), rattling along the housetops, and fiercely agitating the scanty flame of the lamps that struggled against the darkness. ”*
>
> *‘Paul Clifford’ E.G. Bulwer-Lytton (1830)*

The first laboratory test of this amplifier’s output stage was made on a typical British summer afternoon – the sky was dark and it was about to bucket with rain. Once suggested, the Bulwer-Lytton name stuck.

Background

It is traditional for amplifiers and loudspeakers to be universal – any loudspeaker can be used with any amplifier. But such versatility carries the price of forcing compliance with an interface standard of zero source resistance rather than allowing the loudspeaker and amplifier to be designed together for the optimum performance of both. By contrast, the Bulwer-Lytton amplifier has some unusual characteristics, and rather than engineer them out, the author chose to design a loudspeaker that could capitalise on them. That loudspeaker uses the Fostex FE166E full-range driver and full design and construction details of the author’s Arpeggio loudspeaker may be found in the ‘Articles’ section of diyAudio.com . Experiment showed that a pair of push–pull 6S4As could produce 2 W (see Figure 6.48).

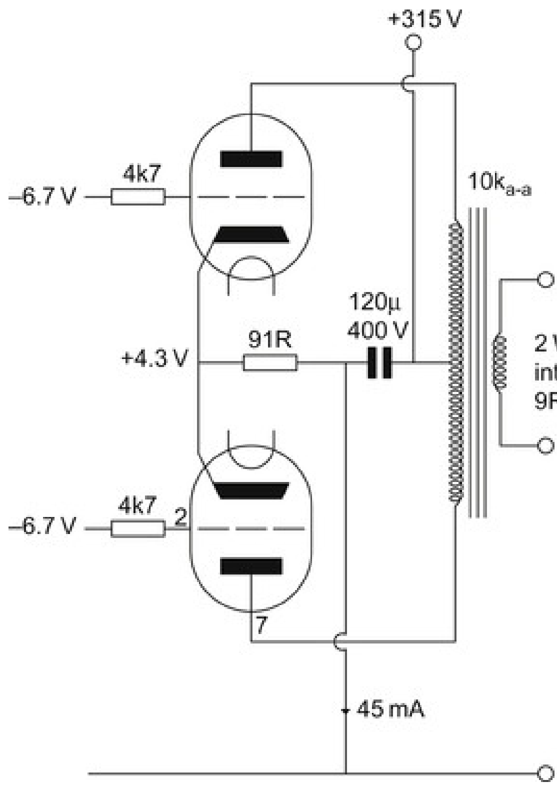


Figure 6.48 6S4A output stage.

More significantly, the inclusion of a harmonic equaliser resistor enabled that 2 W be obtained with distortion as shown in Table 6.7 .

Table 6.7 Distortion of Push–Pull 6S4A Output Stage Using Harmonic Equaliser

	H2	**H3**	**H4**	**H5**	**H6**	**H7**
Level (dB)	−42.6	−68.6	−66.6	−78.6	−97.6	−80.6

Note that not only is the distortion below 1%, but that it is dominated by second harmonic.

Designing the Followers to Drive the Output Valves

The amplifier was required from the first to be scalable, with driver circuitry capable of driving many pairs of output valves, and that immediately implies cathode followers.

Having made the decision to use small output valves but drive them DC coupled from cathode followers, it seems terribly wasteful to have a −300 V HT supply simply for cathode followers that only need to swing to 26 V $_{\text{pk–pk}}$. If we replace the resistive loads in the cathode followers with a semiconductor constant current sink not only do we improve linearity, but also we improve efficiency because the sink only needs a few more volts than the maximum required peak negative swing to ensure correct operation. Thus, if we need to swing to −26 V, a –35 V supply would be perfectly adequate and far cheaper than −300 V at the same current. We still need at least 100 V across the cathode follower, but we've avoided wasting voltage across a resistive load. Even better, a −35 V supply would enable a cascode constant current sink as the tail for the input differential pair – slightly cheaper but far better than the Crystal Palace's 334Z.

We saw in the Crystal Palace amplifier that DC coupling to the output valve grids means that the cathode follower's cathode is slightly negative, so we must either choose a valve that can tolerate V_{ak} >300 V (perhaps the expensive *SN7GTA), or we drop the 300 V HT from the output stage down to a more suitable voltage for the cathode follower. Driving a large number of 6S4A output valves implies a large Miller capacitance that must be charged and discharged, so this implies a substantial quiescent current through the cathode follower and, combined with the V_{ak} problem, this suggests a lot of power is being wasted as heat. The obvious solution is to determine the ideal positive voltage for the cathode followers then provide it from a dedicated mains transformer winding.

Having accepted that the cathode followers need dedicated transformer windings for their positive and negative supplies, it's time to question the superiority of a cathode follower, and whether or not an FET source follower might be just as good (but a lot cheaper to implement). It was time for some measurements.

Comparing Cathode and FET Source Followers

We know that a cathode follower has 100% negative feedback, so high open-loop gain (high-μ) implies more distortion-reducing negative feedback. Further, we would want to minimise distortion before feedback, and that implies a low distortion valve. We know that we need a low output resistance, so that means we also need high g_m. There is only one valve that meets all of these requirements and it is the 6C45Π. We know that distortion is minimised by ensuring that R_L / r_a >50, and this condition is easily satisfied by the cascode constant current sink that we already need to avoid wasting negative HT voltage. Thus, a 6C45Π cathode follower sitting on a cascode CCS is as good as a thermionic follower gets.

We always need the efficiency of the cascode CCS, so an FET source follower should have the same load. The reverse biassed drain/gate depletion layer within a power FET has capacitance that varies with the inverse square root of V_{dg}, potentially causing nonlinearity, so we need to choose a low capacitance FET, such as the 500 V 1.4 A Fairchild FQP1N50. Unfortunately, these (like many other modern FETs) are optimised for switching rather than linearity, and the author soon found that the FQP1N50 has a feature not revealed by the manufacturer's data sheet (see Figure 6.49).

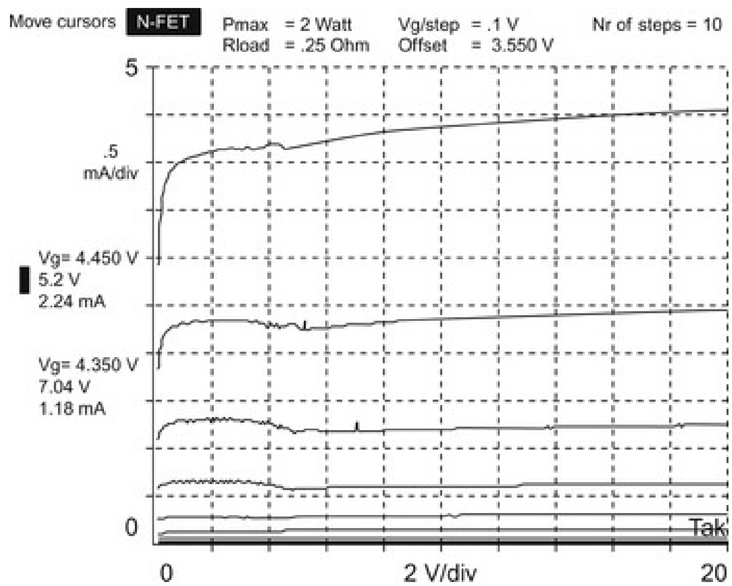


Figure 6.49 FQP1N50P drain characteristics; note the negative resistance tetrode region at $V_{DS} \approx 25V$

As can be seen from the measured curves, there is a tetrode region of negative resistance at $V_{ds} \approx 5$ V, and subsequent distortion measurements showed that V_{ds} should not be allowed to drop below 30 V. Nevertheless, this is still less onerous than the 150 V required by the 6C45П. It's just a shame that the ninth FQP1N50 to be tested provoked smoke from the author's Tek 571 curve tracer (again).

Whether the follower is a cathode follower or source follower, it still needs a quiescent current of 10 mA, and possibly more in order to be able to supply the current needed by the input capacitance of the output stage. Both devices will oscillate given half a chance, so power supply decoupling is important and carbon grid or gate stopper resistors are essential. Further, we know that the followers are likely to be driven from a high source resistance, and that this could cause distortion due to grid current in the 6C45П or due to signal-dependent depletion region capacitance in the FQP1N50, so a test source resistance of 100 kΩ was chosen to reveal either of these problems. Thus, the test circuit for the Device Under Test (DUT) comparison can be drawn (see Figure 6.50).

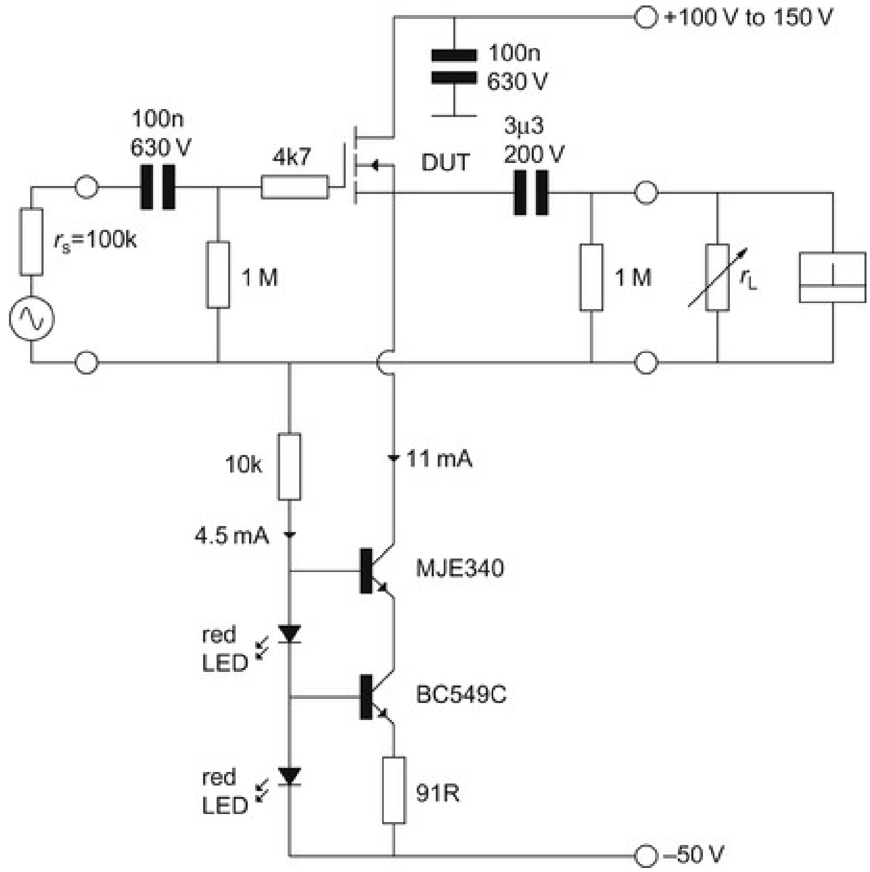


Figure 6.50 6C45Π vs FQP1N50 follower test circuit.

The only circuit change when exchanging between DUTs was that the triode was allowed +150 V, whereas the FET was fine with +100 V. Each follower was driven with +23 dBu (31 $V_{pk–pk}$) because this was the maximum single-ended voltage that the author's audio test set could provide. 1 kHz distortion was indistinguishable from the test set's own distortion, so measurements were made at 10 kHz in the hope of revealing distortion due to signal-dependent depletion region capacitance. With only the 60 kΩ of the test set to load the followers, both achieved <0.01% THD, dominated by second harmonic.

In an effort to provoke some clearly measurable distortion, each follower was loaded by a resistance box adjusted to increase distortion in 3 dB steps (see Figure 6.51).

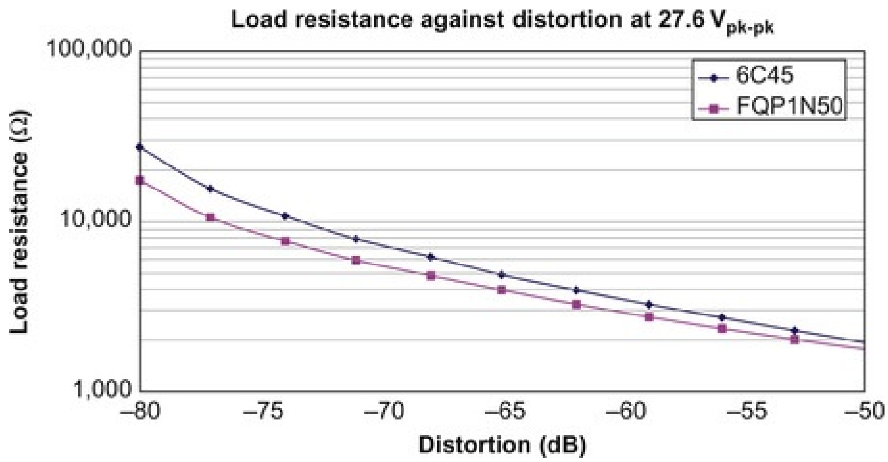


Figure 6.51 Comparision of 6C45Π vs FQP1N50 distortion against load resistance.

We can finally see a difference between the two followers, and the FET is superior. This series of measurements shows that if appreciable load current is needed, the most important property of the follower device is high g_m, and the FET has higher g_m than the valve. Thus, the Bulwer-Lytton amplifier uses FET source followers because they allow lower distortion.

Output Stage Bias, Balance and Coupling

Just like the Crystal Palace, we must make provision for adjusting output stage quiescent current and balancing the anode currents in the output transformer. The required harmonic equaliser resistor for each pair of 6S4As was determined empirically as being 91 Ω. At the optimum current of 45 mA per pair of 6S4As, this shared cathode resistor drops 4.3 V. But the 6S4A needs $V_{gk}=-13$ V to set 45 mA per pair from the 320-V HT, so if −4.3 V is provided by cathode bias, the remaining −6.7 V must be provided by grid bias. Thus, the Bulwer-Lytton amplifier is an example of mixed bias and is more stable than the fixed bias Crystal Palace.

With only 9 V_{RMS} needed at each output valve's grid, the Bulwer-Lytton needs very little gain, and two differential pairs are unnecessary. This implies that a single differential pair will provide the gain and that the driver circuitry's single Low Frequency time constant must be the coupling capacitors between the differential pair and the FET source followers. Thus, output stage quiescent current and DC balance must be adjusted at the inputs of the FET source followers. The traditional way to do this is to make a pair of potential dividers using the ends of a variable resistor with the wiper connected to a negative voltage. Movement of the wiper adjusts relative attenuations of the two potential dividers so that one voltage rises while the other falls. Quiescent current is adjusted by a variable resistor in series with the wiper that controls the current sunk into the potential dividers and therefore the voltage developed across them (see Figure 6.52).

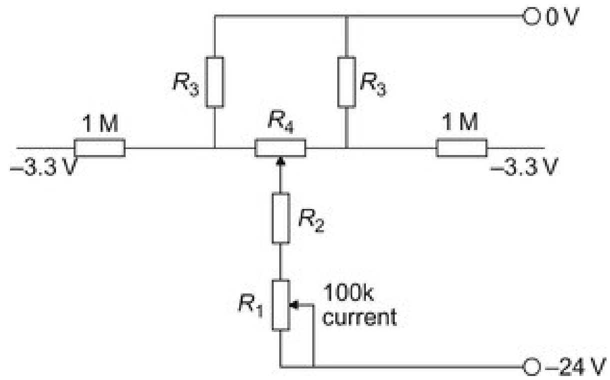


Figure 6.52 Adjusting an output stage's DC balance and bias.

The best way to design such a bias arrangement is with a spreadsheet. The equations are best determined by treating the quiescent current variable resistor R_1 and its associated limiting resistor R_2 as fixed resistors in a Norton current source, then apply the current divider equation to each arm of the bias balance resistor R_4 and its associated limiting resistors R_3 . Once the currents down each arm of the current balance resistor are known, the resistance of each arm of R_4 can be added to its R_3 and Ohm's law is applied to that resistance to determine the output voltage. Finally, copy the equations down for different proportions of R_4 , then plot a graph of the two output voltages against these proportions (see Figure 6.53).

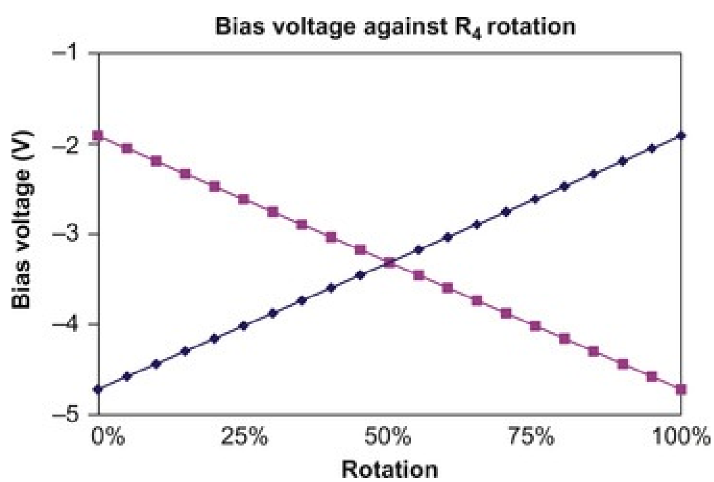


Figure 6.53 Careful adjustment of values allows adequate but not excessive range of bias adjustment.

Having set up the spreadsheet, set R_1 =0 and choose R_2 to give the highest voltage at 50% of R_4 rotation that you think you will need, then adjust R_1 to give the expected quiescent voltage. R_4 and R_3 can then be adjusted to give adequate variation. The process tends to be somewhat iterative, but once the equations have been written, experimentation is quick. Finally, adjust R_1 to give the most negative voltage you think you will need, and this will be the value of this variable resistor.

Although the previous scheme works well at DC, the resistances seen looking into the circuit from the coupling capacitors become unequal once the bias balance control is adjusted from its centre position, and unequal load resistances disturb differential pair AC balance. Furthermore, the time constants of the two coupling capacitors and load resistances become unequal, causing imbalance and consequent distortion towards their cut-off frequency.

The problem can be solved by buffering the DC from the ends of the bias balance control with emitter followers so that the output resistance becomes a few tens of Ohms. Once we add our 1 MΩ gate-leak resistor, any variability or nonlinearity in the emitter follower's output resistance becomes inconsequential and a constant load resistance is seen irrespective of bias balance adjustment. The emitter followers drop 0.6 V across their base–emitter junctions, so it may be necessary to take account of that and revisit the spreadsheet. In addition, it is now possible to include some power supply smoothing to the circuit by putting a capacitor across each R_2 . However, it is not a good idea to make these capacitors too large as they will slug the speed of adjustment, which is very irritating. A time constant of 100 ms allows some smoothing without noticeably slugging adjustment speed (see Figure 6.54).

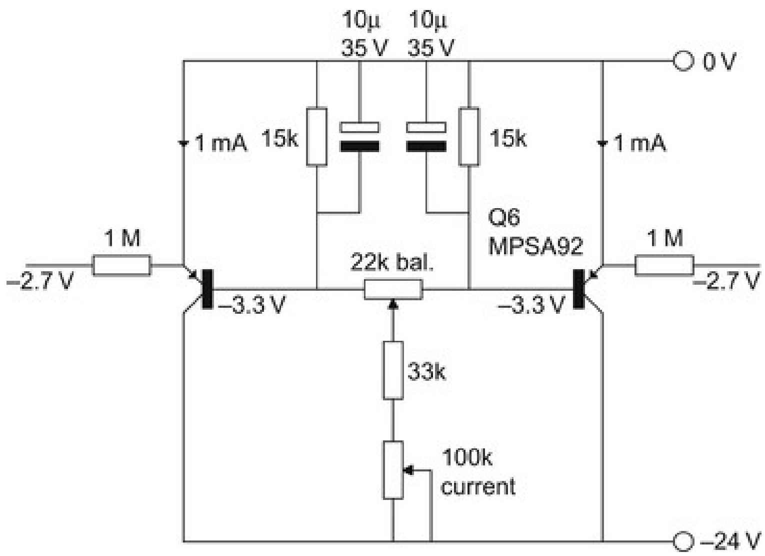


Figure 6.54 Buffering the bias adjustment prevents interaction with AC conditions.

The FQP1N50 has its drain connected to its tab, and as the drain is connected to the +39 V rail, there is no problem in heatsinking it to the chassis via an insulating washer. However, the MJE340 transistors in the constant current sink have their collector connected to their tab, so heatsinking these transistors to the chassis via an insulating washer would add ≈8 pF in parallel with their output, which is not ideal. Thus, the MJE340s have their own heatsinks.

Since the amplifier is designed to drive a dedicated loudspeaker having a known $f_{-3\text{ dB}}=83$ Hz, there seems little point in saturating the output transformer by applying full amplitude bass down to 10 Hz, so the PTFE coupling capacitors are only 10 nF and feed 1 M gate-leak resistors, setting the amplifier's $f_{-3\text{ dB}}=16$ Hz.

Providing Gain

Having designed an output stage that doesn't generate higher harmonics and drivers with low distortion FET source followers, it would be unfortunate to spoil this performance with poor gain circuitry. Thus, although we might consider 0.6% distortion in the output stage to be acceptable on its own, it must be driven by a perfect undistorted signal. In this context, 'perfect' would mean that any driver distortion should be >20 dB below that of the output stage and be composed solely of low orders. Although the differential pair can produce low distortion, it does so by cancelling even harmonics and summing odd, so the triode's dominant H2 tends to cancel but the lesser H3 is summed. We therefore need a valve that inherently produces very small amounts of H3, and that takes us straight back to the *SN7*/N7 family chosen for the Crystal Palace.

Each output valve is biassed so that $V_{gk}=-13$ V, so it requires a swing of ≈9 V_{RMS} to drive it into grid current ($V_{gk}=0$ V). The source followers have very nearly unity gain, but not all 6S4As are the same, so we ought to assume that each anode of the differential pair must swing ≈10 V_{RMS}. Experience suggests that a *SN7/*N7 differential pair should be capable of producing H2≈−69 dB (0.035%) and H3≈−80 dB at this level – which is good enough. The 7N7 was biassed to <8 mA in order to allow a flatter loadline that would allow low distortion and good swing from the limited HT voltage.

Gain Stage CCS and Gain Balance

The gain stage uses the author's standard cascode CCS but it is supplied from a 337L (low power TO92 variant) regulator to guarantee low hum.

Because the test amplifier was designed to operate without global negative feedback, it is entirely possible that component variation (especially valves) could cause an inter-channel gain error. A simple method of compensation is to allow adjustment of DC tail current in each differential pair by adjusting the CCS programming resistor; adjusting tail current changes r_a and thus slightly changes gain. Only a little adjustment can be obtained in this way, but if both stereo channels have the adjustment it becomes possible to match them.

Balanced Inputs on Power Amplifiers

Unlike the signal leaving a microphone or a moving-coil cartridge, signals can be transferred easily to power amplifiers with negligible risk of interference, yet power amplifiers with balanced inputs are becoming increasingly fashionable. Popular reviews comparing balanced and unbalanced inputs on the same amplifier have often claimed that the balanced input sounded better, yet it is hard to see how a domestic length of interconnecting cable could pick up sufficient interference to make balanced operation desirable. However, there is another (and much more likely) possibility, and that is interference *within* the power amplifier.

Fundamentally, there are two ways of achieving a balanced input: either with a transformer, or with a differential pair. Transformers having good winding balance, wide frequency response, and low distortion tend to be expensive, so the differential pair is understandably more popular.

Consider the differential pair driven from a balanced source. Two equal amplitude signals of opposing polarity are applied to the grids, resulting in perfect balance within the differential pair, and *no audio* on the cathodes.

Now consider the differential pair driven from an unbalanced source. One input grid is grounded, and the other has a signal, but the differential pair maintains output balance because a large cathode resistance (whether explicit or a CCS) prevents audio current out of one anode from flowing anywhere but into the other anode. If one grid is grounded but the anode of that valve carries a signal, that valve must be acting as a grounded grid stage and the input signal must be at its cathode.

However, there is capacitance between the cathodes and the heater ($C_{hk} \approx 7$ pF per triode for 7N7), and the heater transformer could easily have 1 nF capacitance to the mains winding, picking up common-mode interference. Thus, by unbalancing the inputs of the differential pair, we force it to amplify signals on its cathodes, rendering it sensitive to mains interference coupled through heater transformer inter-winding capacitance – and that is why the balanced input sounds better, not because of better rejection of interference on audio cables.

Note that the previous argument hinges on a very particular definition of 'balanced.' The overriding requirement of balanced audio is that impedances to earth are equal from each leg – no mention of audio voltages, yet this differential pair 'balanced' input also requires equal signals of opposing polarity.

Standalone digital to analogue converters have once again become popular because of their versatility in dealing with disparate digital sources (computer music server, broadcast receiver, conventional CD player). Many digital to analogue converters improve their distortion/noise by internally operating a pair of DACs in push–pull because taking the difference between these two signals increases the (correlated) signal by 6 dB, but the sum of the two DACs' distortion/noise (which is assumed to be uncorrelated) only rises by 3 dB, thus improving dynamic range by 3 dB over a single DAC. Thus, many modern DACs inherently produce balanced audio, and it is cheap to make a feature of it by bringing it out onto an XLR connector.

Given that domestic balanced audio is now readily available and that a push–pull amplifier with a balanced-capable input works better with a balanced signal, it now makes a great deal of sense for a valve amplifier to expect balanced line-level signals – but not because of the traditional cable interference issues.

The Volume Control and Baffle Step Compensation

Having decided that the Bulwer-Lytton amplifier will be driven primarily from a balanced output DAC, we now know that the required input sensitivity is 4 V_{RMS} (differential). This doubling of voltage occurs because balanced output

DACs invariably produce 2 V_{RMS} on each leg, so if the other leg is inverted polarity, the differential signal must be 4 V_{RMS}. There's no engineering necessity to choose this convention – it is just the way it seems to be done.

The 6S4As need ≈10 V_{RMS} each, equivalent to ≈20 V_{RMS} (differential), and the gain of a 7N7 differential pair is ≈14, resulting in an input sensitivity at the grids of 1.4 V_{RMS} (differential). We therefore need ≈9 dB attenuation from the 4 V_{RMS} source, which is just as well because the Arpeggio loudspeaker needs 2.4 dB baffle step compensation at the input of its dedicated amplifier (causing 2.4 dB attenuation), and balanced volume controls have to be Type C (see Chapter 7) that cannot achieve 0 dB attenuation. Type C volume controls are always a delicate balance between excessive output resistance and low (and variable) input resistance loading the source. Since it has been assumed that the Bulwer-Lytton amplifier will primarily be driven from a DAC having low output resistance, the balance can be swung towards low input resistance, enabling lower output resistance. The issue of low output resistance is important because the volume control precedes the baffle step equaliser, so minimising output resistance minimises the (unavoidable) changes in equalisation as the volume control is adjusted (see Figure 6.55).

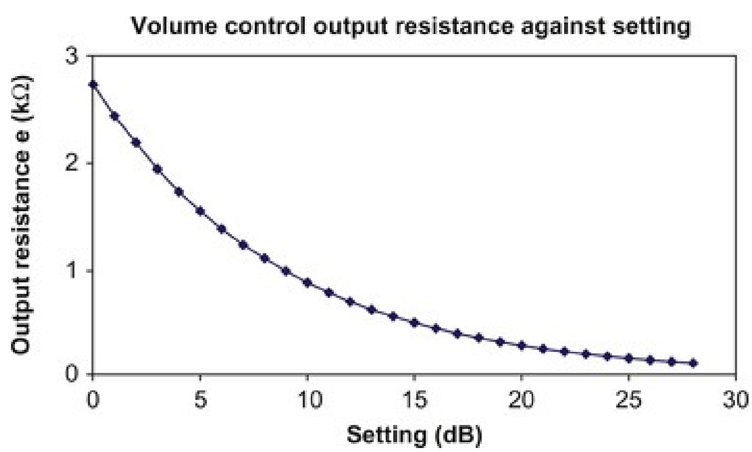


Figure 6.55 Bulwer-Lytton volume control output resistance against attenuation.

The volume control's output resistance changes from 2.73 kΩ to 109 Ω, and this changing source resistance added to the equaliser's series resistors changes its attenuation by 0.17 dB, manifesting itself as a shelf error. Fortunately, although shelf errors are the most audible, discriminating between a 0.2 dB error and no error at a constant level is marginal, so a 0.17 dB error between full and minimum volume should pass unnoticed.

Type C attenuators habitually select an individual resistor for their shunt but it is much easier to obtain a good logarithmic law if the wiper moves up and down a series chain of resistors so that it picks up the sum of a number of series resistors. Using this technique, the Bulwer-Lytton's volume control deviates from logarithmic linearity by <0.05 dB and has perfect channel matching (Table 6.8).

Table 6.8 Bulwer-Lytton Volume Control Resistor Values

Attenuation (dB)	Individual resistor
0	–
1	2k7
2	1k8
3	1k5
4	1k1
5	820 Ω
6	680 Ω
7	560 Ω
8	430 Ω
9	390 Ω
10	330 Ω
11	270 Ω
12	240 Ω
13	200 Ω
14	160 Ω
15	150 Ω
16	130 Ω
17	120 Ω
18	91 Ω
19	91 Ω
20	75 Ω
21	68 Ω
22	56 Ω
23	51 Ω
24	47 Ω
25	39 Ω
26	36 Ω
27	33 Ω
28	27 Ω
Tail resistor	221 Ω 0.1%+2 Ω 1%

We can now draw the entire circuit diagram of the Bulwer-Lytton amplifier (see Figure 6.56).

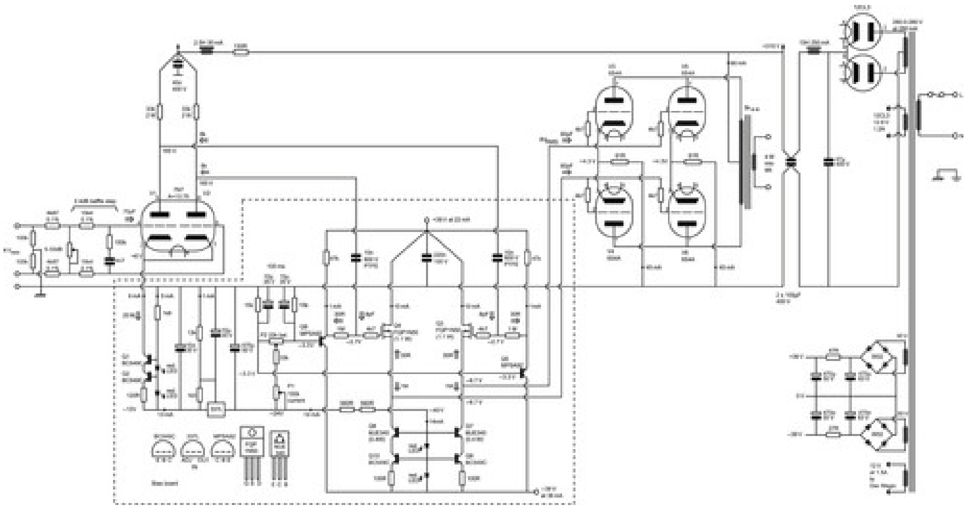


Figure 6.56 Entire Bulwer-Lytton circuit diagram.

Audio Circuit Comments

The circuit appears complex, but most of it is housekeeping circuitry to enable the simple audio department of differential pair and source followers to work at their best. Although the amplifier as a whole was designed to complement the author's Arpeggio loudspeaker, the housekeeping circuitry was designed to be generic. Each channel's entire housekeeping circuitry was built on a single PCB; just add a chassis-mounting valve socket to suit the gain valve and you have a complete driver. Further, the FET source follower is specifically designed to be able to cope with the higher voltages needed by valves such as the KT88, hence the 300 V MJE340 upper transistor in the CCS and the 500 V FET. Obviously, detailed resistor values would need to be changed and the MJE340 would probably need bigger heatsinks because the higher voltages would increase dissipation, but the board layout would remain unchanged.

Power Supplies

Each pair of 6S4As needs 45 mA, and there are four pairs, so that means 180 mA, and each differential pair requires 8 mA, giving a total requirement of 196 mA at 315 V_{DC}. Traditionally, a GZ34 would have been the obvious HT rectifier but these are now expensive, so a pair of 12CL3/12CK3 was chosen, necessitating a 12.6 V 1.2 A heater supply. The FET source followers and housekeeping circuitry require +39 V_{DC} at 44 mA and −39 V_{DC} at 72 mA. The author detests 'wall-warts,' so rather than leaving an ugly plug-mounted mains transformer permanently powered (getting hot and wasting electricity), a

12 V 1.5 A winding was added to power his DacMagic audio DAC.

A CLC filter was ideal for HT to the output stages, but the gain stage needed as much HT as possible to minimise distortion, so it also used LC smoothing. As is now the norm, PSUD2 was used to analyse the supplies and check for low frequency stability – which revealed that the inductor feeding the gain stage had too high a *Q* and needed a small series resistor to prevent Low Frequency ringing. Similarly, two 100 μF 400 V Kelvin capacitors were needed to smooth the HT to the output stages not for any reasons of stereo separation (since they are in parallel) but because 200 μF was needed to prevent Low Frequency ringing.

All of the preceding supplies can be derived from a common transformer, but the audio valve heaters need their 6.3 V at 6 A from a separate transformer (ideally a split bobbin EI type) to avoid interference from rectification spikes. Unfortunately, the author didn't have room on the top of his chassis for a canned EI heater transformer, so he was forced to use a 6 V 50 VA toroid under the chassis that gave 6.2 V because it was slightly under-run (under-running a 6 V transformer to obtain 6.3 V is a useful trick).

Global Negative Feedback

We saw in Chapter 3 that Baxandall's analysis of feedback showed that provided open-loop distortion is <1% and dominated by second harmonic, *any* amount of negative feedback from 0 dB upwards proportionately reduces the amplitude of all harmonics and self-generation of higher harmonics is negligible. On test, the 6S4A push–pull output stage produced 0.6% THD just below grid current, dominated by second harmonic, so this amplifier is an ideal candidate for global negative feedback. In the Bulwer-Lytton amplifier, global feedback would have been undesirable because it would have reduced the output resistance required by the Arpeggio loudspeaker, but a more forceful version employing sixteen 6S4As per channel (BL) [(2)] to give 16 W could drive a more conventional loudspeaker and 20 dB of negative feedback would reduce its maximum distortion to 0.06% – still dominated by second harmonic.

The task of a power amplifier is to amplify a processed signal and deliver *power* into a load such as a loudspeaker. It should do this without introducing spurious signals, such as hum, noise, oscillation or audible distortion, whilst driving a wide range of loads. Additionally, it should be tolerant of abuse, such as open or short circuits. It will be appreciated that this is not a trivial objective, and will therefore require careful design and execution if it is to be achieved.

[1] Hafler, D; Keroes, HI, An ultra-linear amplifier , *Audio Eng* (November) (1951) 15 – 17 .

[2] Blumlein A. Improvements in/or relating to thermionic valve amplifying circuits. British Patent 496,883; 1937.

[3] Molloy E, editor. High fidelity sound reproduction. 2nd ed. George Newnes; 1958. p. 52.

[4] McIntosh, FH; Gow, GJ, Description and analysis of a new 50 W amplifier circuit , *Audio Eng* (December) (1949) .

[5] Futterman, J, A practical commercial output transformerless amplifier , *J Audio Eng Soc* (October) (1956) ; [also US Patent 2,773,136; December 1959] .

[6] Schmitt, OH, Cathode phase inversion , *J Sci Inst* **15** (136) (1938) 100 – 101 .

[7] Clare, JD, The twin triode phase-splitting amplifier , *Electron Eng* (February) (1947) .

[8] Carpenter REH. Improvements relating to electron discharge apparatus. UK Patent 325,833; 1928.

[9] Masatsugo Kobayashi. Electric oscillator. US Patent 1,913,449; 1929.

[10] The Williamson Amplifier. A collection of articles reprinted from 'Wireless World' April 1947, May 1947, August, October and November 1949, December 1949, January 1950, May 1952.

[11] Hafler, D; Keroes, HI, Ultra-linear operation of the Williamson amplifier , *Audio Eng* (June) (1952) .

[12] Mullard: Tube circuits for audio amplifiers; 1959 [now reprinted by Audio Amateur Press, Peterborough, New Hampshire, USA; 1993].

[13] Walker P. Private communication; 1993.

[14] Williamson, DTN; Walker, PJ, Amplifiers and superlatives , *J Audio Eng Soc* **2** (April(2)) (1954) .

[15] An approach to audio frequency amplifier design. General Electric Company Ltd; 1957 [now reprinted by Audio Amateur Press, Peterborough, New Hampshire, USA; 1994].

[16] Colloms, M, *High performance loudspeakers* . 6th ed. (2005)Wiley ; pp. 327–43 .

[17] Hedge LB. Cascade AF amplifier. Wireless World; June 1956.

[18] Langford-Smith, F, *Radio designer's handbook* . 4th ed. (1957)Iliffe .

[19] Philips QE 5/40 valve data sheet; 12th December 1958.

[20] LM134-LM234-LM334 data sheet. SGS-Thomson; October 1997.

References

Further Reading

- Spicer, S, *Firsts in high fidelity* . (2001)Audio Amateur Press , New Hampshire, US .

Chapter 7. The Pre-Amplifier

The traditional pre-amplifier performed a number of functions:

- Input selection
- Volume control
- Balance control
- Cable driver
- Tone control
- RIAA stage.

If the last two functions are not needed, then the first three can be implemented passively within the power amplifier, negating the need for the extra electronics of a cable driver and making an external pre-amplifier entirely redundant.

Different situations require a different mix of the previous functions, so we will investigate how each function may be implemented, allowing us to mix and match to a particular application. In general, the problem is defined either by the need to put analogue controls conveniently to hand at the listening position or to limit the length of cable carrying a low-level signal. Thus, there are four common situations that require a powered box before the power amplifier:

- *Traditional pre-amplifier with full controls* : Placed adjacent to vinyl turntable, and both placed conveniently to hand.
- *Limited control pre-amplifier* : Input selector, volume control, plus balance control with compensating gain – placed conveniently to hand.
- *Volume control plus unity-gain line driver* : Although this could be an even simpler version of the above, it is more likely to be a six-channel stepped attenuator placed *after* a digital active crossover as a means of achieving stereo volume control whilst maintaining each DAC’s full dynamic range.
- *RIAA stage* : Vinyl cartridges need their amplification nearby.

Input Selection

Although infrequently used, when the input selector *is* used, attention is focussed upon it. Thus, input selection should not offend, and ways it can offend are:

- Disparate levels between sources
- Crosstalk between sources
- Noise.

Analogue switches such as the 4016 (but ideally not the 4066 because of crosstalk from its control lines) can switch audio signals. However, these semiconductor switches are based on Field Effect Transistors (FETs) having 'off' capacitances proportional to the inverse square root of the voltage across their depletion region, so they need virtual earth techniques to minimise distortion, implying semiconductor op-amps. Valve electronics therefore uses mechanical switches, so we will investigate their defects and how to minimise them.

Disparate Levels between Sources

Strictly, there's no switch defect if source levels do not match, but that fine distinction is lost on the non-technical user. It ought to be possible to switch between Bach arriving via analogue radio, remote digits, optical disc player, computer server or LP without hearing any appreciable volume difference between the sources.

Most signal switching is now done in the digital domain, and there are now only a few analogue sources. CD is the oldest source of domestic digital audio, and its 2 V_{RMS} maximum undistorted sine wave definition of analogue level has become the de facto standard. Thus, we should ensure that all sources arrive at the input selector matching this standard. In this instance, broadcast terminology has been adopted across the industry and most quality sources have internal engineering adjustments marked 'line level', so it is these that should be adjusted (if necessary) to correct level.

Quarter-inch tape machines have become fashionable once again, and European ex-broadcast machines typically have their line levels set to reproduce peaks either 8 dB (UK) or 9 dB (Germany) above 0.775 V_{RMS} or 0 dBu. The significance is that +8 dBu=1.95 V_{RMS}, so no adjustment is necessary.

Sadly, FM tuners commonly produce quite small signals, and this is a tuner rather than pre-amplifier problem because we should only send robust signals down cables. If, as is common, there isn't a line-level adjustment, a piggyback line driver board will be needed to provide the additional gain.

Adjacent Contact Capacitance (Crosstalk Between Sources)

The simplest way to select between sources is with a rotary switch. Unfortunately, all such switches suffer adjacent input crosstalk due to capacitance between adjacent contacts (typically 0.4–0.8 pF). If we had a number of sources plugged into the pre-amplifier, but selected an inactive or unplugged source, that ≈0.6 pF would form a high-pass filter in conjunction with the typical 100 kΩ input resistance of the volume control. The ear is most sensitive at ≈4 kHz, and at this frequency the RC combination would cause crosstalk ≈53 dB below an expected signal. On high-quality traditional pre-amplifiers, this irritation was solved by having two ganged switches: one selected the source and the other deselected the short-circuit to ground on that source. Sadly, such wafer switches are no longer available, but a good alternative is to use alternate contacts for inputs and connect intermediate contacts to ground to guard between signal contacts (see Figure 7.1).

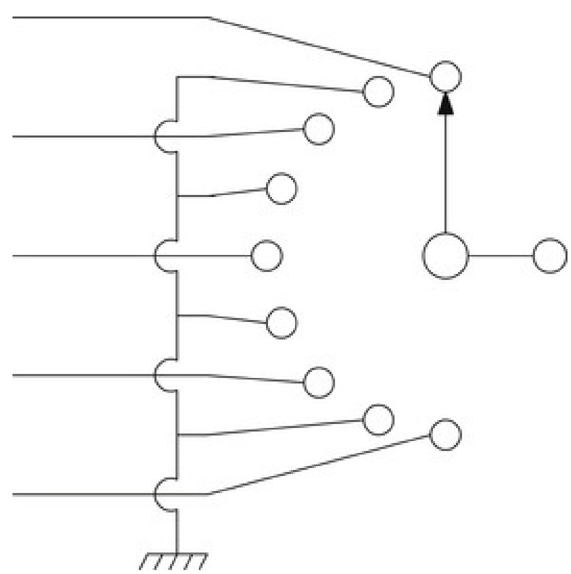


Figure 7.1 Block diagram of pre-amplifier.

Although crosstalk on an open-circuit input could easily be at −53 dB with a traditional switch, once an active source having a low source resistance is selected, the crosstalk falls rapidly. As an example, selecting a source having r_{out} ≈300 Ω would cause the crosstalk to fall to −107 dB. Thus, the grounded intermediate selector switch contacts are mainly there to avoid crosstalk if an unused input is selected, but they also provide a mute setting between sources without adding contacts to the signal path that the author has subsequently found to be very convenient.

Contact and Leakage Resistance (Noise)

Ideally, contact resistance would be zero, but this is never achieved. Contact resistance is not a problem provided that it is low (milliohm) and constant, but causes noise when it rises and particularly if it fluctuates. Contact resistance rises over time as a result of oxidation of the contacts, waning contact pressure and contact wear. Because of its resistance to atmospheric corrosion, gold-plated contacts are sometimes used, despite the fact that the resistivity of gold is considerably higher than that of silver or copper, but beware that gold is softer than silver, so the wiping contacts of a rotary switch wear gold faster than silver.

All switches have leakage resistance, which is usually specified by the manufacturer, and will be slightly worsened by increasing humidity, but surface leakage currents can easily be generated by failing to deflux the wafer after soldering (flux remaining after soldering contains solder droplets and is electrically leaky). Surface leakage currents invariably generate noise and, on a selector switch, crosstalk.

Solutions and Problems Peculiar to Electromechanical Switches (Relays)

Although a changeover relay could be used for input selection, allowing crosstalk due to stray capacitance, practical implementations use one relay (and associated coil) per input. Because the coil is relatively large, the enforced physical separation between source wiring effectively eliminates stray capacitance across the distributed switch – although this could always be restored by spectacularly incompetent Printed Circuit Board (PCB) layout.

The ideal signal switch was the mercury wetted relay because a droplet of mercury wetted the contacts to ensure minimum, and constant, contact resistance. Because the relay had to be hermetically sealed to prevent the (poisonous) mercury vapour from escaping, the contacts did not oxidise. The contacts were non-wiping, so wear was minimised. Unfortunately, they are no longer manufactured, mostly because their main use was in telegraph repeaters (long obsolete), but also because safety legislation concerning substances hazardous to health has made it all but impossible to sell any product incorporating legislated substances no matter how great their benefit or how small the health risk.

Restricted availability of mercury wetted relays notwithstanding, relay switching is an excellent solution. If we had a series relay on each source, we could precede it with a shunt relay to ground that would normally be closed, and this would ensure almost perfect attenuation of unwanted sources. To protect the source from the short-circuit, a 1 kΩ series resistor is normally fitted before the shunt relay, which further improves attenuation (see Figure 7.2).

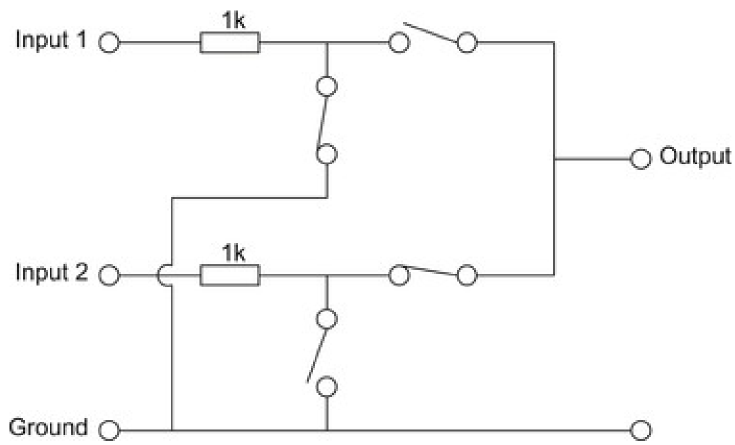


Figure 7.2 Modifying a typical pentode input stage for triode operation to reduce noise.

Each relay should be mounted as close as possible to its associated input socket, allowing the signal wiring to be as direct as possible (minimising stray capacitance), and the front panel selector switch merely carries DC. 5 V relays are an obvious choice, since these could be powered by a 5 V regulator fed from the same transformer winding as the valve heaters. However, whilst this would be a perfectly valid design choice if we used simple DC or combinational logic to drive the relays, it is more likely that we would use a microcontroller and add remote control. The key virtue of combinational logic is that it is quiet as a mouse when not being controlled, whereas the sequential microcontroller must continually poll its inputs to see if anything has changed, so its clock and firmware run continuously, polluting the heater winding and capacitively coupling noise into cathodes of the audio valves. If you use a microcontroller, either filter its supply very carefully or give it a dedicated mains transformer (also enabling remote control of power ON/OFF).

Volume Control

The volume control is the most frequently used control, so it is essential that it does not offend. Ways that it can offend as it is adjusted are:

- Disturbing frequency response
- Perceived volume not changing smoothly with rotation
- Disturbing channel matching.

We will investigate how to minimise each of these problems in turn.

Limitations on the Control's Value (Disturbing Frequency Response)

The maximum output resistance of a 100 kΩ volume control is 25 kΩ. This maximum output resistance may easily be verified by moving the wiper to the *electrical* mid-position of the track. The resistance to each end must be half the total resistance, and assuming zero source resistance, each end is at AC ground. Looking back into the potentiometer, we see the two halves in parallel, and therefore the output resistance is equal to the total resistance of the potentiometer divided by four. If the wiper is at either end of the track, output resistance will be zero because it is connected to ground either directly or via the (zero resistance) source. Maximum output resistance therefore occurs when the wiper is as far away from *each* end as possible, which is the centre position.
The question of potentiometer maximum output resistance is crucial because it forms a low-pass filter in conjunction with loading capacitance whose $f_{-3\ \mathrm{dB}}$ cut-off frequency we can calculate from:

$$f_{-3\ \mathrm{dB}} = \frac{1}{2\pi CR}$$

However, we would like the high frequency roll-off within the audio band to be far less than 3 dB, so we need to be able to relate an $f_{-3\ \mathrm{dB}}$ frequency to a given loss at a given frequency, which we can find from the following formula:

$$f_{-3\ \mathrm{dB}} = \frac{f_{(\mathrm{dB\ limit})}}{\sqrt{(1/10^{\mathrm{dB}/10}) - 1}}$$

where

$f_{(\mathrm{dB\ limit})}$ =the outermost frequency of interest dB=the deviation from flat response at that frequency.

The difference between 0 dB and 0.1 dB roll-off at 20 kHz is negligible, so if we wanted a volume control that caused negligible frequency response errors over its range, we would set our limit to 0.1 dB at 20 kHz, resulting in $f_{-3\ \mathrm{dB}}$ =131 kHz. Thus, when the volume control is at its maximum output resistance of 25 kΩ, we find that its maximum tolerable loading capacitance is ≈49 pF.
Obviously, a lower-resistance volume control tolerates higher loading capacitance but at the expense of loading the source more heavily. Modern valve electronics can tolerate a 100 kΩ load without any problem, but distortion rises as the load resistance falls. Thus, 100 kΩ has become the de facto standard volume control because it is a standard value that offers a usable compromise between high frequency roll-off due to loading capacitance and increased distortion due to a steep loadline.

Logarithmic Law (Perceived Volume Not Changing Smoothly with Rotation)

In common with other human senses, the ear has a logarithmic response to sound pressure level, so if we want a volume control that has a uniform perceived response to adjustment throughout its range, we need a logarithmic potentiometer. This is the root cause of all our problems.
It is not a problem to make a linear potentiometer. We simply deposit a strip of carbon of uniform width and thickness onto an insulator, put terminals at each end, and arrange for a contact to scrape its way along. If we don't bother with a casing, it is known as a skeletal type. In an attempt to produce a logarithmic law, the coating thickness is made variable, then, in deference to audio sensibilities, a pressed metal screening can is fitted, and two potentiometers are ganged together on one shaft onto which we can fit a big, shiny, spun aluminium knob. Making the coating thickness continuously variable would be expensive, so the logarithmic law is approximated by a series of straight lines (see Figure 7.3).

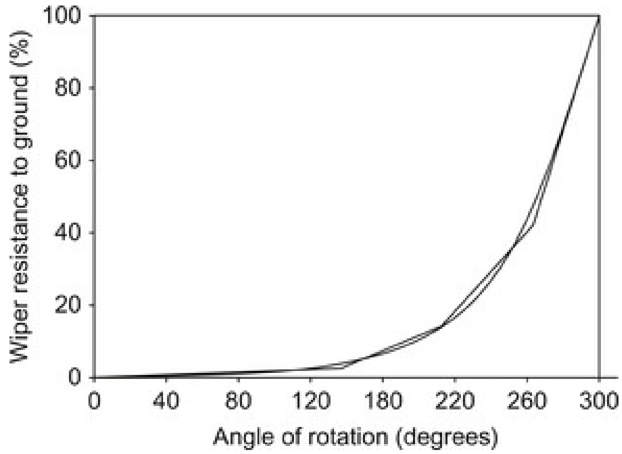


Figure 7.3 Approximation of logarithmic law by straight lines.

It is amazing how good a fit to the ideal logarithmic curve can be made using only four different resistance tracks, but it will come as no surprise to learn that this still results in steps in the response as the knob is rotated. Worse, when we mechanically gang them we expect the channels to match all the way from 0 dB to 60 dB and that just isn't going to happen. Some are quite good, but ganged carbon track logarithmic volume controls belong in landfill.

Switched Attenuators (Disturbing Channel Matching)

If quality is paramount, and we can accept a control that is not continuously variable, we could use a switched attenuator that works by selecting fixed resistors in order to control volume. The logarithmic law can now be perfect, as can the much more important channel matching. However, just because a volume control has detents, this *does not* guarantee that it is a true switched attenuator – it could be a carbon track potentiometer in masquerade. Real switched attenuators tend to be quite large. A quick test, rather than dismantling it in the shop, which might have you thrown out, is to measure the resistance of the lower arm of each gang at the maximum attenuation setting with a digital multimeter. If there is any measurable difference between gangs, it is likely to be a carbon track potentiometer.

The switched attenuator has a long and noble history. The BBC used quadrant faders (switched attenuators without detents) on sound desks until the 1970s because there is nothing worse than gently fading out a programme and hearing an abrupt change in the rate of attenuation – the ear expects to hear an exponential decay akin to reverberation decay. Once Penny & Giles slider attenuators using conductive plastic tapped linear tracks became available, the (far more expensive) quadrant fader could be safely relegated to broadcasting history.

Surprisingly, switched attenuators with a mere nine steps were made by Erie for consumer use in 1949 [1], but even then the advantage of the superior law was realised.

Commercial switched attenuators having thick film resistances inked directly onto the ceramic substrate of the switch wafer are available (at a price), and their performance is far better than a carbon track volume control, although not quite as good as you might expect owing to the poor tolerance of the inked resistors. Nevertheless, an Alps stepped attenuator guaranteed its step error as <±0.5 dB and its far more important channel matching error as <0.5 dB, implying 10% tolerance inked resistors. More recent commercial solutions solder a PCB populated with surface mount resistors directly to the switch wafer and the ready availability of closer tolerance resistors potentially allows much better channel matching errors to be legitimately claimed.

The practical disadvantage of the switched attenuator is that we can only have as many different volume levels as switch positions. Although common rotary switches have detents at 30° (11 usable positions), Elma makes a popular miniature 15° (23 usable positions) stud switch, UK Type 72 stud switches are 12° (29 usable positions), and the author recently bought some very splendid Soviet dual wafer 10° (35 usable positions) stud switches that looked entirely appropriate for detonating explosives yet had a light action that belied their size. Shallco's E and F series of rotary switches offer a range of detent angles down to 6°, potentially allowing excellent resolution *and* wide range.

The designers of the 15° detent Alps-stepped attenuator assumed that we needed a −60 dB position, then provided coarse steps followed by 2 dB uniform steps up to 0 dB. The 60 dB attenuation range allowed for switching between sources having disparate levels. Thus, the previous requirement for all sources to reach the input selector control at the same level was not only ergonomic, but also reduced the range needed by the volume control, making switched attenuator volume controls much more practical. 2 dB steps are a little too coarse, and the author prefers 1 dB. Provided that levels are matched at the input selector and correctly matched to power amplifier sensitivity (only just able to overload the amplifier on the last two or three steps), 1-dB steps on a switch having 12° detents (or better) allow a perfectly practical volume control.

Switched Attenuator Design

Assuming that we have a suitable switch for the attenuator, we need to calculate the required resistor values. We could do this by hand, but a computer makes life much easier. There are three fundamental attenuator types (see Figure 7.4).

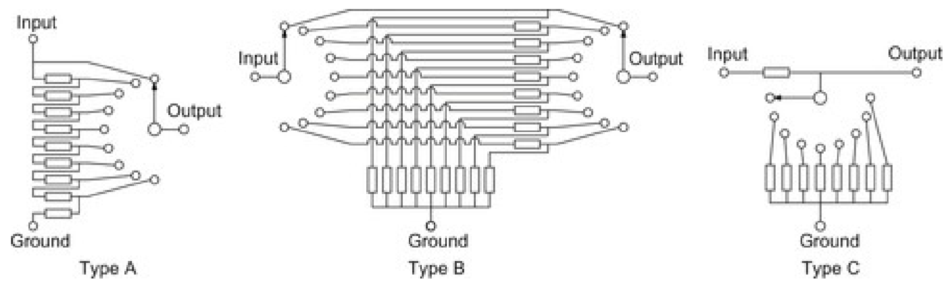


Figure 7.4 μ -Follower as pre-amplifier output stage.

The Type A attenuator in Figure 7.4 is similar to the carbon track attenuator in that it has a chain of resistors from which we choose an appropriate tap. The Type A attenuator has a good logarithmic law and excellent channel matching because for most attenuations each potential divider resistor is made up of a number of resistors, improving the tolerance of each composite resistor. As a result, even 2% tolerance resistors allow an average channel matching error of <0.1 dB provided that the tail resistor at the bottom of the ladder is 0.1% tolerance.

The Type B attenuator in Figure 7.4 uses individual potential dividers for each volume setting, which dramatically reduces the number of soldered joints and components in the signal path at the expense of twice as many wafers and resistors. Unfortunately, maintaining constant input resistance yet providing the exact attenuations required to give a good logarithmic law, requires awkward values and therefore E96 resistors. Further, because each attenuator is made of only two resistors, there is none of the averaging effect that reduced errors in the Type A attenuator, so the resistors must be 1% tolerance (preferably 0.1%) to minimise deviations from logarithmic law and achieve close channel matching. Thus, the Type B attenuator is unpopular because it costs more than double compared to the Type A, yet has worse logarithmic law and channel matching solely to reduce the number of components and joints in the signal path.

The Type C attenuator in Figure 7.4 uses a single fixed series resistor and a selection of shunt resistors in order to achieve the same signal path component count as the Type B but at the cost of the Type A. However, input resistance is no longer constant, and the series resistor must be equal to the maximum tolerable output resistance because when this attenuator is set to maximum attenuation, its input resistance is equal to that of the series resistor. Thus, input resistance falls to a minimum of 25 kΩ, whereas the Type A and B attenuators had a constant input resistance of 100 kΩ and a maximum output resistance of

25 kΩ. Provided stray capacitances are minimised and the attenuator only drives a low input capacitance cathode follower, the series resistor may be increased to 100 kΩ to tame its input loading. Despite its loading issues, the Type C attenuator is far more popular than the Type B because it is so much cheaper.

The following QBASIC programs generate the resistor values for the attenuators in Figure 7.4 . They are not miracles of programming, but they are quick and easy to use, and can easily be modified for other programming languages, or the key equations can be extracted and used in a spreadsheet.

The programs ask for the load resistance across the wiper; this is the grid-leak resistor of the following valve. It is tempting to try to use the potentiometer as the grid-leak, but this is poor practice and can cause noise problems when the contacts bounce, and it is also unnecessary, since the programs account for its loading in designing the attenuator.

This program finds individual resistor values for the Type A attenuator. The final value given by this program is connected between the last usable switch contact and ground, and it is often convenient to use the spare contact on the switch as a ground terminal. The tail resistor is invariably an awkward value, yet it must be 0.1% tolerance, but it is perfectly permissible to make it up from the series combination of a 0.1% tolerance resistor and a 1% tolerance resistor provided that the 0.1% tolerance component is more than 10 times the value of the 1% component.

```
CLS
A=0
B=0
N=0
PRINT "This program calculates individual values of resistors between"
PRINT "taps of the Type A attenuator."
PRINT "How many switch positions can you use";
INPUT S
PRINT "What step size (dB)";
INPUT D
PRINT "What value of resistance will be across the output of the potentiometer";
INPUT L
PRINT "What value of potentiometer is required";
INPUT R
DO UNTIL N=S - 1
Y=((R - L/10 ^ (-A/20))+SQR((L/10 ^ (-A/20) - R) ^ 2+4 R L))/2
C=R - Y - B
PRINT A; "dB "; C; "ohms"
B=B+C
```

```
A=A+D
N=N+1
LOOP
PRINT A; "dB "; R - B; "ohms."
```

The following program is for the Type B attenuator.

```
CLS
A=0
N=0
PRINT "This program calculates upper (X) and lower (Y) arms of"
PRINT "individual potential dividers for the Type B attenuator"
PRINT "How many switch positions can you use";
INPUT S
PRINT "What step size (dB)";
INPUT D
PRINT "What value of resistance will be across the output of the potentiometer";
INPUT L
PRINT "What value of potentiometer is required";
INPUT R
DO UNTIL N=S
Y=((R - L 10 ^ (-A 20))+SQR((L 10 ^ (-A 20) - R) ^ 2+4 R L)) / 2
X=R - Y
PRINT A; "dB "; "Y ="; Y; "ohms "; "X ="; X; "ohms"
A=A+D
N=N+1
LOOP
```

The final program calculates shunt resistors for the Type C attenuator. Note that it *never* achieves zero attenuation, and therefore the program also predicts the minimum unavoidable loss through the volume control and grid-leak resistor (basic loss). In effect, this volume control must be considered to be a fixed attenuator plus a variable attenuator.

```
CLS
N=0
PRINT "This program calculates shunt resistors for the Type C attenuator."
PRINT "How many switch positions can you use";
INPUT S
PRINT "What step size (dB)";
INPUT D
PRINT "What value of resistance will be across the output of the potentiometer";
INPUT L
PRINT "What value of series resistor is required";
INPUT R
B=((-100 LOG(L/R+L)) 8.686) \ 1)/100
REM THE 8.686 FACTOR ARISES BECAUSE QBASIC USES NATURAL LOGS
```

```
PRINT "Basic loss ="; B; "dB, added shunt is infinite"
PRINT "Added attenuation:"
A=B
DO UNTIL N=S - 1
A=A+D
C=R * 10 ^ (-A/20))/(1 - 10 ^ (-A/20))
Y=1/(1/C - 1 / L)
N=N+1
PRINT N * D; "dB, shunt="; Y; "ohms"
LOOP
```

Spreadsheets and Volume Controls

As mentioned, the fundamental equations from the previous QBASIC programs can be used in a spreadsheet. Although harder to debug, the advantage of a spreadsheet is that it can be set up to predict exact values, substitute nearest standard E24 values, and then plot the consequent design errors on a graph.

The most common requirement is for a Type A attenuator loaded by a 1 MΩ grid-leak resistor that emulates a perfect 100 kΩ logarithmic potentiometer. Fortuitously, once practical E24 values were substituted for the calculated values, this combination produced quite low design errors (10 other combinations were tried but all were worse), and some tweaking away from obvious values improved it further – hence the resistor change to 910 Ω from 984 Ω (chain) and to 3k92 from 3,995 Ω (tail). Alternatively, we could use E96 resistors, rendering the design errors entirely negligible (see Figure 7.5).

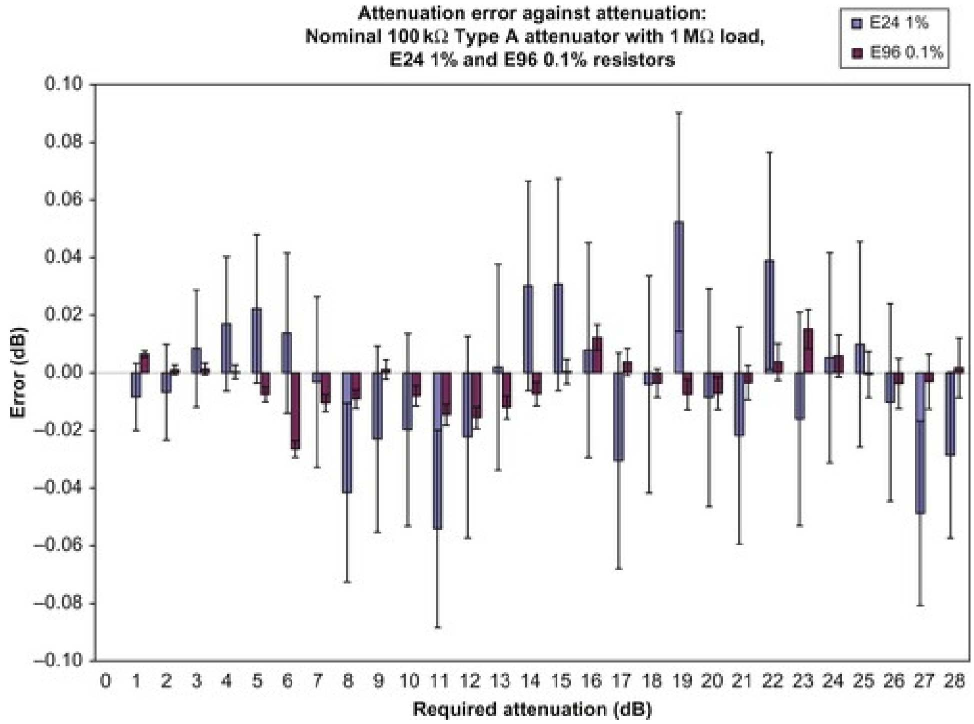


Figure 7.5 Attenuation error against attenuation for E24 and E96 Type A attenuator.

The error bars associated with each design step error in the graph assume the use of E24 1% or E96 0.1% tolerance chain resistors and a 0.1% tolerance tail resistor, resulting in an entirely negligible maximum channel matching error of 0.053 dB and an average channel matching error of 0.045 dB even for the E24 1% option.

The reason for including the E96 0.1% option is not to further improve the already entirely adequate channel matching but because E24 1% tolerance resistors tend to be thick film, whereas E96 0.1% tolerance resistors are thin film. The significance is that the distortion of surface mount thin-film resistors is better than −100 dB for the values needed (provided that 0805 or 1206 case size are used), whereas thick-film resistors are somewhat worse [2]. If surface mount resistors are used, they should be the 1206 case size to minimise distortion, irrespective of whether they are thick or thin film.

Note that the E24 values in the table 7.1 are subtly different to those given in the third edition of this book, enabling slightly lower design errors (maximum 0.053 dB) and a (more convenient) single-tail resistor.

Table 7.1 Resistor Values for 100 kΩ Type A Attenuator

Loss (dB)	*R* (ideal)	E24 1%	E96 0.1%
0	0	0	0
1	10068	10 k	10 k
2	9261	9k1	9k31
3	8456	8k2	8k45

4	7675	7k5	7k68
5	6932	6k8	6k98
6	6237	6k2	6k34
7	5594	5k6	5k49
8	5005	5k1	4k99
9	4470	4k3	4k42
10	3987	3k9	4k02
11	3553	3k6	3k57
12	3164	3k0	3k16
13	2816	2k7	2k80
14	2506	2k4	2k49
15	2229	2k2	2k21
16	1983	2k0	1k96
17	1764	1k8	1k78
18	1569	1k5	1k58
19	1396	1k3	1k40
20	1242	1k3	1k24
21	1105	1k1	1k1
22	984	910 Ω	976 Ω
23	875	910 Ω	866 Ω
24	779	750 Ω	787 Ω
25	694	680 Ω	698 Ω
26	618	620 Ω	619 Ω
27	550	560 Ω	549 Ω
28	490	470 Ω	487 Ω
Tail	3996	3k92 (0.1%)	3k92 (0.1%)+75 R

Volume Controls for Digital Active Crossovers

Versatile digital active crossovers intended for the professional sound reinforcement market are deservedly becoming popular for domestic Hi-Fi loudspeakers because economies of scale mean that they are surprisingly affordable. However, to achieve that price, the DACs and analogue electronics in such crossovers tend to be good rather than excellent quality, with no dynamic range to spare, so the Hi-Fi solution to retaining their full dynamic range at all volume settings is to place the volume control after the crossover.

Although driver deviations with frequency are far greater, inter-driver level errors of 0.2 dB are just perceptible when setting up an active crossover loudspeaker because the ear is sensitive to the averaged response and notices if there is a shelf in the relative response between adjacent drivers. The significance of a 0.2 dB error being just perceptible is that if the volume control is to be placed after an active crossover, its channel matching errors must be <0.1 dB, implying that a Type A-stepped attenuator using 2% tolerance resistors is permissible, but 1% preferable.

Sadly, a Type B attenuator would be unsuitable unless all resistors were 0.1% or better, substantially increasing cost.

Volume Control Values and Their Effect on Noise

A volume control using resistance to achieve attenuation must produce noise. A resistive volume control's noise can be calculated easily by assuming that it is driven from a source of zero resistance. This assumption folds the top arm of the potential divider over, and the (zero resistance) signal source can be removed, leaving two resistors in parallel, which can be combined into one resistor producing noise (see Figure 7.6).

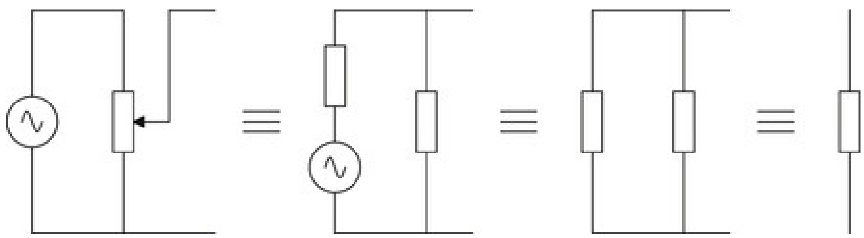


Figure 7.6 The noise sources in a potential divider volume control.

The thermal noise is calculated using $\sqrt{4kTBR}$. If we set bandwidth to 19,980 Hz (20 Hz to 20 kHz) and temperature to 20°C, this simplifies to:

$$v_n = 1.798 \times 10^{-8} \times \sqrt{R}$$

The absolute level of thermal noise generated at the output of a volume control is not especially useful because the output level changes with volume control setting. What is needed is to apply a standard level (2 V_{RMS}) to the input of the volume control and determine the S/N (Signal-to-Noise) ratio at the output of the volume control for each attenuation setting and plot it as a graph (see Figure 7.7).

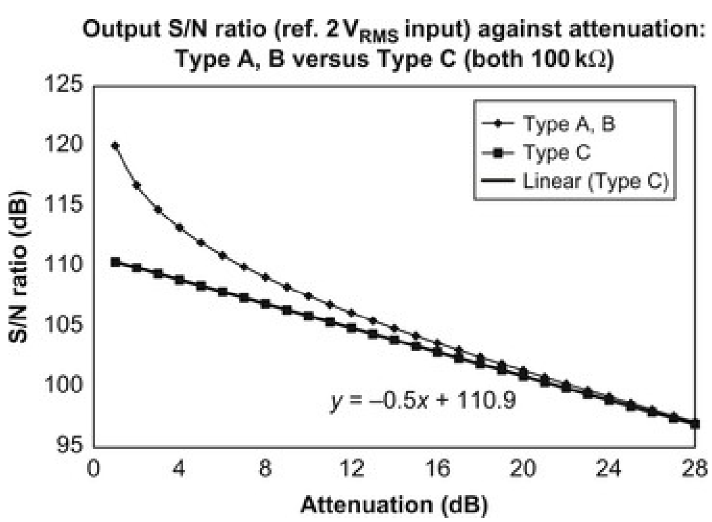


Figure 7.7 The effect on S/N ratio of self-generated noise in a volume control.

The graph compares the S/N ratio of a 100 kΩ Type A (tapped chain) against a Type C (switched shunt) using a fixed 100 kΩ series resistor, and shows that as attenuation increases the two controls become comparable, but that the Type A has a clear advantage at low attenuations. The shape of the curves is always the same, so they can be used to predict S/N ratio for any resistance value or signal level.

Resistors produce noise proportional to the square root of their resistance, so it follows that higher value volume controls produce more noise, and the change in noise caused by changing the value of a volume control can be found using:

$$\text{change S/N} = 20\log\sqrt{\frac{R_1}{R_2}} = 10\log\left(\frac{R_1}{R_2}\right)$$

Scaling a volume control's value by a factor of 10 changes the noise by 10 dB and a factor of 5 by 7 dB, so exchanging a 100 kΩ control for a 1 MΩ control increases noise by 10 dB, whereas exchanging a 100 kΩ control for a 20 kΩ control reduces noise by ≈7 dB.

Reducing the input level reduces the S/N proportionately, so transistor electronics standardised on 400 mV $_{RMS}$ uses a 5 kΩ volume control to achieve almost identical volume control S/N to valve electronics using a 100 kΩ volume control but standardised on 2 V $_{RMS}$.

Another way of generating noise at a volume control is to allow DC onto it. This should never occur because the preceding stage will always have a coupling capacitor to block its DC from the next stage, but a leaky capacitor would cause noise, as could significant grid current from the next stage.

Grid-Leak Resistors and Volume Controls

The grid-leak resistor of the following stage causes a problem for two reasons. Firstly, although the previous stepped attenuator programs were designed to account for it, any error in its value increases attenuation errors. Secondly, it causes the input resistance of the volume control to vary as attenuation is changed from maximum to minimum. These two potential problems warrant further investigation.

The Type A stepped attenuator in Table 7.1 was designed to be loaded by a 1 MΩ grid-leak resistor. Changing the value of this load resistance affects the maximum attenuation error (0.053 dB), so +10% causes the maximum error to rise to 0.060 dB, and −10% causes it to rise to 0.073 dB. Although this is an almost negligible degradation, it is not difficult to ensure that the load resistance is within 1 MΩ±10%.

When the volume control feeds a self-biassed cathode follower, the grid-leak resistor is bootstrapped (see Figure 7.8).

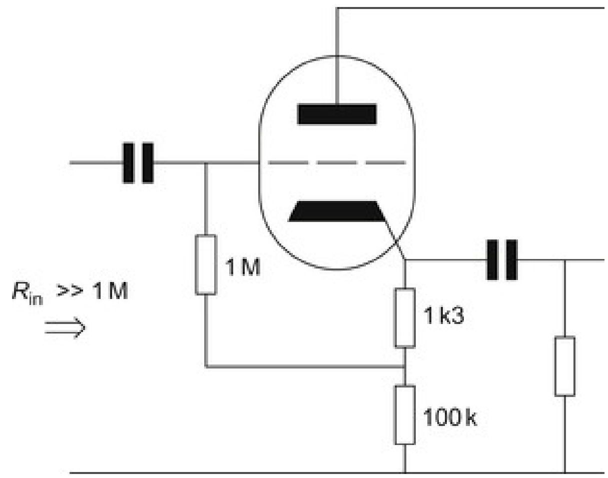


Figure 7.8 A self-biassed cathode follower bootstraps its grid-leak resistor, greatly increasing input resistance.

Bootstrapping increases the effective resistance of the grid-leak resistor (see Chapter 2) and therefore the loading resistance imposed on the volume control. For minimum error, the loading resistance should be measured or calculated and, if necessary, a resistor added in parallel with the output of the volume control to maintain the expected loading resistance of 1 MΩ.

A 1 MΩ grid-leak resistance in parallel with a 100 kΩ volume control set to minimum attenuation causes the input resistance of the volume control to be 91 kΩ, but as maximum attenuation approaches, its input resistance rises towards its nominal value of 100 kΩ. If the volume control is part of a filter, this changing resistance upsets the accuracy of that filter. If the filter is simply the arbitrary 7,950 μs (20 Hz) high-pass filter of an RIAA stage, then a 10% change is not a problem, but if it were part of the critical 75 μs section, a 10% change would be completely unacceptable. There is therefore a great temptation to discard the formal grid-leak resistor, recalculate the volume control and gain the benefit of an unchanging input resistance (see Figure 7.9).

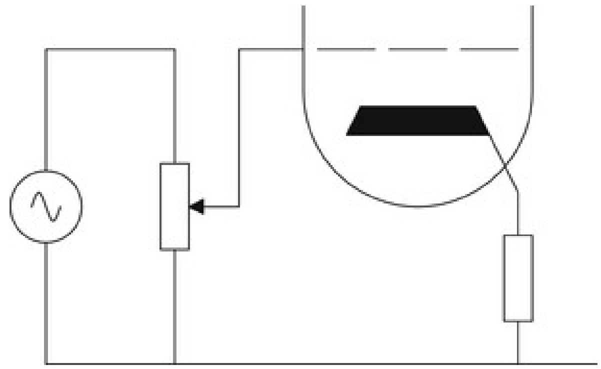


Figure 7.9 Beware using the volume control as the grid-leak.

The switch contacts are typically projecting studs, and the wiper is pressed against them by a spring. 'Make before break' switches are designed so that the wiper contacts the next stud before leaving the first (thereby maintaining contact with the resistor chain at all times), but as the wiper rotates and transfers from one stud to another there is a possibility that it may bounce and break contact with both studs. If the wiper does bounce and we have omitted the formal grid-leak resistor, the following valve's grid floats and it becomes a diode with only the anode load resistor to limit current flow. The anode voltage falls sharply, and when the wiper resumes contact, the anode returns to its design voltage equally sharply. Depending on the design of the stage, wiper bounce could easily cause >50 $V_{pk–pk}$ spikes at the anode. Because the spikes are (hopefully) of short duration, they are composed mostly of high frequencies and will be attenuated by any shunt capacitance; nevertheless, they could be sufficient to damage a succeeding transistor amplifier or ribbon tweeter.

Balanced Volume Controls

It may be that you want to build a balanced pre-amplifier. In which case, the volume control should be balanced too. Curiously, people persist in placing notionally identical controls in *each* path of each channel. This is wrong. The two mechanically ganged attenuators cannot have perfect channel matching, so a common-mode noise signal such as hum must be attenuated unequally, converting a proportion of it to differential mode, to which a balanced pre-amplifier is sensitive. The correct way to construct a balanced volume control, using the minimum number of components, is to use a balanced Type C attenuator having a fixed series resistor in each leg (see Figure 7.10).

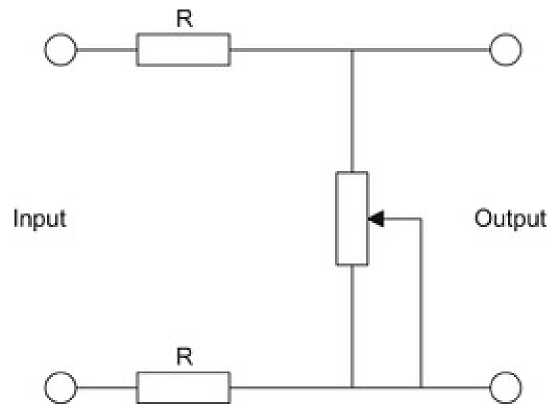


Figure 7.10 Balanced volume control.

The reason that the balanced Type C configuration in superior is that degradation of common-mode rejection is reliant solely on the matching of the two fixed series resistors, which should therefore be 0.1% tolerance or better. If you use the computer program to determine values, remember that the series resistor that the program uses is *twice* the value of the series resistor in each leg. Unfortunately, the Type C attenuator has the disadvantage of a high output resistance when set for a sensible input resistance and, in combination with the input capacitance of the following stage, this causes high frequency loss if ignored.

Light-Sensitive Resistors as Volume Controls

Cadmium sulphide (CdS) light-sensitive resistors such as the ORP12 are occasionally mooted as possible volume control elements.

As an experiment, the author fitted a pair of ORP12 light-sensitive resistors into a carefully machined 2"-long aluminium tube, one at each end, and added light-tight seals. A 28 V 40 mA incandescent lamp was fitted axially midway between the two resistors with a light-tight seal. One ORP12 had a dark resistance of ≈5 MΩ, and the other 100 kΩ. When fully illuminated, one had a light resistance of 69 Ω, and the other 63 Ω. Clearly, we would need to select these devices to have any hope of making a stereo volume control that tracked.

When a 1 kHz sine wave at +30 dBu was applied via a 10 kΩ series resistor, with the resistors dark, 1% distortion resulted (pure second harmonic). Once the input level was reduced to +8 dBu (2 V_{RMS}), the distortion fell to 0.02%. As the resistors were illuminated, both output voltage and distortion fell (Table 7.2).

Table 7.2 Comparison Between Light-Sensitive Resistor and Type A Switched Attenuator

	CdS attenuator	Type A switched attenuator

Distortion	Benign (second harmonic)	Barely measurable
Channel matching	Poor, needs to be selected	Easily betters 0.05 dB
Logarithmic law	Poor	Essentially perfect
Ease of construction	Difficult to make light-tight	Slightly fiddly

Provided that light-sensitive resistor selection could reduce the channel matching error to acceptable levels, the logarithmic law problem could be solved by adding a correction signal to the illumination control current derived via a microcontroller's internal 8-bit DAC driven by a look-up table; look-up table values would be adjusted until the law was acceptable. Beware that LEDs easily respond to 1 μs pulses and although CdS resistors tend to be sluggish by comparison, noise could still be coupled into the audio from the control current, so incandescent illumination could be a better choice.

Classic valve compressors successfully used illuminated light-sensitive resistors as a means of controlling gain because 0.02% of pure second harmonic distortion is quite good performance for an electronically controlled gain element. A volume control doesn't need fast-responding electronic control of attenuation, so why tolerate even 0.02% distortion and poor channel matching?

Transformer Volume Controls

The previous volume controls have all attenuated the signal by interposing a variable resistance, but in doing so their output resistance has always been higher than the source and they have added noise.

A recently fashionable method of controlling volume has been to use a switch to select between different tappings on a transformer. The stated advantages are that the transformer actually reduces source resistance because it is divided by the square of the selected turns ratio (although winding resistances are in series) and that there is no loss of signal energy (true, but irrelevant). Perfect channel matching is a given with transformer volume controls because attenuation is determined by the turns ratio (counted by a machine), and the significantly reduced internal resistance reduces noise compared to a resistive attenuator. Just like the resistive switched attenuator, the control has a limited number of steps, not just because of switch limitations, but also because of the practical difficulty of making a large number of transformer taps.

Fortuitously, the large number of taps tends to force a large coil former and correspondingly large core, and the levels to be coupled are quite low, so core flux densities are low, enabling a mu-metal core, which greatly increases primary inductance over grain oriented silicon steel. Taken together, low core flux density and high primary inductance mean that we should not expect the usual transformer bugbear (distortion at low frequencies (LFs) and high amplitudes) to be noticeable. Low frequency distortion falls with level but rises with source impedance, so the <0.1% distortion at 20 Hz from a 5 Ω source even at +30 dBu (24.5 V_{RMS}) achieved on test by a Stevens & Billington TX102 transformer volume control confirms that low frequency distortion is not a problem.

However, high frequency response is somewhat more problematic because as attenuation increases, the secondary tapping moves to use less and less of the coil, so coupling becomes poorer and leakage inductance rises, causing either ultrasonic ringing or high frequency loss, dependent on volume control setting and source resistance. With good design, these defects can be minimised, but they remain larger defects than found in a resistive attenuator.

The real advantage of transformer volume controls is not that they attenuate the signal more faithfully than a resistive attenuator (they don't), but that because they are a transformer, they break the earth loops between equipment that add broadband noise. Noting the eyebrow-raising price of volume control transformers and their limited number of coarse taps, the author would rather break earth loops with a well-balanced input transformer, and then use a resistive attenuator.

Balance Control

It is perfectly possible to make a balance control using a pair of reduced range switched attenuators, but we will see shortly that law-faked balance controls using conductive plastic linear potentiometers can be so good that it's simply not worth the trouble.

Law Faking

One way of approximating a logarithmic law with a linear potentiometer is to add a stand-off resistor between its lower leg and ground and a law-faking resistor between wiper and ground (see Figure 7.11).

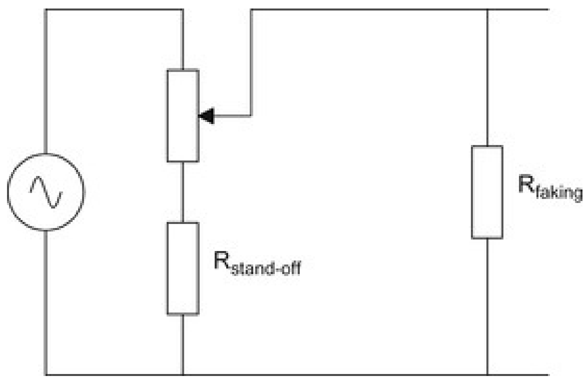


Figure 7.11 Faking a logarithmic law from a linear potentiometer; arrangement of resistors.

This simple solution can't turn a linear potentiometer into a perfect logarithmic potentiometer, but as Table 7.3 shows, it works very well for balance controls where only a limited range of attenuation is required.

Table 7.3 Required Resistor Values for Law Faking

Maximum attenuation (dB)	Worst error (dB)	Stand-off resistor ratio	Faking resistor ratio	Minimum input resistance ratio
26	2.4	0.0582	0.1674	0.145
±8	0.51	0.25	0.5916	0.402
±6.1	0.22	0.4706	1	0.595
±4	0.085	1	1.988	≈1
±3	0.025	1.5	3.204	1.404

Example : A 100 kΩ linear potentiometer is available and needs to be law faked to give the best approximation to logarithmic attenuation over 0–26 dB of range. The stand-off resistor becomes 100 kΩ×0.0582=5.82 kΩ (5k6 Ω in series with 220 Ω), and the law-faking resistor becomes 100 kΩ×0.1674=16.74 kΩ (18 kΩ in parallel with 240 kΩ). However, the input resistance of the volume control is no longer constant and falls to a minimum of 100 kΩ×0.145=14.5 kΩ

– we pay a heavy price for attempting to law fake over such a wide range.
The remaining four attenuators are intended to be used as balance controls, so they are defined by their deviation from mid-position rather than maximum attenuation, and it is assumed that each is followed by gain equal to its deviation from mid-position.
Although the ±8 dB option still has a somewhat onerous input resistance, the remaining three options are perfectly usable and all have been designed to give symmetrical attenuation about mechanical centre (see Figure 7.12).

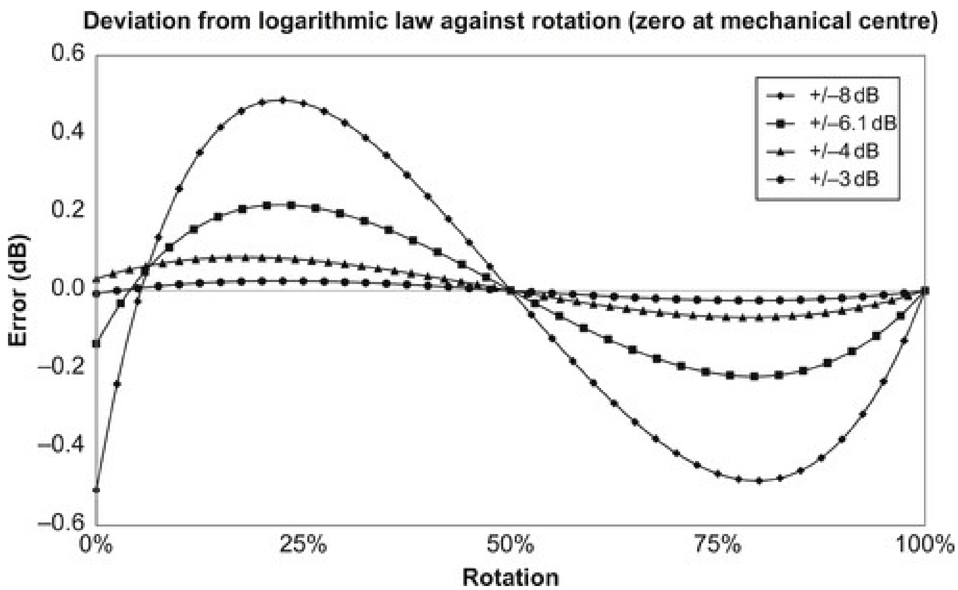


Figure 7.12 Deviation from logarithmic law against rotation for the four law-faked balance controls.

If we pan a camera across a small bright light, we don't expect the light's brightness to change as we pan, so we don't expect an audio source to change its volume if we use a balance control to alter its apparent position between the loudspeakers. Balance controls should be constant volume.
Symmetrical attenuation is an essential requirement for a constant volume stereo balance control because although the attenuator in one channel is wired conventionally (attenuation decreases clockwise), the attenuator in the other channel must be wired in reverse (attenuation increases clockwise). Asymmetrical attenuation could still allow electrical centre to correspond to mechanical centre, but adjustment of the control away from centre would change perceived volume.
The significance of the ±6.1 dB attenuator is that it uses a faking resistor equal to its value, implying that an existing 100 kΩ volume control could be immediately preceded by a tandem 100 kΩ linear potentiometer (perhaps with centre detent) plus 47 kΩ stand-off resistors, making a very simple constant

volume balance control having a range of ±6.1 dB.

However, the addition of such a balance control means that the volume control is no longer driven from a source of zero resistance and that must affect its law. Worse, once the balance control is moved from its centre position, each channel's volume control is driven from a different resistance and there is a danger that rather than the balance control causing the desired fixed inter-channel balance difference, it might cause a difference that varies with volume control setting (see Figure 7.13).

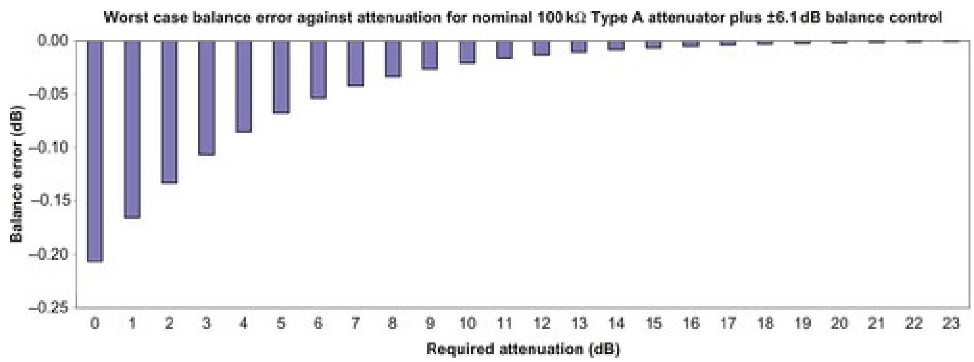


Figure 7.13 Worst case balance error against attenuation for nominal 100kΩ Type A attenuator plus ±6.1 dB balance control.

The graph shows the calculated inter-channel balance error of a nominal 100 kΩ Type A-stepped attenuator preceded by a ±6.1 dB law faked balance control as described. The balance control has been swung hard to one end to produce a 12.3 dB difference in level between the channels and it can be seen that the worst deviation from this is 0.2 dB at full volume. At a far more reasonable setting where the balance control causes a 3 dB level difference between channels, the worst case error falls to an entirely negligible 0.042 dB at full volume. The ±6.1 dB law-faked balance control has two remaining niggles that will probably go unnoticed in practice but should be mentioned for completeness:

• Retrofitting the balance control to a stepped attenuator significantly worsens its deviations from true logarithmic law (typically from ±0.06 dB to ±0.27 dB) because source resistance is no longer zero. However, since the primary justification for a stepped attenuator was good channel matching, and this remains unchanged, so <0.3 dB deviations from strict logarithmic law are entirely acceptable. If absolutely necessary, the stepped attenuator could be redesigned to account for the mid-position output resistance of the balance control.

• Input resistance drops to 59.5% of volume control resistance. This is unlikely to be audible, but the change from 100 kΩ to 59.5 kΩ input resistance just might cause a measurable change in frequency response or distortion of the previous stage. If this is a problem, dropping to ±4 dB balance range by using a 50 kΩ balance potentiometer with 100 kΩ stand-off resistor raises the minimum input resistance to that of the original volume control.

Although the ±4-dB and ±3-dB options can be used as constant volume balance controls, their more likely use is as analogue fine gain trimmers in active crossovers. As an example, if an active loudspeaker had just been built and was being calibrated, it would be very handy to use Vishay T73 trimpots as ±3 dB gain trimmers because each of their markings would correspond to 1/2 dB, so a measured 3/4 dB gain discrepancy on one driver could be quickly corrected with confidence. The ±3 dB option can be particularly conveniently implemented using a 50 kΩ potentiometer plus 75 kΩ stand-off and 160-kΩ faking resistors (or in semiconductor electronics, 5k potentiometer, 7k5 Ω stand-off, 16 kΩ faking).

Finally, remember that the centre attenuation of any balance control must be made up elsewhere, so fitting the ±6.1 dB balance control requires the addition of 6.1 dB gain.

Cable Driver

The most common reason for needing a cable driver is that the power amplifiers have been sited adjacent to the loudspeakers in order to minimise the length and consequent resistance of loudspeaker cable, but that the sources and volume control have been placed conveniently to hand at the listening position, necessitating a long cable to the power amplifiers.

Determination of Required Quiescent Current

Once cable is routed inconspicuously from one point to another, its required length quickly escalates. If we assume that the cable is 20 m long, typical cable capacitance of 100 pF/m would produce a capacitance of 2 nF. If we also assume transistor power amplifiers (!), we should add a further 1 nF, giving a total load capacitance of 3 nF. If we want to restrict the loss at 20 kHz to 0.1 dB when driving this capacitance, we must set f_{-3dB} =131 kHz, which requires a source resistance of ≈400 Ω.

Almost any valve used as a cathode follower can achieve this small-signal output resistance, but the more significant question is whether it can supply the required current without distortion. The reactance of a capacitor falls with frequency, so we take the worst case, and find the reactance at 20 kHz:

$$X_c = \frac{1}{2\pi fC} = \frac{1}{2 \times \pi \times 20{,}000 \times 3 \times 10^{-9}} \geq 2{,}650\ \Omega$$

If we assume that the power amplifiers have a sensitivity of 2 V_{RMS}, driving them to full power at 20 kHz implies applying 2 V_{RMS} across this reactance. By Ohm's law, this requires a current of:

$$i = \frac{V}{X_c} = \frac{2}{2{,}650} \geq 750\ \mu A_{RMS}$$

When we consider loadlines and valve operating conditions, we must work in peak currents and voltages, so 750 μA_{RMS} ≈1 mA_{pk}. Because this current is required by a capacitor, it forces a typical loadline to change from a straight line to an ellipse (see Figure 7.14).

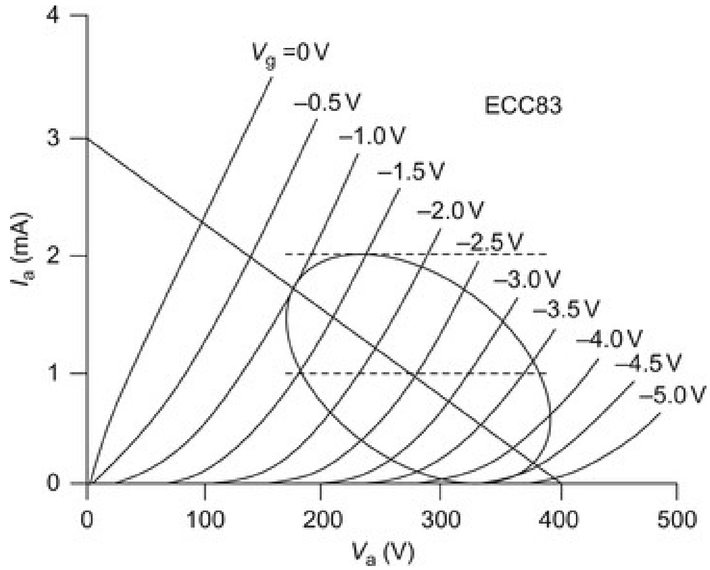


Figure 7.14 Elliptical loadline caused by capacitive load.

The diagram is somewhat exaggerated to aid clarity because our cathode follower will not need to swing as many volts as shown, but the effect of a capacitive load can be clearly seen. Driving the cable capacitance forces our line stage to swing vertically ±1 mA without any change in voltage (this is due to the 90° lag between current and voltage in a capacitor). As an absolute minimum, the (Class A) line stage must pass 1 mA of quiescent current so that it can swing 1 mA to the load without switching off.

Choice of Follower Valve

To achieve good linearity in the face of a heavy reactive load, a cathode follower needs plenty of distortion-reducing negative feedback, so μ needs to be as large as practicable. More importantly, we need g_m to be constant with current because we know that our elliptical loadline forces changing current. Sadly, constant g_m in the face of changing I_a is impossible, but the valve that comes nearest is the Russian 6C45 π single triode (other possibilities are triode-strapped D3a, or 6H30P) (see Figure 7.15).

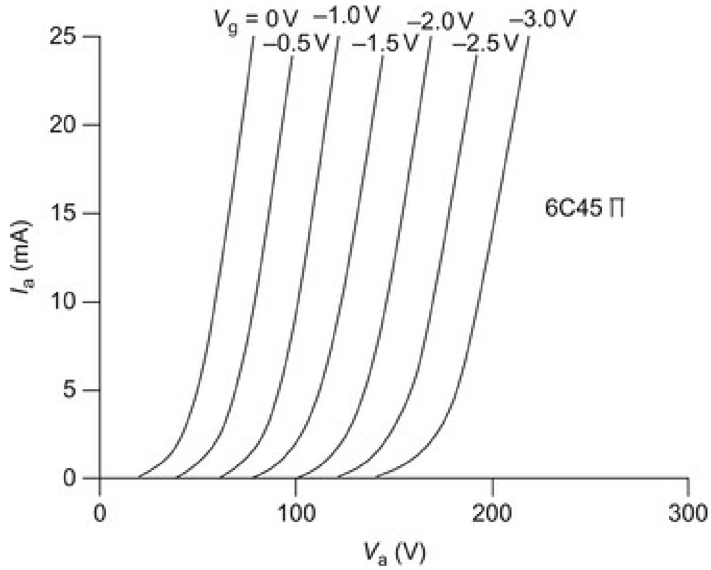


Figure 7.15 6C45 π anode curves.

Looking at the curves, we find that once $I_a > 5$ mA, the curves are nearly straight, and bunching (which reduces g_m and μ) is almost non-existent. We know that our quiescent current must be 1 mA clear of bunching, so we could operate the valve at $I_a = 6$ mA, but this would be a little marginal, and 10 mA would be better.

Now that we have chosen I_a, we need to set V_a. The main limitation on V_a is that it must allow us to avoid grid current at our chosen I_a. If we choose $V_{gk} = -2.5$ V, this is nicely clear of grid current, and sets $V_a = 170$ V. If we use a 390 V HT (in common with the EC8010 RIAA stage), we must drop 390 V−170 V−2.5 V=217.5 V, and if this passes 10 mA, a 22 kΩ load resistor would be required. Even for the 6C45 π, this is a moderately steep loadline, which increases distortion before feedback, so we will use an EF184 constant current sink (CCS) to force the resistive component of the loadline to be horizontal (see Figure 7.16).

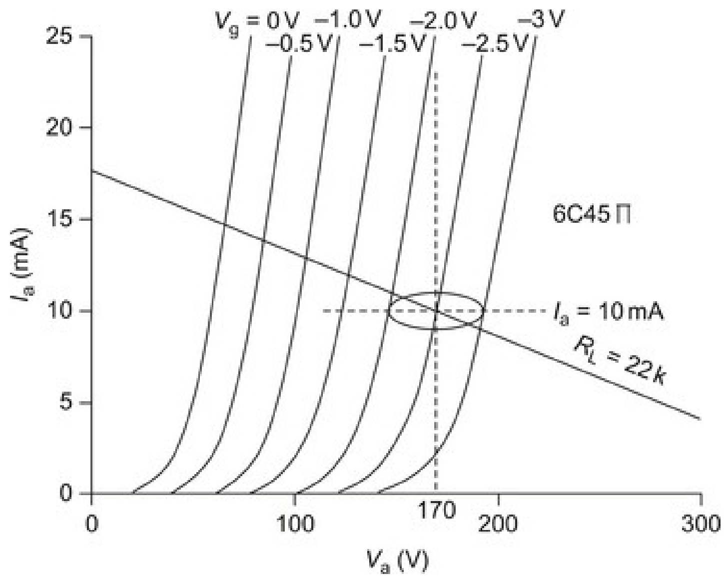


Figure 7.16 6C45 π operating conditions in unity gain cable driver.

Practical Considerations

Adding the EF184 CCS dramatically reduces distortion, but adds practical problems related to g_2. Firstly, under these conditions, to sink 10 mA at its anode, the EF184 requires 4.4 mA of g_2 current, increasing the HT current for a stereo pair to almost 30 mA.

Secondly, if the 6C45 π should fail to draw current for any reason, g_2 of the EF184 would act as an anode and pass the full programmed current of 14.4 mA, causing its dissipation to rise to 2.5 W $(P_{g_{2(max)}} = 0.9\ \text{W})$ which would destroy g_2. This g_2 problem is common to all pentode circuitry, and is commonly solved by supplying g_2 via a resistor, which limits current, so the 39 kΩ resistor protects g_2 in this circuit (see Figure 7.17).

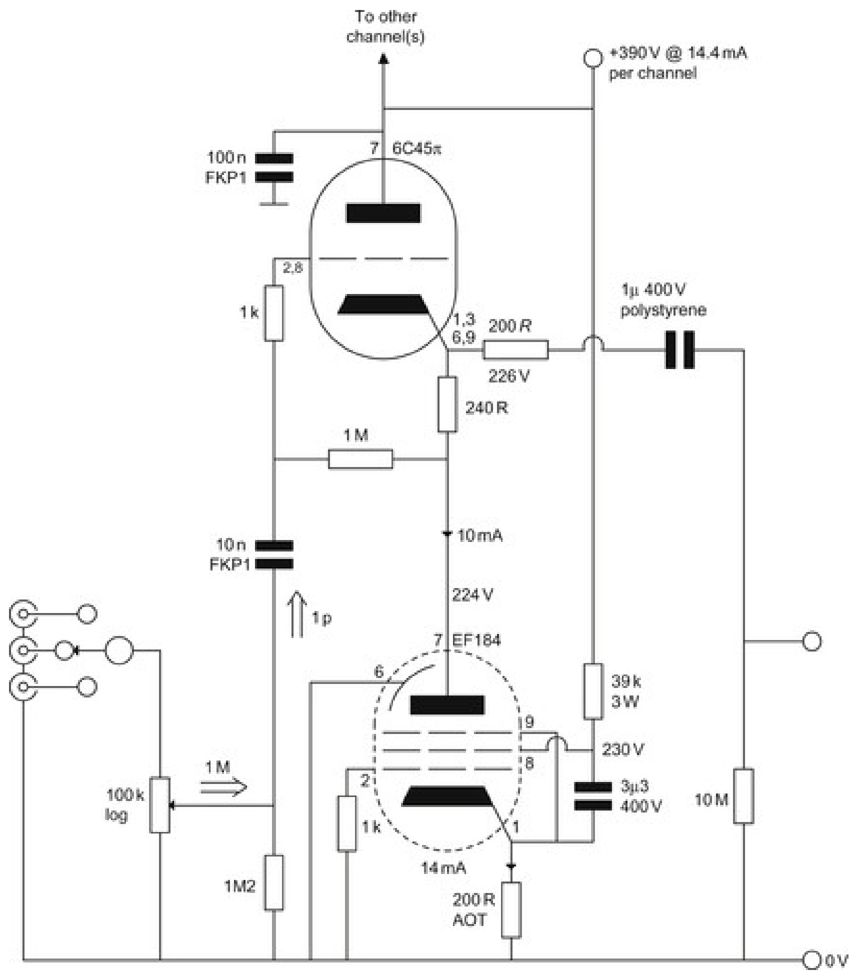


Figure 7.17 Low distortion unity gain cable driver.

However, we can improve the performance of any pentode circuit by supplying g $_2$ from a low impedance supply (because I_a is far more dependent on V_{g_2} than V_a), but this offers no protection against the g $_2$ problem. If a simple regulator such as the THINGY (see Chapter 5) were used to supply g $_2$, the large and expensive 3.3 µF 400 V capacitor could be replaced by a smaller collection of cheap components offering better performance, but there would then be no protection against the g $_2$ problem. Although the EF184 is cheap, we would still prefer them to die of natural causes rather than being mugged.

If this circuit was powered in the traditional fashion using a valve rectifier and all supplies originating from one transformer, the EF184 might heat faster than the 6C45π, leaving the EF184 vulnerable to the g $_2$ problem. Thus, the THINGY was reluctantly rejected, and the lower performance option of supplying g $_2$ via a resistor was adopted.

The 6C45π is self-biassed by the voltage dropped across the 240 Ω cathode resistor because fixed bias from a potential divider chain would inject HT noise into the grid circuit. Because the measured input capacitance was only 11 pF, this factor does not restrict the choice of volume control, but distortion rose slightly (from 0.02% to 0.05% at +20 dBu) when the source resistance exceeded 150 kΩ, suggesting that there was some grid current. A subsequent test on a modified AVO VCM163 confirmed the hypothesis, as DC grid current was found to be constant at ≈0.1 μA when V_{gk} was swept from −1 V to −3 V.

Given half a chance, the 6C45π oscillates enthusiastically when configured as a cathode follower. The HT must be locally bypassed to ground at radio frequencies, hence the 100 nF polypropylene capacitor from the anode pin to ground (polyester might be too lossy at radio frequencies to adequately decouple). In addition, 1 kΩ grid stopper and 200 Ω cathode stopper resistors were essential to suppress oscillation at 70 MHz. If you don't have an oscilloscope that can reliably see such a high frequency at its probe tip, it might be worthwhile to raise the grid stopper to 4.7 kΩ. Similarly, because $g_m \approx 16$ mA/V at the operating conditions, the cathode stopper resistor could be raised to 330 Ω, and still keep r_{out} below the required 400 Ω.

Passing $I_a = 10$ mA through the EF184 with $V_a = 221$ V means that it dissipates $P_a = 2.2$ W, which is close to the 2.5 W limit, but the EF184 is cheap and plentiful, so we need not worry about a reduced lifetime. The DC conditions of the EF184 were determined in the usual way, but the value of the cathode resistor is critical to set the anode current correctly to 10 mA, so it may need to be adjusted on test. The easiest way to measure anode current is to measure the voltage across the 240 Ω 6C45π cathode resistor using a DVM and then adjust the EF184 cathode resistor until 2.4 V is seen.

Adding Gain

Once a balance control has been added, its mid-position attenuation must be corrected, and the easiest way to achieve this is to direct couple a common cathode gain stage to a cathode follower, and then wrap negative feedback around both to achieve the required low gain (probably 6.1 dB). Parallel derived feedback further reduces output resistance, but the feedback could be series or parallel applied.

Series applied feedback to the input valve's cathode forces a very low value of feedback resistor, steepening the cathode follower's loadline and increasing its distortion before feedback. Parallel applied feedback allows a larger feedback resistor but requires a large resistor in series with the input valve's grid that increases thermal noise and potentially converts input valve grid current into distortion.

The input valve must have low grid current, and high- μ allows more feedback, making an ECC83 loaded with a CCS a good choice because its gain is then equal to μ (typically 90). An insipid green LED drops about 1.8 V at 300 μA, conveniently biassing the ECC83. The input resistor loads the volume control, so it must be 1 MΩ, and this will generate 18 μV $_{RMS}$ of noise; −101 dB referred to 2 V $_{RMS}$, so it is just tolerable. Rearranging the standard feedback equation, we have:

$$\beta = \frac{1}{A} - \frac{1}{A_0} = \frac{1}{2} - \frac{1}{90} \geq 0.48$$

From which we can find the feedback resistor:

$$R_{\text{feedback}} = \frac{R_{\text{input}}}{\beta} = \frac{1\,\text{M}}{0.48} = 2.07\,\text{M}$$

The standard E24 value of 2M2 will do and certainly not load the cathode follower. We have reduced gain from 90 to 2, corresponding to a feedback factor of 45, so the cathode follower's output resistance and distortion will be reduced by a factor of 45. Such a large reduction in output resistance and distortion means that, provided it passes sufficient current (≈10 mA), we need not worry too much about its exact operating point (just as well, because the ECC83's anode could be anywhere between 130 V and 200 V), so an E88CC loaded by a CCS will do very nicely (see Figure 7.18).

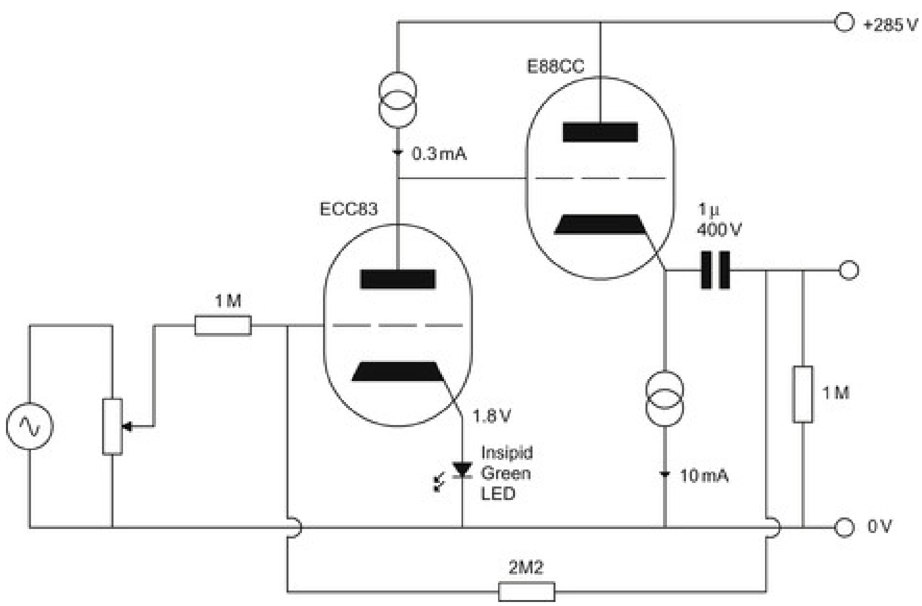


Figure 7.18 ECC83/E88CC anode follower.

It appears as if the volume control is being used as the grid-leak, and this would not be ideal, but the ECC83's grid has a DC path to 0 V via the 2M2 feedback resistor and 1 M output resistor – so it is vital to include that output resistor!

Polarity Inversion

Note that adding gain to the cable driver caused an inversion of output polarity.

Every so often a fuss is made in the Hi-Fi press about 'absolute phase', [1] with claims that some tracks of some recordings sound more realistic when inverted. There *is* a reason why reproduction could sound different with one polarity rather than the other, but it's nothing to do with incorrect recordings. Distortion occurs once a loudspeaker cone changes the volume of air in its enclosure by >5%, and because that distortion differs between compression and expansion, a sound such as a bass drum (which contains a significant DC component) sounds different reproduced on such a loudspeaker dependent on signal polarity. This is another reason for the author's earlier assertion that any loudspeaker smaller than a domestic washing machine is small – a 15″ driver in a washing machine-sized enclosure reproduces bass with far lower distortion than a long-throw driver straining a shoe box.

1. Strictly, it's polarity, but an engraved "Φ" takes up far less room on a microphone body or cluttered control panel.

Polarity is important when recording and particularly mixing. It is common to use two microphones when recording a snare drum, one inside, one outside – the relative balance between the two microphones adjusts the recorded sound's timbre. When the skin is struck, it moves towards one microphone and away from the other, so one microphone must have its polarity inverted to prevent LF cancellation when they are summed in the mixer. But which has the 'correct' polarity? The inside microphone, or the outside?

In short, the author considers absolute polarity to be floobydust when it comes to reproduction, but you might have a different view and wish to ensure that any unavoidable polarity inversion is corrected.

Tone Control

If a tone control is to be included to compensate for poor recordings, then it must be implemented carefully. Experimentation with the Tone Stack Calculator freeware (available from the Duncanamps website) will quickly convince you that although the James [3] control is the least worst of the passive RC bunch, it requires logarithmic controls to even approximate to a flat electrical response at mechanical mid-position. The only plausible Hi-Fi tone control (as opposed to electric guitar effects) is the negative feedback Baxandall [4] control because it achieves a flat response with linear controls at their mechanical mid-position (see Figure 7.19).

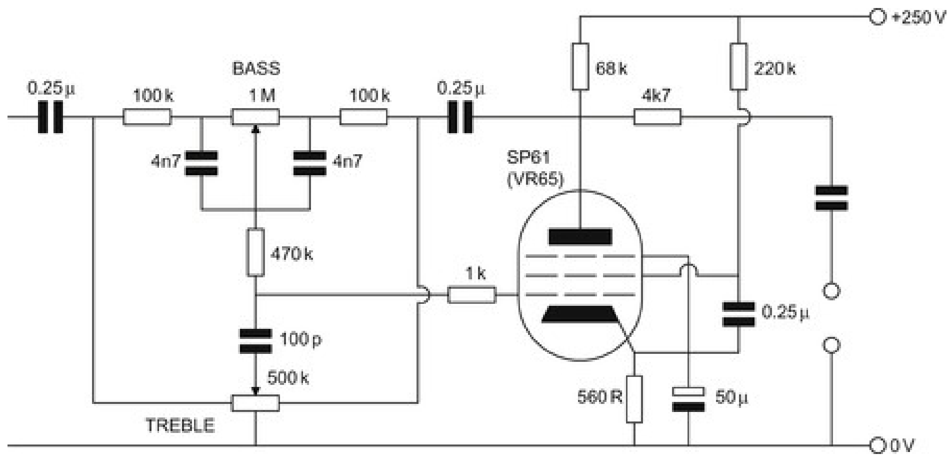


Figure 7.19 Noise in the input stage.

As originally presented, Baxandall's tone control offers an absurd amount (theoretically 20 dB) of boost and cut. Although such an amount might be used to achieve a particular sound when recording, 20 dB of 'correction' during reproduction is excessive and ±10 dB is quite enough to allow the sweet spot to be found quickly by applying too much correction then backing off.

Note that there isn't an explicit grid-leak resistor – the control grid's path to 0 V is via the wiper of the bass control and the centre tap of the treble control. This is not a good practice as it will ultimately lead to noise as the bass control's wiper is moved.

Remembering that the stage is essentially a unity gain inverter, Baxandall stated that a high-slope (high g_m) pentode was necessary for two reasons:

- C_{ag} is part of the feedback impedance, so it should be minimised, otherwise the full treble boost will not be available. EF37A: C_{ag} <0.02 pF; 6SN7: C_{ag} ≈4 pF.

- It ensures that $A_0 \geq 100$ in order that there is still enough feedback available even when full boost ($A = 10$) is invoked to keep distortion low.

Reducing the maximum correction from 20 dB to 10 dB slightly relaxes the open-loop gain requirement, but minimising C_{ag} still makes sense. Baxandall recommended the EF37A (precursor to the EF86) because its top-cap grid connection avoided the grid to heater leakage resistance of the phenolic octal base that would otherwise cause hum. Top-cap grids are a nuisance, so possible alternatives are shown in Table 7.4.

Table 7.4 Comparison of Possible Pentodes for Baxandall Tone Control

	C_{ag} (fF)	Typical g_m (mA/V)	Hum at grid (μV) [a]	Comments
EF37A	<20	1.8	Not specified	NOS available
EF86	<50	2.2	2	NOS rare, expensive and often noisy
E180F	<30	7	100	NOS available
EF184	<5.5	15	Not specified	NOS cheap as chips

[a] Manufacturer’s specification using centre-tapped AC heater.

Both the EF37A and EF86 have helically wound heaters to reduce the magnetic hum field. Sacrificing an E180F and an EF184 to science revealed that both had conventional loop heaters, explaining the far poorer E180F hum specification and suggesting that both EF184 and E180F would need DC heaters. Hum aside, the EF184 is the stand-out modern contender because not only does it have superior specifications to the other valves, but it’s also cheap. It was time for some EF184 pentode measurements.

Perusing the Mullard EF184 pentode mutual characteristics suggested that $V_a = V_{g_2} = 170\ \text{V}$, and $V_{gk} = -2$ V ought to result in $I_a \approx 10$ mA and $I_{g_2} = 4.5\ \text{mA}$. For a sensible HT voltage, this indicated a 10 kΩ anode load and a 27 kΩ screen grid resistor; the author had 11 kΩ 5 W and 27 kΩ 2 W, and a good fit to the forward drop of a typical insipid green LED is:

$$V_{\text{insipid green}} = 0.0337 I_{\text{LED (mA)}} + 1.837\ \text{V}$$

Thus, an insipid green LED drops ≈2 V at 14.5 mA and can be used to cathode bias the EF184 (see Figure 7.20).

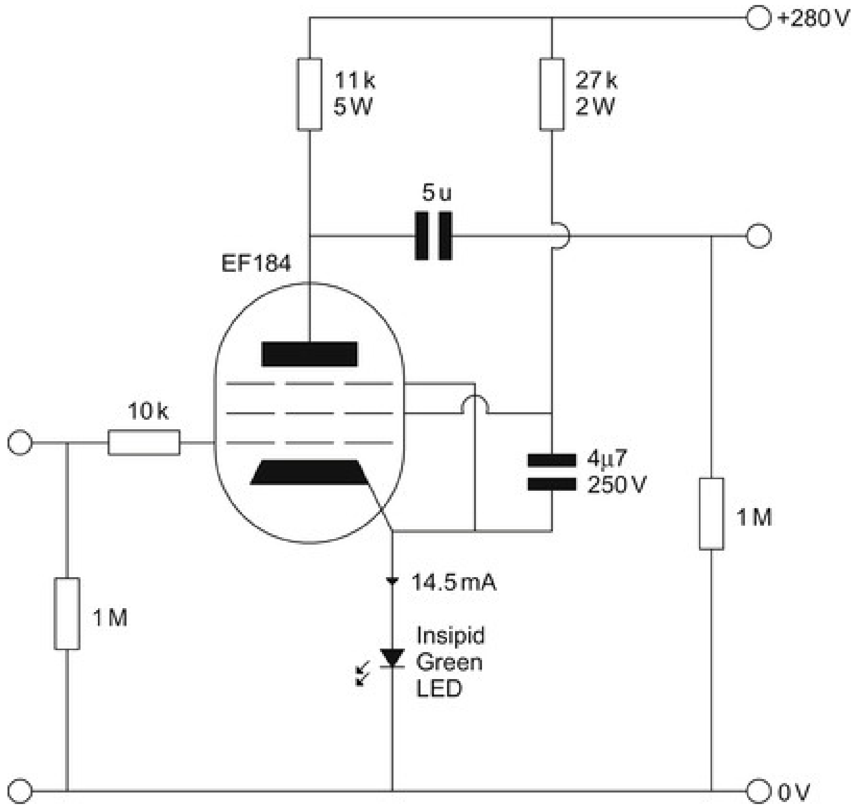


Figure 7.20 EF184 pentode distortion test circuit.

A bogey EF184 having DC characteristics that reasonably matched the Mullard data sheet and average distortion was selected from a stock of 33 EF184s. The stage produced 3% distortion at the author's standard test level of +28 dBu, dominated by third harmonic, which wasn't at all promising. The gain of 42 dB was a little lower than expected, so a 4,700 µF cathode bypass capacitor was added to see if the LED's predicted 112 Ω slope resistance was significantly reducing gain. Adding the capacitor increased the gain by 1 dB but rose distortion by 2 dB, so it was promptly discarded.

Since we know that a tone control invariably follows the volume control and precedes the power amplifier, we also know that it is unlikely to have to develop more than 2 V_{RMS} (+8.2 dBu) at its output, so the output test level was dropped to +8 dBu. At +8 dBu, the distortion was 0.28% (which is what one should expect, dropping level by 20 dB), but more importantly distortion was dominated by second harmonic. It was clear that it was necessary to measure individual harmonics against level (see Figure 7.21).

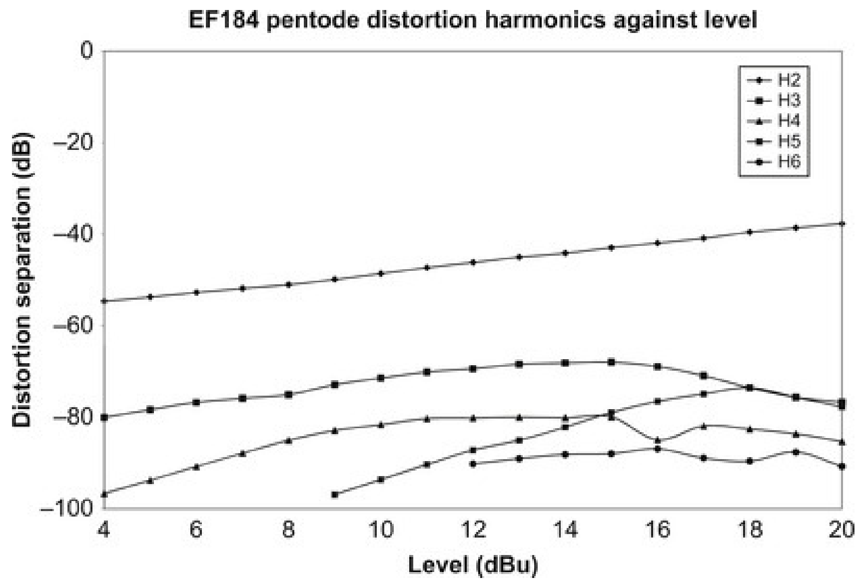


Figure 7.21 EF184 pentode distortion harmonics against level.

It's best not to look at EF184 pentode distortion after looking at *SN7*/N7 distortion, but provided level is <+10 dBu, it's not too bad. More significantly, most of the 42 dB open-loop gain will be used as negative feedback to reduce distortion. As a rough example, if 30 dB of the gain was used as feedback, at +8 dBu we should expect H2≈−80 dB and H3≈−105 dB (because THD<1%, H3 generated by negative feedback will be negligible). Testing the stage as a unity-gain inverter at +8 dBu, the author measured H2=−83 dB and H3=−95 dB (the H3 was probably test set residual). More significantly, the stage didn't burst into RF oscillation with a capacitive load directly on its anode, presumably because of the EF184's astonishingly low C_{ag}.

Like all frequency-selective feedback amplifiers, the Baxandall tone control has an input impedance that changes horribly with frequency, so unless the preceding stage has plenty of feedback and current capability, this frequency-dependent loadline is converted into frequency-dependent distortion, which is almost certainly why tone controls are so frequently vilified. It's not that the tone control stage itself distorts, but it makes the *previous* stage distort, and worse, it causes frequency-dependent distortion that is especially fatiguing because it forces the ear to attempt to reject a complex pattern that changes with frequency rather than a simple fixed pattern. The Baxandall tone control stage must be preceded by a stage capable of driving a changing impedance with constant or negligible distortion.

Given that the Baxandall tone control inverts polarity, it would be useful if the preceding stage also inverted polarity because this would restore correct polarity. Combined with the current drive and distortion requirement, this suggests that the ideal preceding stage might be another unity-gain pentode inverter identical to that in the tone control. Admittedly, this becomes expensive in HT current, but that's the price of a decent tone control. Self [5] offers a much fuller discussion of the relative merits of the different forms of Baxandall tone control, so this author will simply content himself with presenting a version that he believes is appropriate for domestic reproduction (see Figure 7.22).

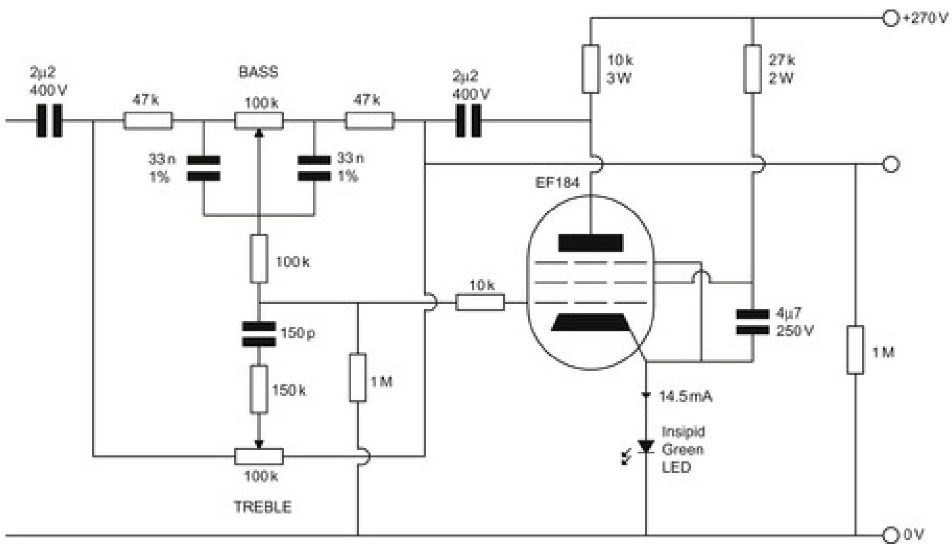


Figure 7.22 Contemporary implementation of Baxandall tone control.

The high g_m of the EF184 allowed a much smaller anode load resistor than Baxandall's, enabling a lower impedance frequency-selective network using 100 kΩ potentiometers. Thus, a fully featured pre-amplifier could use the same model of conductive plastic centre detent linear 100 kΩ potentiometer for bass, treble and stereo balance. The 33n capacitors associated with the bass control need to be 1% tolerance to ensure a flat response at the control's mid-position. Note that the loss of the treble control's centre tap requires the inclusion of an explicit grid-leak resistor.

The tone control has been configured to be asymptotic to ≈9.6 dB of boost and cut for both bass and treble, although a slightly smaller range is seen within the 20 Hz to 20 kHz audio bandwidth (see Figure 7.23).

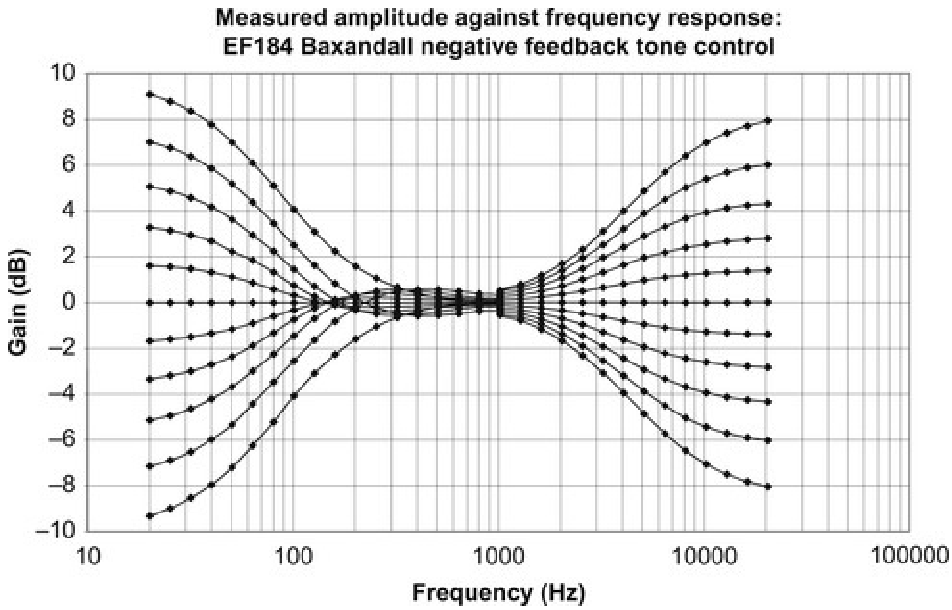


Figure 7.23 Basic RIAA pre-amplifier.

Rather than using infinitely variable potentiometers, the measured response was obtained by constructing two chains of ten 10 kΩ 1% resistors to make 100 kΩ controls. Thus, accurate tappings were available corresponding to 10% intervals of mechanical rotation. In addition, the EF184 gain stage used a 5 μF output coupling capacitor in order to place its low frequency loss at a sufficiently low frequency that it could not disturb the Baxandall curves. As can be seen, not only does the control have a flat response at mechanical centre, and very nearly symmetrical boost and cut, but also the amount of boost and cut nearly follows a logarithmic law with mechanical rotation, implying that the control will feel natural to use. The discontinuities at 1 kHz occur because bass responses were plotted up to 1 kHz and treble responses down to 1 kHz, but both responses had tails that extended a little further in frequency that were not plotted.

Because the control has plenty of feedback at all settings, output resistance is low, and with 10 mA of anode current the stage should be capable of driving cable capacitance. The controls were set for maximum bass and treble, and the stage driven at 20 kHz to produce 2.5 V_{RMS} at its output with a load of ≈300 pF, and this waveform was stored as a reference. When loaded with 6 nF, the difference between this and the reference was barely visible, but subtracting one from the other revealed a small sawtooth waveform indicative of the onset of slew rate limiting (see Figure 7.24).

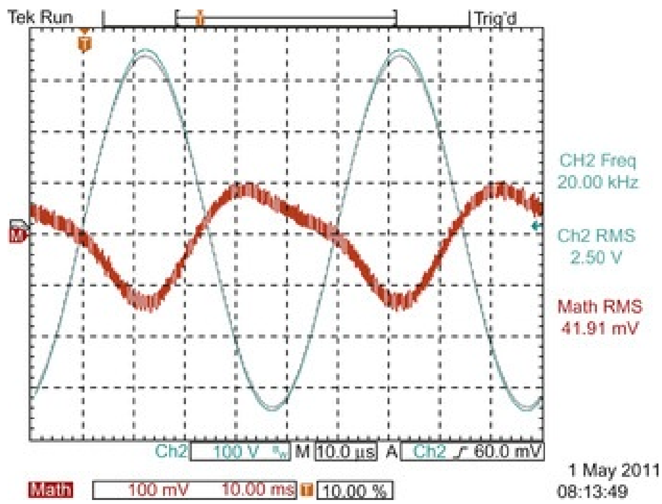


Figure 7.24 Tone control slow rate limiting at 2V $_{RMS}$ at 20 KHZ caused by 6 nf load.

The subtraction waveform reverted to a sine wave when load capacitance was backed off to 3 nF, implying that the stage is capable of driving full output at 20 kHz into 3 nF under its most onerous conditions. Interestingly, although a 20 kHz square wave provoked overshoot, no value of load capacitance provoked ringing even though the output was taken directly from the anode. It would probably be wise to check whether a different construction needed a small resistor (a few hundred ohms) between anode and cable.

Obtaining a Clean Signal from Analogue Disc

Before we can investigate the design of RIAA stages, we need to quantify levels and look at the practicalities of obtaining a clean signal from the cartridge.

Comparison of Analogue Levels between Vinyl and Digital Sources

To allow switching between sources with negligible volume change, the RIAA stage should match digital sources by producing 2 V_{RMS} at maximum level. Although digital sources define their 2 V_{RMS} output as being the maximum amplitude of undistorted sine wave that they can reproduce, analogue cartridges define their output voltage referred to a recorded (sinusoidal) velocity of 5 cm/s (sometimes 3.54 cm/s) at a frequency of 1 kHz. How can we reconcile these entirely different methods?

Vinyl is not nearly as tightly conformed as CD, and maximum recorded level is totally dependent on the skill and audacity of the cutting engineer. Recorded level may be reduced by as much as 6 dB in order to allow 40 min rather than 20 min to be recorded per side of a 33⅓ record. Since vinyl noise and dust clicks are constant, this automatically means 6 dB has been subtracted from that recording's dynamic range, so Beethoven's ninth symphony requires four vinyl sides, not two. RIAA equalisation further complicates matters, but when equalised vinyl typically has peaks (measured with a Peak Programme Meter) reaching 12 dB above the nominal 5 cm/s line-up level.

The significance of this is that we could now consider a cartridge specified to produce a nominal 2 mV_{RMS} at 5 cm/s as capable of producing musical peaks of 8 mV_{RMS}, which when multiplied by the 1 kHz gain of a suitable RIAA stage would give a signal level directly comparable to digital sources.

RIAA and Replay Rumble

RIAA is the abbreviation for 'Recording Industry Association of America' and is the de facto worldwide post-1954 standard for equalisation of microgroove records, as opposed to the numerous standards for 78s. Because the RIAA standard was not invented in Europe, but a worldwide standard *was* needed, the IEC invented an LP equalisation standard that was almost identical. The only difference is that the IEC standard recommends bass cut on replay only, with a −3 dB point at 20 Hz (7,950 μs) in order to reduce rumble. Most manufacturers of high-quality pre-amplifiers assume that their products will be complemented by equally good turntables and that replay rumble will not be a problem, so they

ignore the IEC recommendation. Their equalisation is therefore RIAA.
Nevertheless, there is considerable pressure to modify RIAA stages to include a low-frequency roll-off because:

- Some valve power amplifiers are susceptible to output transformer core saturation if high-amplitude signals are applied at low frequencies (<50 Hz).
- Bass reflex loudspeakers are easily overloaded at low frequencies because there is negligible damping of cone motion below their roll-off frequency. Bookshelf reflex loudspeakers tend to roll off below 100 Hz, whereas freestanding reflex loudspeakers could improve this to 50 Hz, or less, but this still leaves both vulnerable to low frequency noise.
- Vinyl records contain low frequency (<20 Hz) noise due to warps and rumble.

It is therefore argued that these problems could be avoided by implementing some form of LF roll-off within the RIAA stage. One possibility is to implement the IEC 7,950 μs recommendation, but a more sophisticated approach is to incorporate a properly designed high-pass filter having a final slope of 12 dB/octave, or more, set at ≈10 Hz.
The author firmly believes that neither of the preceding electrical approaches is correct and that RIAA equalisation should be reserved solely for correcting the record equalisation applied by the manufacturer at the time of cutting. CD players do not add a 10 Hz high-pass filter to solve the problems of poorly designed loudspeakers or questionable output transformers, so why adulterate vinyl? Warps and rumble are mechanical problems, and should have *mechanical* solutions, not electrical ‘fixes’.

The Mechanical Problem

Fortunately, a 12 dB/octave high-pass mechanical filter is unavoidably formed by the compliance of the cartridge suspension and the effective mass of the arm plus cartridge. The low frequency arm/cartridge resonance may be found using the standard resonance equation:

$$f = \frac{1}{2\pi\sqrt{Cm_{\text{total}}}}$$

where

C =cartridge dynamic vertical compliance (often different to the static value)

m_{total} =total effective mass.

Typical values might be:

Cartridge mass	5 g
Mounting hardware (screws and nuts)	1.5 g
Arm effective mass	12 g
Total effective mass (m_{total})	18.5 g
Cartridge dynamic vertical compliance (C):	15×10^{-6} dyne/cm

The previous figures applied to a unipivot arm designed for an Ortofon Quattro moving coil cartridge with its outer body removed, and resulted in the ideal resonant frequency of 10 Hz. The significance of determining this resonant frequency is that it is also the cut-off frequency of the high-pass mechanical filter.

It has been suggested that a higher resonant frequency (12–15 Hz) should be set, as this would be more effective at filtering LF noise, and this is quite true. However, we live in a practical world and dramatically reducing an arm's effective mass (to raise the resonant frequency) inevitably produces a flimsy structure only suitable for cartridges that do not transfer much vibration into the arm. Unfortunately, such cartridges are high compliance, and we are back where we started. Additionally, even setting the resonant frequency as low as 10 Hz means that the reproduced response (when RIAA equalised) is likely to be −1 dB at 20 Hz, depending on damping.

As an aside, the best way of reducing effective mass is to remove mass at the headshell. Modern arms are usually fixed headshell, so this leaves the cartridge. A moving coil cartridge often has a heavy outer shell that can save valuable grammes *if* it can be removed without damaging the internal workings. Even better, provided that the magnet system is firmly anchored, removing the body eliminates box resonances. The stylus cantilever is now completely exposed, which makes cueing and alignment easy, but leaves it frighteningly vulnerable to damage. All factors considered, the trade is well worthwhile, which is why some cartridges are sold naked.

Even if the arm/cartridge resonant frequency is correct, the mechanical high-pass filter can operate correctly only if the resonance is correctly damped, and some cartridges require extra damping. The general principle is that the moving pick-up arm is fitted with a paddle which is forced to move through a viscous liquid, thus damping the motion of the arm. Ideally, damping would be applied at the headshell because this would also attenuate high frequency energy transferred from the cartridge into the arm, and thereby reduce excitation of unavoidable high frequency structural resonances, but damping near the pivot, as used by almost all unipivots, damps the low frequency resonance equally well.

Additional damping has to be set by trial and error, and commonly, far too much damping is applied – the fluid is either too viscous, or there is too much of it. One way to set damping is to play a badly warped record with no damping applied, and observe cartridge movement as the warps are traversed. If the cartridge appears to bounce relative to the record surface, add a little fluid, and try again. Use as little damping as possible, as too much will *increase* low frequency noise and cause tracking problems at higher frequencies as undistorted stylus displacement is squandered by tracking warps rather than recorded signals.

Setting the mechanical high-pass filter correctly has two major consequences. Firstly, it means that we no longer need an electrical high-pass filter, but more importantly it means that stylus vertical deflection is considerably reduced and distortion generated by the cartridge falls.

Arm Wiring and Moving Coil Cartridge DC Resistance

Once universally ignored, the wiring resistance of a pick-up arm can become significant, particularly when a low output moving coil cartridge is stepped up by a transformer, because the transient response of that transformer is critically dependent on source resistance.

The author measured the resistance of a 5 m length of the fine wire used within his pick-up arm, and found that it had a resistance of 0.45 Ω/m, so a typical 9″ arm requiring 600 mm of wire for each channel (loop to cartridge and back) contributes 0.27 Ω resistance. The 1 m loop of 0.7 mm silver twisted pair from the arm base to the pre-amplifier added 0.12 Ω, bringing the total arm resistance to 0.39 Ω. Pick-up arm wiring modifications are popular, and to reduce the number of soldered joints, some take the fine wire required within the arm tube all the way to the pre-amplifier input plug. 600 mm leads are typical, so this would increase the loop resistance to 0.81 Ω. To put these values into perspective, the previously mentioned Ortofon Quattro cartridge had a specified coil resistance of 3 Ω.

High levels at high frequencies imply high acceleration of the stylus tip. To minimise the force required by the vinyl to accelerate the tip ($F = ma$), cartridge manufacturers strive to minimise stylus tip mass. The most effective way of reducing tip mass is to use a smaller diamond, because so little of the diamond actually constitutes the stylus tip but this makes the diamond harder to grip whilst grinding and lapping, and increases the chance of trapping dust between record surface and cantilever. However, the mass of the coils in a moving coil cartridge is also significant, so a useful improvement can be obtained by reducing wire diameter. Unfortunately, cartridge manufacturers do not always update their technical information as they develop a product, so the measured DC resistance of the coils can be higher than specified. The author recently measured a nominal 6 Ω cartridge, and found its DC resistance to be 10.5 Ω.

The ideal point to measure cartridge DC resistance is at the pre-amplifier input plug because this takes arm wiring resistance into account. The fine wire within a cartridge would be ruptured by passing a high current through it, so the author first measured the current supplied by his meter when set to the lowest resistance range, and found it to be 0.1 mA – quite low enough not to disturb a cartridge. The safest measurement method is to use a component bridge to measure the AC resistance (and inductance).

Once the source resistance seen by the step-up transformer is accurately known, its optimum output loading can be found (see Chapter 4).

Hum Loops and Unbalanced Interfaces

All unbalanced input stages are susceptible to the noise current circulating in a hum loop, but the problem can be reduced even if the loop cannot be broken:

- Always bond the 0 V signal ground to chassis at the input of the RIAA stage's first active device – this is an unbreakable rule.
- $V = IR$ forces the hum voltage to be proportional to the resistance of the 0 V signal conductor, so a thicker conductor reduces hum. Unfortunately, this implies that turntables need heavy earth bond wires that would tend to short-circuit any acoustic suspension they might have. One way of avoiding the problem of finicky cable dressing to a suspended sub-chassis is to take the fine arm wires to the plinth before connecting to the heavier external cable needed to connect to the RIAA stage.
- Stop treating cartridges as unbalanced generators – balanced interfaces are far more forgiving of hum problems.

Balanced Working and Pick-Up Arm Wiring

A balanced source is simply one where each terminal of the source has balanced, or equal, impedances to ground. Frequently, the only path to earth from the terminals is via stray capacitances (no DC path), and the source is then said to be *floating* . Connecting cables for balanced systems therefore require two identical signal wires, or *legs* , to maintain this balance, plus an overall screen. To maintain balance, the input stage of the following amplifier must have its stray impedances carefully balanced to ground and is based on either a differential pair (cheap) or a carefully designed transformer (best, but more expensive).

Professional audio is invariably balanced in order to protect audio signals from external electromagnetic interference. TV studios and stages are particularly harsh environments because microphone cables may unavoidably have to be laid adjacent to lighting cables fed from triac dimmers (which generate copious mains harmonics).

When we immerse a balanced connecting cable in a changing electromagnetic field, an identical noise current is induced into each wire. The series resistance of the cable is the same on each leg, and the shunt capacitances and resistances to ground are also equal, so the noise current develops a voltage of identical amplitude and phase on both legs at the amplifier input. Because this is a common-mode signal, it is rejected by the amplifier, whereas the wanted audio signal is differential mode and is amplified.

A typical moving coil cartridge produces ≈200 μV at 1 kHz 5 cm/s, but before RIAA equalisation the level at 50 Hz is ≈17 dB lower at only 28 μV. Achieving our goal of inaudible hum on a signal at this level is not trivial, and we need all the help that we can get. The cartridge is inherently a balanced device, so why unbalance it?

We should immediately rewire the cable leaving the base of our pick-up arm to maintain this balance by throwing away any coaxial cable and replacing it with a twisted pair having an overall screen for each channel. (A pair of coaxial cables for each channel would not be a good idea because the increased spacing between the inner conductors would cause slightly different noise currents in each leg, greatly reducing cancellation.)

The author’s pick-up arm uses an interconnect from the arm base comprising a twisted pair of 0.7 mm solid core silver threaded down a polytetrafluoroethylene (PTFE) sheath, covered with a braid electrostatic screen. Both cables are then threaded down one overall braid screen, which also serves to hold the cables together. The braid must not have voids, so domestic aerial cable is unsuitable; broadcast quality video cable and multicore umbilical cable are both ideal sources of non-voided braid. Once the plastic outer sheath has been removed, the braid easily concertinas off the inner conductors. Finally, the cable should be sleeved with nylon braid to prevent the noise that would otherwise result from the screen intermittently touching another earthed metal part.

All three screening braids should be firmly bonded to the metalwork at the base of the pick-up arm and a link taken from there to the turntable chassis in order to earth turntables with low voltage motors that would otherwise be floating. At the other end, make sure the interconnect’s screen is connected firmly to the RIAA stage’s chassis.

Turntables using mains motors will/should already have their chassis bonded to earth by their own mains earth safety bond, and the first active device in the RIAA stage will have its 0 V signal ground connected to chassis (also bonded to earth) to minimise hum. Connecting the two earth bonds together via the interconnect’s screen creates a hum loop, but because the balanced audio signal does not pass down the screen, the circulating current in the loop does *not* cause a problem. In the event that the loop does cause a problem, it is the link between arm and turntable chassis that should be broken.

Phono plugs should *not* be used for connecting the audio cable to the pre-amplifier as they are unbalanced connectors, and a ‘professional’ quality metal bodied five-pin DIN plug or XLR is ideal, although its cable entry will almost certainly need to be enlarged. Alternatively, and more clumsily, a pair of three-pin XLRs could be used, but this requires individual (double-screened) cables from the arm base, or a ‘Y’ split in the cable near the pre-amplifier, which is quite difficult to make neatly.

Within the arm tube, most pick-up arms lightly twist all four (thin, non-screened) wires from the cartridge together because this makes the harness easier to handle. Crosstalk between channels, and hum rejection, can be improved by tightly twisting each individual channel pair as it passes down the arm tube, but perhaps reverting to the four-wire twist (often required for low friction) as the wires pass through the bearings to the output cable. Because this modification primarily affects longitudinal currents, it tends to be of more value to pre-amplifiers with balanced inputs, but it is still worthwhile on unbalanced ones. Martin Bastin (of Garrard modification fame) reports that he has been using this method for years.

Bear in mind that the *primary* requirement for pick-up arm internal wiring is low friction, and that pick-up arm wire is fiddly stuff and difficult to strip without damaging the internal conductors, although the specialist adjustable wire strippers having calibrated settings down to 0.25 mm are very effective. Leave well alone unless you really are competent with such fine wire, and beware that it becomes brittle with age.

Balanced wiring is particularly beneficial for moving coil cartridges, and even helps hum rejection when the pre-amplifier is unbalanced.

RIAA Stage Design

The RIAA stage has to satisfy so many contradictory requirements simultaneously that its design and execution is fraught with problems – microphone amplifiers and RIAA stages are the two hardest analogue audio problems.

When we investigated power amplifiers, we looked at some classic designs to see how the goals were achieved. There were *no* classic RIAA stages, they varied from mediocre to plain awful. This was not always due to incompetence on the part of the designers. They had poorer quality components and could not use regulated HT supplies as is habitual today. However, the main factor was that there was simply no incentive to design superb RIAA stages because the signal leaving the turntable was not very good. Vinyl was regarded as a very poor quality source, often requiring low-pass filtering at 8 kHz to reduce dust/click disturbance. Amazingly, good vinyl, turntables and cartridges were all available, but the appalling mechanical/acoustical failings of most of the arms and chassis/plinths meant that the electronics engineers were never exposed to criticism.

Determination of Requirements

We need to define the RIAA stage's detailed requirements before we can begin design:

(1) *Low noise and no hum* : We have to concede that valves are not quite as quiet as the latest generation of low-noise IC op-amps, but DC heater supplies eliminate hum and slightly reduce valve noise. Pentodes are complete non-runners, and we will need to be careful in our use of triodes.

(2) *Constant input resistance and capacitance* : This might seem obvious, but high output moving coil cartridges can be sensitive to changes in electrical loading.

(3) *Accurate RIAA* : The author is delighted to report that wildly inaccurate RIAA is now becoming a rarity on new designs. Nevertheless, it's easy to make a mistake. One mistake is to use polyester (polyethylene terephthalate) capacitors – their capacitance changes with frequency. Polystyrene is the best practical dielectric, but polypropylene is acceptable.

(4) *Low sensitivity to component variation* : Valves wear out, and as they do so, r_a rises. Similarly, when a valve is replaced, the new value of C_{ag} may not be the same as that of the old valve. Neither of these effects should noticeably

affect the accuracy of RIAA equalisation.

(5) *Good overload capability* : But what capability is necessary? Using a Tektronix TDS420 oscilloscope, the maximum output of LPs was investigated in conjunction with a high quality record playing system. The TDS420 was first used in ‘envelope’ mode to find the maximum output of the cartridge and monitored an entire day of listening to music. The largest musical peaks were found whilst playing a Mobile Fidelity pressing of Beethoven’s Ninth Symphony. Before equalisation these peaks rose to +16 dB above the nominal 5 cm/s level, but clicks due to dust or scratches rose to about twice this level at +22 dB (see Figure 7.25).

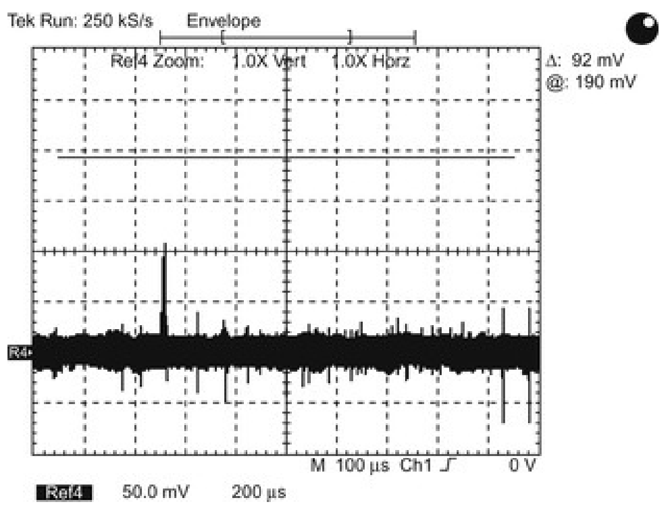


Figure 7.25 Unequalised enveloped music output from cartridge (peaks are dust/clicks).

Individual clicks were then captured, and it was found that the vinyl/tip mass resonance was being excited and that this produced a heavily damped oscillation at 56 kHz for this particular (moving coil) cartridge (see Figure 7.26).

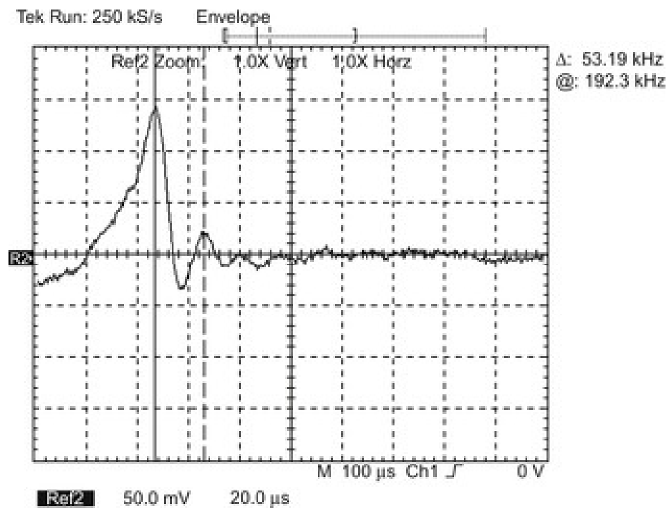


Figure 7.26 Unequalised output from cartridge showing excitation of vinyl/tip mass resonance.

Ultrasonic overload could either generate intermodulation products that come back down into the audio band, or worse, it could cause blocking. Blocking is particularly undesirable because it converts a momentary overload that might have been almost unnoticeable into a protracted low frequency disturbance whose severity is amplified by RIAA equalisation. If a power amplifier blocks, the user can turn the volume down, but this is not possible in an RIAA stage, so blocking must be avoided at all costs.

We should now allow for variable cartridge sensitivity of about 6 dB – if we need more than this, we should reconfigure the RIAA stage.

A good design should not operate permanently at its limits, so a further 6 dB margin is desirable, to give a total headroom of 28 dB in the audio band, *rising* to 34 dB or more at ultrasonic frequencies. Very few RIAA stages of any age achieve this requirement and low noise simultaneously.

Worn/old discs generate more ultrasonic energy than a new disc. This may be due to dirt ground into their groove, or because they were previously played by a cartridge that mistracked, causing wall damage as its stylus flailed helplessly from side to side of the groove. Inadequate ultrasonic overload margin is the reason why a poor RIAA stage exacerbates a worn record's defects, yet a good one renders them palatable.

(6) *Low distortion* : This is an obvious requirement, and is linked to (5).

(7) *Low output resistance* : The RIAA stage should ideally be able to drive the input resistance of a semiconductor recording device such as a computer sound card.

(8) *Low microphony* : Valves are always microphonic, but it is possible to make them worse. A low value of anode load combined with high anode

current reduces noise, but increases the *power* gain of the stage, which causes microphony to rise. Ultimately, we have to isolate the valves mechanically. This is less of a problem than it seems, since most of the structural resonances of the electrodes are above 1 kHz, and a mechanical filter to deal with this can be made by floating the RIAA stage sub-chassis on springs of knicker elastic. (Metal springs tend to ring at the same frequencies as the internal structures of valves.)

Implementing RIAA Equalisation

Now that we know our requirements, we can consider our topology.

The low noise and constant input resistance requirements eliminate shunt feedback. Low noise rules out the pentode. We are therefore left with a combination of triode stages having active equalisation determined by series feedback, or with passive equalisation. Either of these contenders may be further broken down into performing the equalisation 'all in one go', or splitting it over a number of stages.

To see how we need to tackle the problem of RIAA equalisation, we need to define RIAA equalisation. The equalisation is specified in terms of time constants: 3,180 μs, 318 μs and 75 μs. The equation that generates the gain G_s required by the RIAA replay equalisation standard is:

$$G_s = \frac{(1 + 318 \times 10^{-6} \times s)}{(1 + 3.18 \times 10^{-3} \times s)(1 + 75 \times 10^{-6} \times s)}$$

where

$$s = j\omega \quad \text{and} \quad \omega = 2\pi f.$$

This is not a friendly equation, and a spreadsheet is the easiest way of subduing it. The graph shows the required amplitude response of the equaliser *only* a perfectly pre-equalised signal passed through a perfect equaliser would yield an amplitude response deviation of 0 dB and a phase response of 0° at all frequencies (see Figure 7.27).

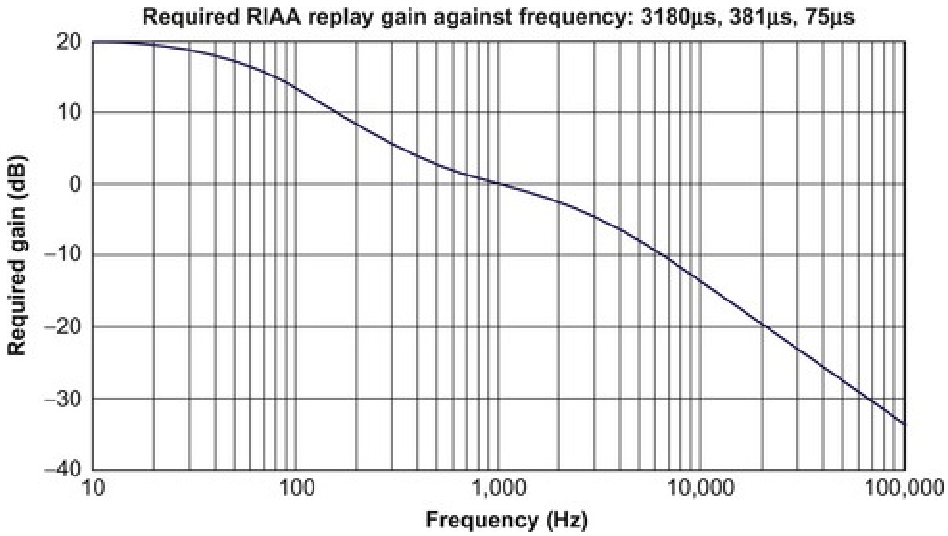


Figure 7.27 Required RIAA replay gain against frequency (includes 3.18µs).

The graph shows that 19.9 dB of gain is needed at low frequencies, whilst high frequency attenuation must continue indefinitely, which apparently excludes the series feedback ‘all in one go’ topology, because its gain cannot fall below unity. However, this failing can be compensated after the feedback amplifier, but it does mean that the response before compensation is rising, which compromises distortion and ultrasonic headroom within the amplifier.

Since the RIAA replay curve appears in almost all discussions about vinyl equalisation, there is a temptation to think that it refers to the vinyl cut, but this is not quite true. The RIAA replay curve also corrects for the fact that virtually all cartridges are magnetic and therefore velocity transducers that need a correction that falls at 6 dB/octave. If we subtract velocity correction from the RIAA replay curve and invert it, we obtain a response that reflects the vinyl cut (see Figure 7.28).

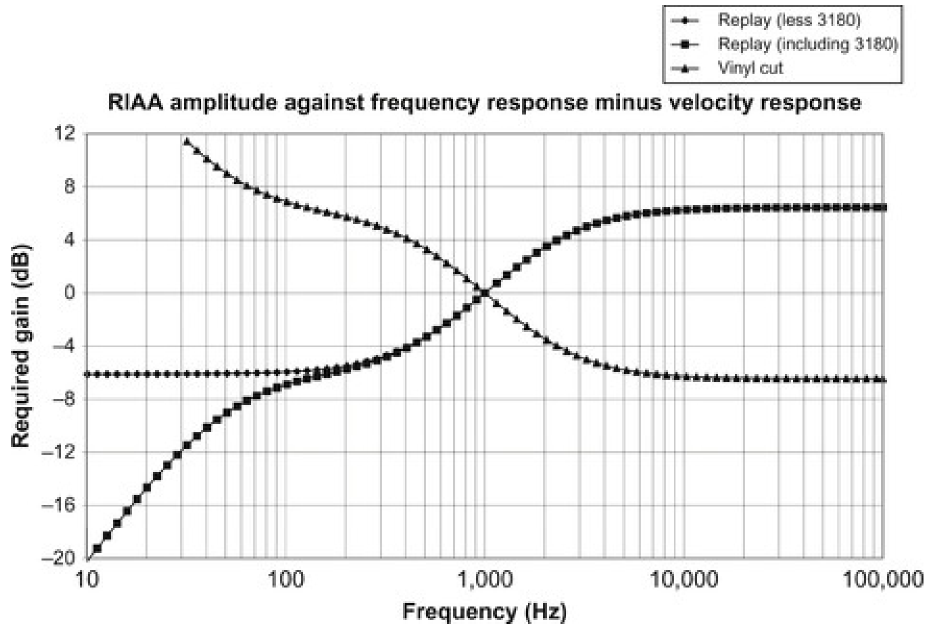


Figure 7.28 RIAA amplitude against frequency response minus velocity response.

Note that the replay (less 3,180 μs) response is constant amplitude except for the 318 μs/75 μs 12.5 dB shelf equaliser that corrects for the recorded equalisation that tamed recording velocities. The significance of the RIAA curve less velocity correction is two-fold:

- If we had a perfect displacement transducer (perhaps based on strain gauge or electrostatic principles), we would require replay equalisation conforming to the replay (including 3,180 μs) curve if we were to match the frequency response of an equalised magnetic cartridge. However, the few strain gauge cartridges that the author has seen have incorporated the 318 μs/75 μs equalisation mechanically, yet do not appear to include 3,180 μs equalisation, effectively giving them significant bass boost below 100 Hz compared to a magnetic cartridge plus electrical RIAA.
- The vinyl cut curve gives the relative amplitudes of the groove widths at differing frequencies and, in principle, a good USB microscope with measurement capabilities supported rigidly just above the surface of a test record should be capable of calibrating it to within 1 dB (remember that 1 dB is 12%, so it's actually quite a sizeable uncertainty even for a pixellated optical measurement).

'All in One Go' Equalisation

Because the 1 kHz level is ≈19.9 dB below the maximum level at low frequency, any ‘all in one go’ passive network *must* also have ≥19.9 dB of loss. In general, an RIAA stage having an acceptable balance of noise and overload capability using this technique requires a high gain input stage. Beware that if the grid-leak resistor of the following valve is across the output of the network, this causes additional attenuation (this can be avoided by moving the grid-leak capacitor to be across the input of the network).

If we should decide to use ‘all in one go’ equalisation, the relevant formulae are given in the definitive JAES paper by Lipshitz [6] .

Of the four possible networks that Lipshitz gives, these reduce to two for passive equalisation. Of these two, only one has a capacitor in parallel with the lower arm of the network. This feature is important because it allows us to account for stray and Miller capacitance and is therefore the only feasible network in a valve pre-amplifier (see Figure 7.29).

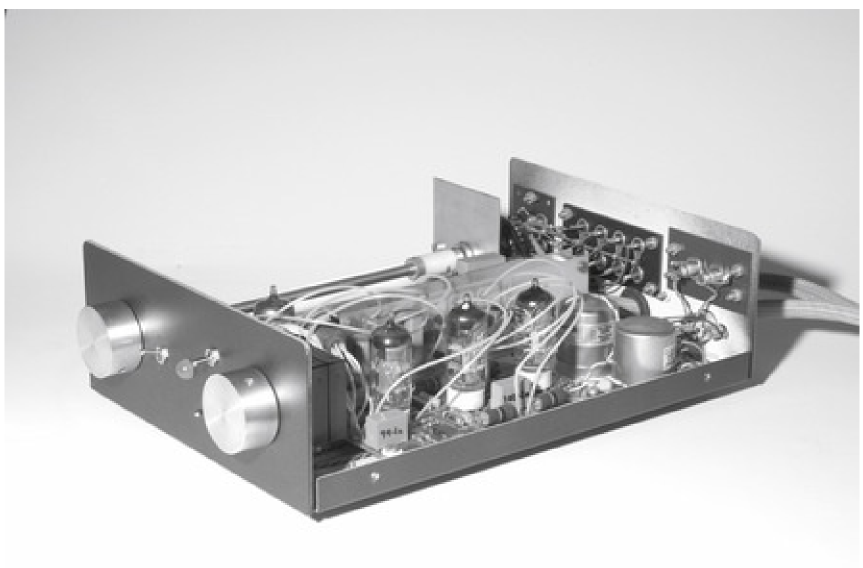


Figure 7.29 Passive RIAA de-emphasis network.

The relevant equations for this *passive* network are:

$$R_1C_1 = 2187\ \mu\text{s}$$
$$R_1C_2 = 750\ \mu\text{s}$$
$$R_2C_1 = 318\ \mu\text{s}$$
$$\frac{C_1}{C_2} = 2.916$$

These numbers are exact and have not been rounded.

Remember that any grid-leak resistor in parallel with the lower arm of the network, or non-zero output resistance of the driving stage, changes the effective value of R_1 as seen by the network. Therefore, the values for the network must be calculated using the Thévenin resistance seen by that network.

Likewise, any stray, or Miller, capacitance must be subtracted from the calculated value of C_2.

For any 'all in one go' topology other than the above network, it is essential to refer to the Lipshitz paper, and read it thoroughly before embarking on design.

Split RIAA Equalisation

We are now left with only two possibilities for equalisation, split active and split passive, so we must define how to split the equalisation. Fortunately, there is only one rational way to split the equalisation, and that is to pair the 3,180 μs with the 318 μs, but to perform the 75 μs separately.

The 75 μs time constant defines a low-pass filter whose $f_{-3\ \mathrm{dB}} \approx 2{,}122$ Hz and rolls off at 6 dB/octave thereafter. This would be an ideal filter to use early in the pre-amplifier since it allows HF overload capability after that stage to rise at 6 dB/octave above cut-off, which is *exactly* what we need with a magnetic cartridge.

The 75 μs time constant can be implemented passively following the input stage, which has the advantage of ensuring that the load seen by the cartridge is constant with frequency.

Moving-magnet cartridges often use the load capacitance in conjunction with the generator's self-inductance to form a resonant equaliser that corrects the falling mechanical response of the cartridge. Thus, the value of load capacitance is critical, but this can be set quickly and easily by adding a twin-gang ≈300 pF air-spaced variable capacitor salvaged from a (probably valve) medium wave radio (see Figure 7.30).

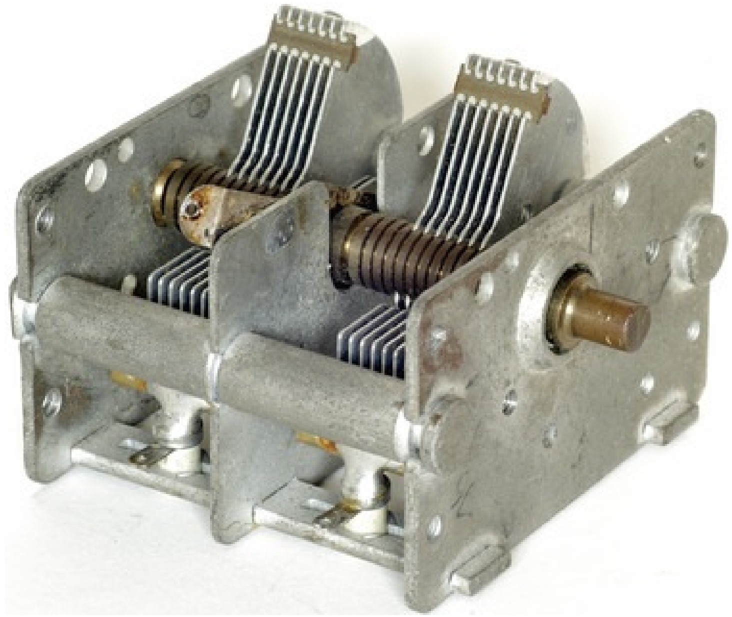


Figure 7.30 Typical variable 300pF air-spaced capacitor.

The main reason for the choice of a passive 75 μs RIAA equalisation network is that a series feedback amplifier cannot make A_v <1, and a shunt feedback amplifier would have noise problems. Additionally, although it was not noted earlier, the output stage of a feedback amplifier attempting this response faces a heavy capacitive load. Briefly, the capacitive load demands a large current at HF and is equivalent to changing the AC loadline to a far lower value of load resistance, which results in additional distortion *before* closing the feedback loop, although this is not a problem with the better IC op-amps.

The 3,180 μs, 318 μs pairing defines a shelf response with a level variation of exactly 20 dB. Using IC op-amps it is equally convenient to perform this actively or passively, but with valves it is more convenient to use passive equalisation.

The Final Choice

Feedback RIAA is best suited to semiconductor op-amps because their extremely high open-loop gain and falling response with frequency allow a fairly constant (and high) amount of feedback at all frequencies. Passive equalisation is best for valves, and the choice between 'all in one go' and split equalisation depends on the signal amplitude leaving the first stage. As a useful rough guide, the decision point tends to be around 200 mV $_{RMS}$ at 1 kHz 5 cm/s – below this amplitude the best balance between second stage noise and distortion is likely to be obtained using split RIAA (passive 75 μs at the input to the second stage, passive paired 3,180 μs, 318 μs at the output), whereas above 200 mV $_{RMS}$ ‘all in one go’ at the second stage’s input is likely to be superior.

A Simplified Example RIAA Stage

Rather than going straight to a detailed practical design, we will introduce and investigate some generic problems using a simplified example having split equalisation because this will allow background understanding before delving into the more complex and detailed issues of practical designs.

Noise and Input Capacitance of the Input Stage

Bearing in mind our low-noise requirement, the first stage is the *crucial* stage and must have low noise above almost every other requirement. This is reasonable because even +34 dB ref. 5 mV is only 700 $mV_{pk\text{–}pk}$, so linearity ought to be a minor problem.

Designing for low noise usually means wringing the utmost gain out of the first stage such that noise considerations in succeeding stages are irrelevant. This would imply a common cathode stage using a high-μ triode such as the ECC83 or ECC808 (electrically almost identical, but ECC808 has lower hum and noise, and completely different pin-out), but with a typical gain $A_v = 70$ this would result in an input capacitance of ≈120 pF including strays.

Most moving magnet cartridges are designed to be loaded by a specific capacitance, and older Shures and Ortofons needed 400–500 pF, but one lasting legacy of CD4 (quadraphonic sound with the rear channels on a supersonic carrier) is that modern cartridges tend to expect 250 pF. Once we include pick-up arm wiring capacitance and connecting cable capacitance to the 120 pF contributed by the ECC83, the loading capacitance seen by the cartridge could rise to 300 pF. The ECC83 is probably now out of the running, unless we are prepared to rewire the arm (which might not be such a bad idea), and we are back to the E88CC with a lower gain and lower shunt capacitance.

Although high-μ Loctal valves such as the 7F7 are a possibility, high-μ octal-based triodes are almost certainly forbidden because of their excessive C_{ag}. The 6SL7GT, which with $\mu = 70$ is the predecessor of the ECC83, has $C_{ag} \approx 2.8$ pF (RCA value: depends on manufacturer/source). With a typical gain of 50, this would result in an input capacitance, including strays, of 160 pF. This is now uncomfortably close to our 250 pF limit, and once arm wiring capacitance is included, could only be achieved by mounting the entire RIAA stage directly below the pick-up arm mounting so that the internal wires of the arm connected directly to the grid.

Mounting the RIAA stage onto the plinth, directly below the arm mounting, has enormous advantages in terms of input capacitance, rejection of induced noise and microphony. It also makes the turntable completely non-standard, and may not even be physically possible due to limited space or limited weight-carrying ability. A delicately suspended sub-chassis turntable will not take kindly to a pound or two (500–1,000 g) of pre-amplifier hanging from the arm mounting. Conversely, a turntable such as the Garrard 301 that *must* be directly mounted onto a very heavy plinth would scarcely notice the extra mass.

The E88CC has an additional advantage in that r_a is low and, as we will soon see, this helps noise performance. Additionally, a low r_a forms a small proportion of the total resistance that defines the 75 µs roll-off, which then satisfies our earlier requirement of reduced sensitivity to component ageing and changes.

Noise in the input stage is determined not only by the valve, but also by the associated resistors, of which R_L is by far the most important (see Figure 7.31 a).

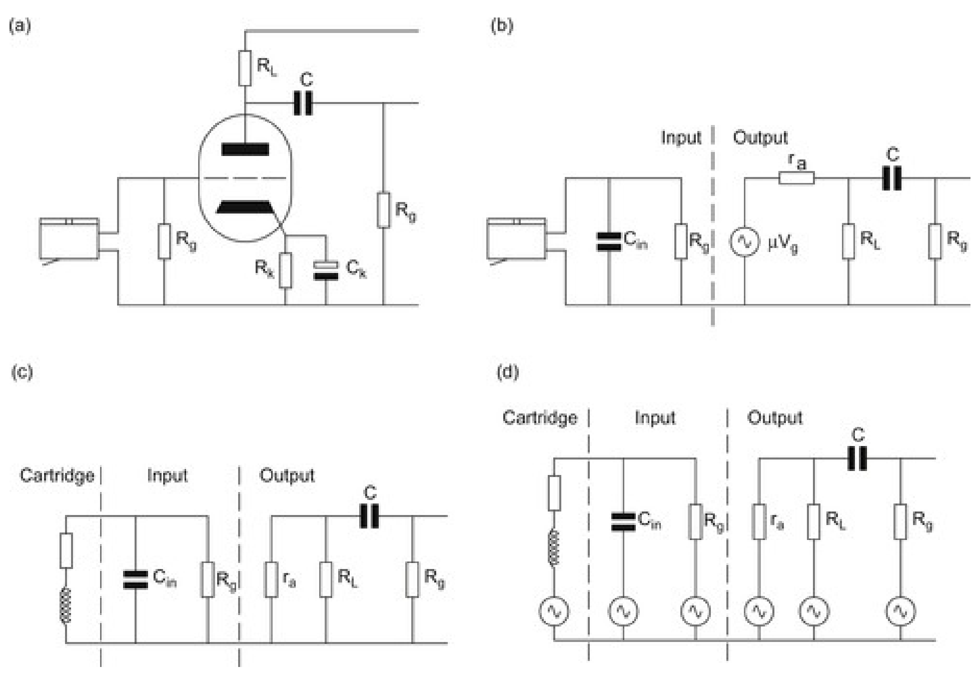


Figure 7.31 Noise in the input stage.

To be able to calculate the noise performance of the stage, we need to redraw the circuit as a simple equivalent circuit, which makes analysis easier (see Figure 7.31 b).

We have replaced the output of the valve with a perfect Thévenin voltage source, and r_a has been included. A moving magnet cartridge can be represented as a resistor in series with an inductor, and since a Thévenin source has zero resistance, we could replace it with a short circuit, and redraw the circuit yet again (see Figure 7.31 c).

We are now in a position to add some noise sources to our equivalent circuit (see Figure 7.31d).

The derivation of this final equivalent circuit was taken in many steps because the final circuit bears very little resemblance to the original circuit. Before embarking on complex calculations, we can make some important, and useful, observations.

All of the noise sources (with their associated resistances) after the valve are in parallel, so a source of zero resistance will short-circuit *any* other source, provided that there is no additional series resistance. Modern designs aim for $R_g \approx 100\ r_a$, and $R_L \approx 10\ r_a$, so r_a tends to shunt these other sources. This should make the contribution of R_g insignificant so that any convenient value of R_g could be used, but the series coupling capacitor reduces the shunting effect of r_a. The reactance of this capacitor is:

$$X_c = \frac{1}{2\pi fC}$$

For a typical grid-leak of 1 MΩ, we might use a coupling capacitor of 100 nF to give a −3 dB frequency of 1.6 Hz. If we assume that the lowest noise frequency of interest is 20 Hz (and this is debatable), then we find that at 20 Hz, $X_C = 80$ kΩ. This is such a high value that it nullifies any possible shunting effect by r_a, until X_c falls to a value lower than r_a.

The result of this is that the usual choice of coupling capacitor does *not* allow r_a to shunt the noise from the grid-leak resistor at frequencies below 1 kHz. The resistor therefore produces noise whose amplitude is inversely proportional to frequency (1/ *f* noise), but that rises to the maximum theoretical thermal noise for that value of resistor ($v_n = \sqrt{4kTBR}$).

To prevent this excess noise, we might decide to use a value of coupling capacitor sufficiently large that r_a is able to shunt R_g at *all* frequencies, which would require a value ≈10 μF. This is a large capacitor, and DC coupling is preferable if possible, but the technique has been used in a number of commercial RIAA stages.

Assuming that we have dealt with the grid-leak resistor and the coupling capacitor, we are left with the anode load resistor R_L, and the valve itself, which leaves us with a simple equivalent circuit (see Figure 7.32).

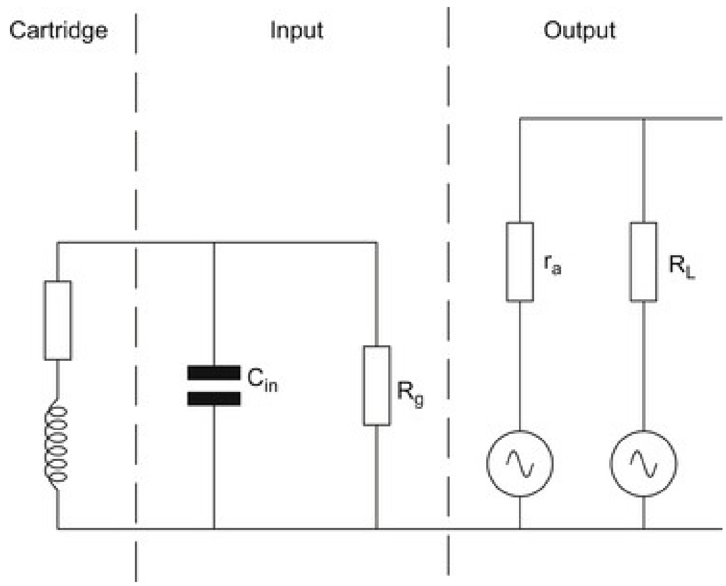


Figure 7.32 Final equivalent circuit for noise sources in the input stage.

R_L generates thermal noise, and unless it is a wirewound resistor, it also generates *excess* noise. Excess noise is generally specified by resistor manufacturers in terms of μV/V of applied DC. We will therefore investigate a typical stage (see Figure 7.33).

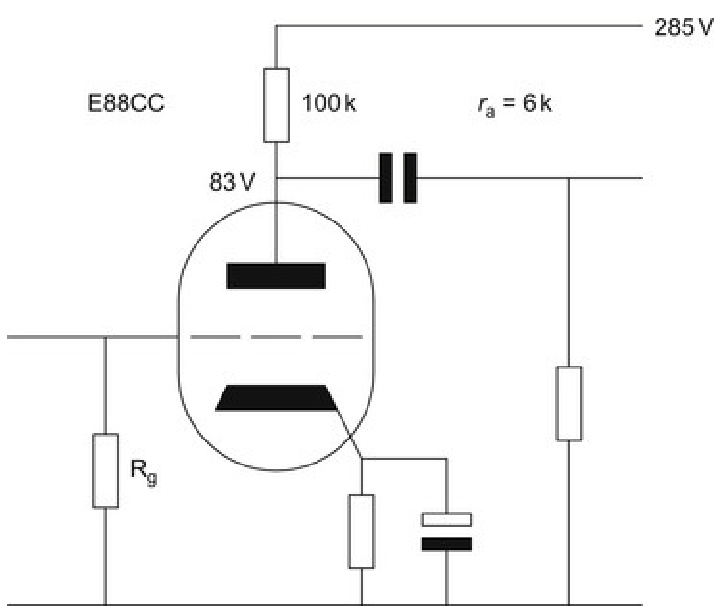


Figure 7.33 Typical input stage for noise analysis.

The DC voltage across $R_L \approx 200$ V. A typical 100 kΩ 2 W metal film resistor generates 0.1 μV/V of excess noise, so 20 μV would be generated in this circuit. The thermal noise of a resistor is given by:

$$v_n = \sqrt{4kTBR}$$

where

k =Boltzmann's constant ≈1.381×10^{-23} J/K

T =absolute temperature in K=°C+273.16

B =bandwidth in Hz

R =resistance in Ω.

For a typical internal temperature of 40°C (313 K), with a bandwidth of 20 kHz, this is more conveniently expressed as:

$$v_n = 1.86 \times 10^{-8}\sqrt{R}$$

Using this equation, we find that a perfect 100 kΩ resistor generates 5.9 μV of thermal noise. In this instance, the resistor's thermal noise has been greatly exceeded by its excess noise. To find the total noise of the resistor, we must add the individual noise *powers* , which, if we remember that $P = V^2 / R$, means that

$$v_{noise(total)} = \sqrt{v_1^2 + v_2^2 + v_n^2 \cdots}$$

This gives a total noise for the resistor of 21 μV, and was rather tedious, but it demonstrates two points:

- For wirewound resistors we need only calculate the thermal noise. (No excess noise.)
- For metal film resistors we need only calculate the excess noise. (This simplification works because in practical circuits, as the DC voltage across the anode load resistor falls, so does its required value, and therefore its thermal noise.)

Now that we have simplified the noise sources in the resistor, we can see how they will be shunted by the r_a of the valve, and redraw the circuit (see Figure 7.34).

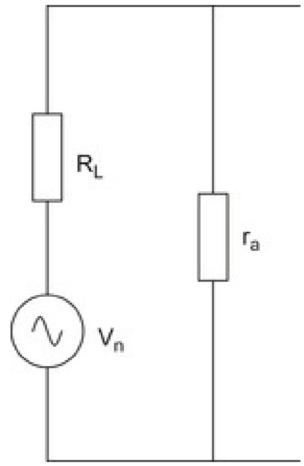


Figure 7.34 Effect of r_a on noise produced by R_L .

It is now easy to see that the circuit is a potential divider and that the actual contribution of resistor noise to the circuit is equal to the open circuit resistor noise multiplied by the attenuation of the potential divider. In our example, this reduces the noise of the resistor from 21 μV to 1.26 μV. It should be noted that if R_k is left unbypassed, r_a rises dramatically and it is no longer able to shunt resistor noise.

If we divide the noise voltage by the gain of the stage A_v =29, we can find the *input referred noise* , which is 43 nV. The significance of this is that it enables us to sum this noise with any noise sources at the grid, such as the grid-leak resistor. In practice, if we calculate the thermal noise generated by R_g and its attenuation by the cartridge, we generally find that it is insignificant compared to the valve noise. In any event, we do not have a choice about R_g since it is set by the cartridge.

Valve Noise

We should now consider the shot noise within the valve itself. Valves produce shot noise because I_a is made up of individual electrons that shower the anode, and also because electrons leave the cathode with random velocities to join the space charge, so this implies that cathode chemistry could affect noise.

For triodes:

$$r_{eq} \geq \frac{2.5}{g_m}$$

This says that the white shot noise generated within the valve is equivalent to the thermal (white) noise generated by a perfect resistor r_{eq} at the input of the valve.

For our example triode, g_m ≈5.3 mA/V, so the equivalent noise resistance would be 470 Ω.

Using $v_n = 1.86 \times 10^{-8}\sqrt{R}$, the input voltage noise produced by the valve is therefore ≈400 nV and swamps the 43 nV (input referred) noise produced by the anode load resistor (as it should, in a good design), and we need not sum the noise powers of the valve and the resistor.

For pentodes [7] :

$$r_{eq} \geq \frac{I_a}{I_a + I_{g_2}} \cdot \left(\frac{2.5}{g_m} + \frac{20 I_{g_2}}{g_m^2}\right)$$

Applying this equation to the low-noise EF86 pentode operating at I_a =1.25 mA, $I_{g_2} = 0.3\ \mathrm{mA}$ predicts a noise resistance of 3.9 kΩ and a noise voltage (20 kHz bandwidth) of 1.2 µV. However, Mullard measured 2 µV for a noise bandwidth of 25 Hz to 10 kHz under the same DC conditions, which corresponds to 2.8 µV for a 20 kHz bandwidth, making it 7.4 dB noisier than the prediction.

For JFETs [8] :

$$r_{eq} \geq \frac{0.7}{g_m}$$

JFETs are common in hybrid valve/transistor circuits, and as the previous equation shows, they are quiet. Their equivalent noise resistance is not only substantially lower than for a triode (effectively 5.5 dB quieter for a given g_m), but they also tend to have higher g_m for a given anode or drain current. The quietest JFETs currently available are the Linear Systems LSK170 (derived from the Toshiba 2SK170) and Philips BF862 – both produce ≈1 nV/√Hz noise, making them entirely suitable for moving magnet cartridges, but not quite quiet enough for moving coils.

1/ f Noise

Unfortunately, the preceding noise equations do not tell the whole story at audio frequencies because they do not account for 1/ *f* noise, but they do indicate that pentodes are much noisier than triodes and that g_m should be maximised. Provided that the cathode has been made properly, 1/ *f* noise is due to grid current noise and, as we saw in Chapter 3 , this is largely down to manufacturing cleanliness, making it sample dependent. However, the clean room conditions needed to manufacture semiconductors mean that manufacturers of low-noise Bipolar Junction Transistors (BJTs) are able to provide graphs that show how noise varies with frequency. The lower the $1/f$ corner frequency, the better it will be.

Connecting Devices in Parallel to Reduce noise

High-g_m valves are so expensive that we might choose to increase g_m by connecting a number of devices in parallel, since the noise falls by a factor of $\sqrt{n}$. The MAT02/LM394 supermatch transistor is an extreme example of this technique, as it contains a pair of composite transistors each made of 100 individual devices to give a 20 dB improvement. Paralleling 100 E88CCs is impractical, but a worthwhile, if somewhat modest, improvement of 4.5 dB can be gained by using three devices in parallel. Note that the input capacitance trebles, outlawing input transformers, moving magnet cartridges, and even some moving coil cartridges. (The high output moving coil Sumiko 'Blue Point Special' specifies maximum load capacitance as 200 pF.)

Unfortunately, the previous examples demonstrate an important point. Although we may improve noise by a better choice of input valve, or valves, we pay dearly for quite small improvements, since obtaining a high g_m is expensive and invariably current hungry. To minimise noise, it is always better to present the input stage with a healthy signal, rather than hope to amplify a weak one cleanly.

Valve Noise Summary

Despite all the previous caveats, qualifications and provisos, we *can* make some useful generalisations that will avoid unnecessary calculations when designing for low noise:

- Pentodes are significantly noisier than triodes, and JFETs can be even quieter.
- Valve sample variation can be large. ($1/f$ noise is largely determined by the cleanliness of the room in which the valve was assembled, so although a given manufacturer tends to be consistent, there can be differences between manufacturers – or more accurately, their assembly rooms.)
- To render the noise of R_L insignificant, there must be no feedback at the cathode, since this reduces the shunting effect of r_a. This is also true for a μ-follower, even though omitting C_k has no discernible effect on gain. The cascode has $r_a \approx \infty$, so the noise from R_L must be considered.
- Maximise g_m for low noise, either with a single excellent device, or with a

number of lesser devices in parallel.

- Maximised g_m invariably raises the input capacitance of the input stage and may preclude using a moving-coil step-up transformer.
- Excess noise dominates in film resistors passing DC. Wirewound and bulk foil resistors do not produce excess noise.
- A very large (typically 100 times normal) coupling capacitor allows r_a to shunt the noise generated by the grid-leak resistor of the following stage, but DC coupling would be even better.

Together, these noise and input capacitance considerations all but eliminate the ECC83, 6SL7GT, and other high-μ, low-g_m valves from the input stage of an RIAA stage.

Noise Advantage due to RIAA Equalisation

Numerical integration of RIAA equalisation (3,180 μs, 318 μs and 75 μs) from 20 Hz to 20 kHz using thirtieth octave bands gives a noise equivalent bandwidth of 94 Hz, which reduces noise by 22.3 dB, but because the equalisation imposes a gain of 19.9 dB referred to 1 kHz, the final advantage due to equalisation referred to the 1 kHz sensitivity is a meagre 3.4 dB (unchanged if the 3.18 μs time constant is implemented). The significance of this 3.4 dB figure is that if the input referred noise and input sensitivity is known, a full post-RIAA signal-to-noise ratio may be predicted.

Stray Capacitances

We now know that our input valve is likely to be a high-g_m valve such as an E88CC, or better, and if we assume a required input sensitivity of 2.5 mV_{RMS} at 1 kHz 5 cm/s, this is likely to result in a signal of ≈65 mV_{RMS} leaving the first stage, requiring split equalisation and necessitating three stages. The second stage can be similar to the first, but the third probably needs to be a follower for reasons that will become apparent shortly. We can now draw a circuit diagram for our example RIAA stage (see Figure 7.35).

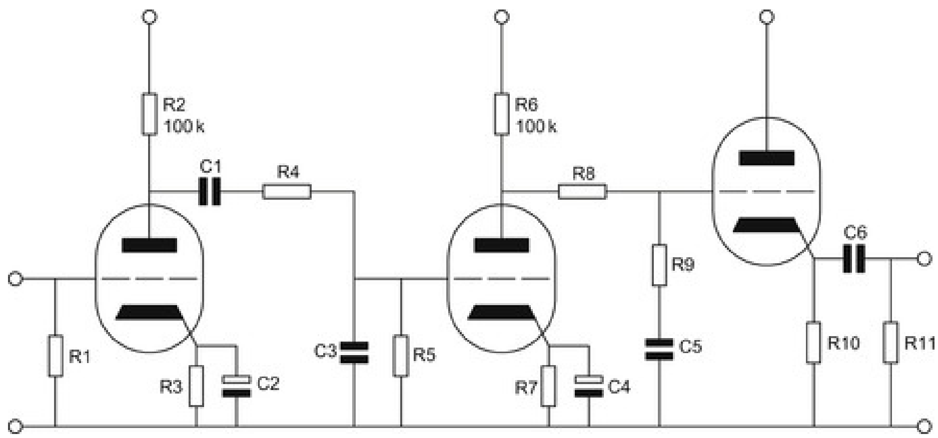


Figure 7.35 Example split equalisation RIAA stage.

The 75 μs high frequency loss is formed by the combination of R_4, R_5 and C_3, whereas the 3,180 μs, 318 μs pairing is formed by R_8, R_9 and C_5. The calculation of these components is simple, but we must remember to account for hidden components such as the output resistance of the preceding valve and the Miller input capacitance of the following valve in parallel with strays.

Calculation of Component Values for 75 μs

We first note that the first and second gain stages are identical, so any calculation applied to one also applies to the other. For the DC conditions chosen for our common cathode triode input stage, $r_a = 6$ kΩ; this is in parallel with the 100 kΩ anode load resistor, so $r_{out} = 5.66$ kΩ.

The gain of the second stage is 29, and $C_{ag} = 1.4$ pF, so the Miller capacitance will be 30×1.4 pF$=42$ pF. In addition to this, the cathode, the heaters and the screen are at earth potential, and are in parallel with this capacitance. $C_{g-k+h+s} = 3.3$ pF, and we ought to allow a few pF for external strays. A total input capacitance of 50 pF would be about right.

To calculate the capacitance needed for the 75 μs time constant, we need to find the total Thévenin resistance that the capacitor sees in parallel (see Figure 7.36).

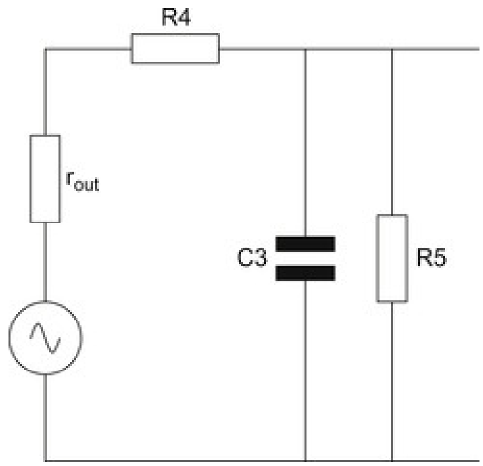


Figure 7.36 Determining 75μs RIAA values.

For the moment, we will ignore C_1, but this will be accounted for later. C_3 sees the grid-leak resistor R_5 in parallel with the series combination of the output resistance of the preceding valve and R_4. As is usual, we will make the grid-leak as large as permissible, so $R_5=1\ \mathrm{M\Omega}$.

We are now free to choose the value of R_4. We need r_{out} to be a small proportion of R_4, otherwise variations in r_a will upset the accuracy of the equalisation, but too large a value of R_4 would form an unnecessarily lossy potential divider in combination with R_5. At high frequency, the capacitor C_3 is a short circuit, and so the additional AC load on the input valve will be R_4. 200 kΩ is a good value for R_4, and it has the bonus of being available both in 0.1% E96 series and 1% E24 series (very few E24 values are common to the E96 series). In combination with R_5, this gives an acceptable loss of 1.6 dB whilst not being an unduly onerous load for the input stage.

The capacitor now sees 200 kΩ+ 5.66 kΩ in parallel with 1 MΩ, which gives a total resistance of 170.58 kΩ. Dividing this value into 75 μs gives the total value of capacitance required, 440 pF. But this network is loaded by the second stage which already has 50 pF of input capacitance from grid to ground, so the *actual* capacitance that we need is 440 pF−50 pF=390 pF, so a 390 pF 1% capacitor would be fine.

We ignored the effect of the coupling capacitor C_1, but this must have some effect on the Thévenin resistance seen by the capacitor. We could use a sufficiently large value to make its reactance negligible compared to the 200 kΩ series resistor, but a more elegant method is to move its position slightly (see Figure 7.37).

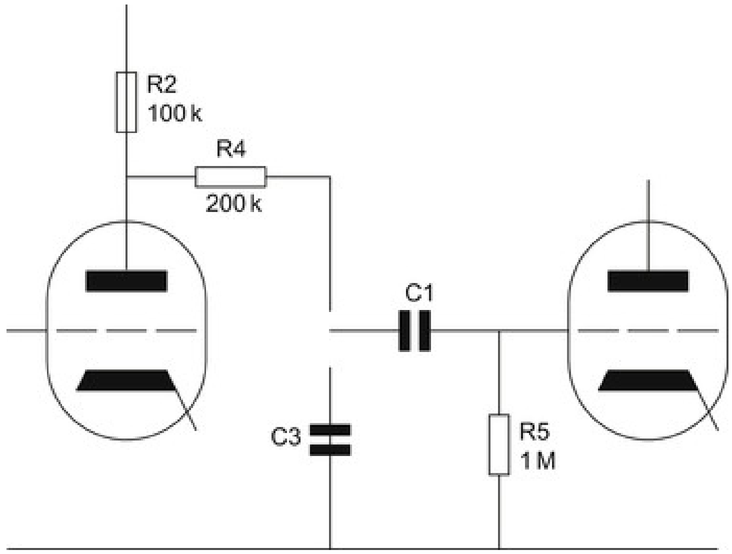


Figure 7.37 Moving the coupling capacitor to reduce interaction.

The capacitor now only has to be negligible compared to 1 MΩ. 75 μs corresponds to a −3 dB point of ≈2 kHz, so it is at this frequency that the values of other components are critical. At 2 kHz, a 100 nF capacitor has a reactance of ≈800 Ω, which is less than 0.1% of 1 MΩ. If we had not moved the capacitor, we would have needed a value >470 nF simply to avoid compromising RIAA accuracy.

Conversely, there is little point in using a very large coupling capacitor in an effort to reduce noise at LFs, since the 200 kΩ series resistance of R_4 swamps the output resistance of the input valve and nullifies its shunting effect on the grid-leak of the second valve.

180 μs, 318 μs Equalisation and the Problem of Interaction

The second stage is direct coupled to a cathode follower in order to eliminate interaction between any coupling capacitor and the 3,180 μs, 318 μs pairing. 3,180 μs corresponds to $f_{-3\ \mathrm{dB}}$ =50 Hz, which is far too close to the typical 1.6 Hz cut-off resulting from 100 nF to 1 MΩ, so they would interact significantly. The other reason for using a cathode follower is its low input capacitance, which causes an additional high frequency roll-off when placed in parallel with the 3,180 μs, 318 μs pairing. In the 75 μs network, we were able to incorporate the value of stray capacitance into our calculations, but in this instance it is not possible, so it is essential that any stray capacitance is so small that it can be ignored. The full equation for the input capacitance of a cathode follower is:

$$C_{\mathrm{input}} = C_{\mathrm{ag}} + (1 - A_v) \cdot C_{\mathrm{gk}}$$

To a good approximation, $A_v = \mu/(\mu+1)$, so for an E88CC ($\mu \approx 32$), $A_v = 0.97$, $C_{ag} = 1.4$ pF and $C_{gk} = 3.3$ pF. The C_{gk} term is thus entirely negligible at 0.1 pF, so the input capacitance is virtually independent of gain at ≈8 pF, including a healthy allowance for strays to ground.

3180 μs and 318 μs Equalisation

The equations that govern the 3,180 μs, 318 μs pairing are delightfully simple, $CR=318\times10^{-6}$, and the upper resistor=9 R, whilst the loss at 1 kHz for this network is 19.05 dB (see Figure 7.38).

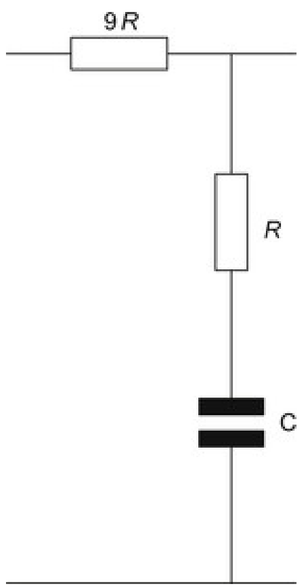


Figure 7.38 Implementing the 3180μs, 318μs RIAA pairing.

We should now check whether our worst case 8 pF shunt capacitance is sufficiently small not to cause a problem. To do this, we need to employ a slightly circular argument.

We first say that it will *not* cause any interaction. If this is true, then the frequency at which the cut-off occurs will be so high that C in the network is a short circuit. If it is a short circuit, we can replace it with a short circuit, and calculate the new Thévenin output resistance of the network. Since the ratio of the resistors is 9:1, the potential divider must have a loss of 10:1, and the output resistance is therefore one-tenth of the upper resistor. If we assume that our upper resistor will again be 200 kΩ (neglecting r_{out} of the previous stage), the Thévenin resistance that the stray capacitance sees at high frequency is 20 kΩ; combined with 8 pF, this gives an a high frequency cut-off of 1 MHz.

As a rule of thumb, once the ratio of two interactive time constants is ≥100:1, the response error caused by interaction is inversely proportional to that ratio, so a ratio of 100:1 causes an error of ≈0.1 dB.

In our example, the ratio of 1 MHz to the nearest time constant of 318 μs (500.5 Hz) is 2,000:1, so we can safely ignore interaction and go on to calculate accurately the values for the 3,180 μs, 318 μs pairing.

If we were driving the network from a source of zero resistance, ideal values for the resistors would be 180 kΩ and 20 kΩ (perfect 9:1 ratio), since both of these are members of the E24 series, and the capacitor could then be 16 nF with only 0.6% design error. Unfortunately, our source has appreciable output resistance, so we will choose 200 kΩ as the upper resistor and must accept whatever values this generates for the lower two components.

The second stage is identical to the input stage, so output resistance is 5.66 kΩ, making a total upper resistance of 205.66 kΩ. The lower resistor will therefore be 22.85 kΩ, and the capacitor 13.92 nF. 22.85 kΩ can be made out of a 23k2 0.1% resistor in parallel with a 1M5 1%. 13.92 nF can be inconveniently made out of a pair of 6n8s in parallel with a 330 pF, or 10n in parallel with 3n9 and 20 pF. We can now draw a full diagram of the example RIAA stage with component values (see Figure 7.39).

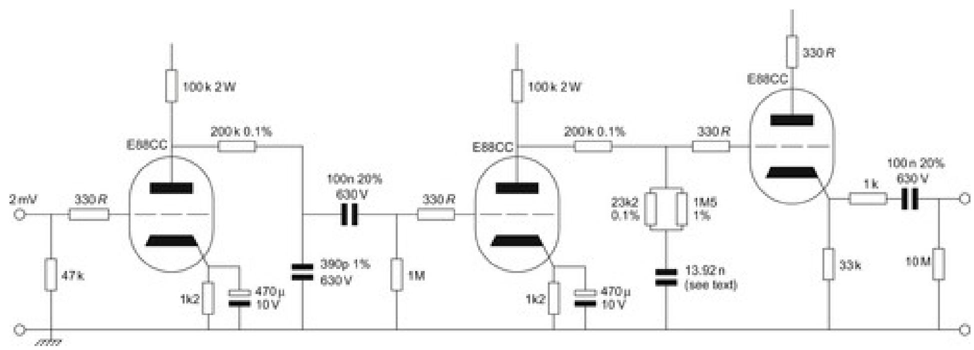


Figure 7.39 Final circuit with component values.

Awkward Values and Tolerances

Equalisation networks and filters *invariably* generate awkward component values, and much manoeuvring is required to nudge them onto the E24 series. Sadly, this is usually wasted effort, since, although 0.1% resistors are readily available, capacitors are only readily available in 1%, and often only in E6 values. Therefore, for best accuracy, we measure the value of the largest capacitors on a precision component bridge (or perhaps a digital multimeter having an *accurate* capacitance range), and add an additional capacitor to achieve the required value.

For the 13.92 nF capacitor needed earlier, we might measure the 6n8 capacitors, and find that they were actually 6.74 nF, so we would actually need a 430 pF, rather than 330 pF. This is not a problem, but suppose we had chosen the more obvious 10 nF//3n9 option, but when the 10 nF 1% capacitor was measured, it was found to be 10.1 nF. We can hardly file a bit off the end!

Close tolerance components are expensive, but they are not always necessary. If we combine a close tolerance component with a looser tolerance component, then the resulting component will *still* be close tolerance, *provided* that the ratio of the values is greater than the ratio of the tolerances. Clearly, the close tolerance component must be the main component, whilst the trimming component can be looser tolerance. As an example, if we need a 22.85 kΩ resistor to a close tolerance, we could choose 23k2 0.1%, and parallel it with 1M5 1%. 1.5 MΩ/23.2 kΩ=65:1, and is greater than the 10:1 ratio of the tolerances, so this combination will be fine. Similarly, for the 13.92 nF capacitor needed earlier, the ratio of the main component to the trimming component is 16:1, so even a 430 pF 10% would be fine. We could probably only buy a 1% component, so there is no need to measure it.

Just because we have adjusted component values on test to meet our exact required value *does not* mean that we now have zero tolerance components. Real components drift with time and temperature, so the values will change. What we have done is to remove the initial error so that the practical value equals the calculated value, which places us in a better starting position for overall tolerance due to drift.

The EC8010 RIAA Stage

Investigating the example stage introduced the two main problems of noise and achieving accurate RIAA in the face of real-world gain stages, but gave only passing consideration to distortion. This single-ended split equalisation RIAA stage uses μ -followers to minimise distortion and transformer coupling from the moving-coil cartridge to the input stage to minimise noise and interference.

The Input Stage

The overriding requirement of the input stage is that it should produce low noise, requiring high g_m, so Table 7.5 sorts valve groups by g_m.

Table 7.5 Comparison of Triode g_m

Type	**Achievable g_m (mA/V)**
E810F (triode connected), EC8020	≈50
3A/167M, 437A	≈42
D3a	≈34
EC8010, 5842, 417A	≈20
EC86, PC86, EC88, PC88	≈11
ECC88/6DJ8, E88CC/6922	≈8

The values given in the table are somewhat lower than manufacturers' quoted values because they reflect usable values that can be achieved in a real design. As a very rough rule of thumb, valves designed for high g_m typically achieve a mutual conductance of between one to one-and-a-half times their anode current. In other words, the E810F requires $I_a \approx 35$ mA to achieve $g_m \approx 50$ mA/V, making it expensive to use, so the choice narrowed to the $g_m \approx 20$ family.

Having chosen the valve family, we must choose I_a. Since $g_m \propto I_a$, we set I_a as high as is practical, so the author chose to operate the valve at $I_a \approx 15$ mA because this current attains most of the achievable g_m. We next need to choose V_{gk}. Many designs have set $V_{gk} < 1$ V, but when the author investigated the distortion spectrum of a 5842 whilst sweeping V_{gk}, he found that if $V_{gk} < 1.3$ V, tiny changes in bias completely changed the distortion spectrum. Once $V_{gk} > 1.5$ V, the higher harmonics subsided and became stable, so the valve was biassed by a cheap red LED (setting $V_{gk} \approx 1.7$ V), which set $V_a = 126$ V for $I_a = 15$ mA.

The value of the anode load R_L can now be chosen. Theoretically, a high value of R_L increases self-noise ($v_n = \sqrt{4kTBR}$), but as this is mostly attenuated by the potential divider formed by r_a and R_L, changing the value over an extreme range only changes the final S/N ratio by ≈1 dB. The factor that determines R_L is the available HT voltage. To have a sufficiently large HT dropping resistor to allow adequate HT smoothing, we should keep the first stage HT<300 V, and since 126 V is dropped across the valve, 174 V is available for R_L, so Ohm's law determines that $R_L \leq 11.6$ kΩ. R_L dissipates significant power in this stage, and because a wirewound type is necessary to eliminate excess noise, the nearest E6 value of 10 kΩ was chosen. (Wirewound resistors are commonly available in E6 values only.)

We now know R_L, and the current through it, so we can determine the precise HT voltage. The resistor drops 150 V, and V_a =126 V, so we need 276 V of HT for the input stage (see Figure 7.40).

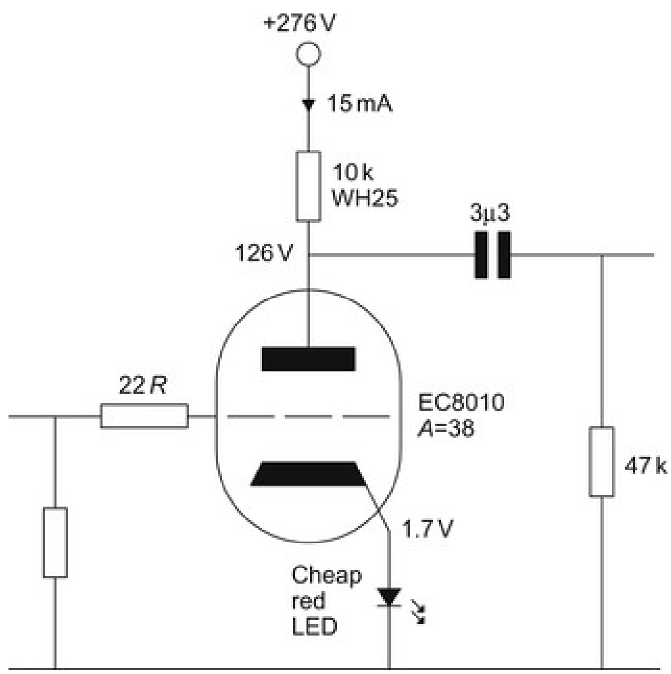


Figure 7.40 EC8010 input stage with LED bias.

Once the design of the input stage had been set, it could be tested for distortion. The circuit was tested at an output of +18 dBu, which lifted the distortion harmonics clear of the noise floor but was well below clipping. Twenty-six samples were tested from the EC8010, 5842, 417A family, and they were very consistent both for total THD+N and for the individual levels of their harmonics, so a typical example is shown (see Figure 7.41).

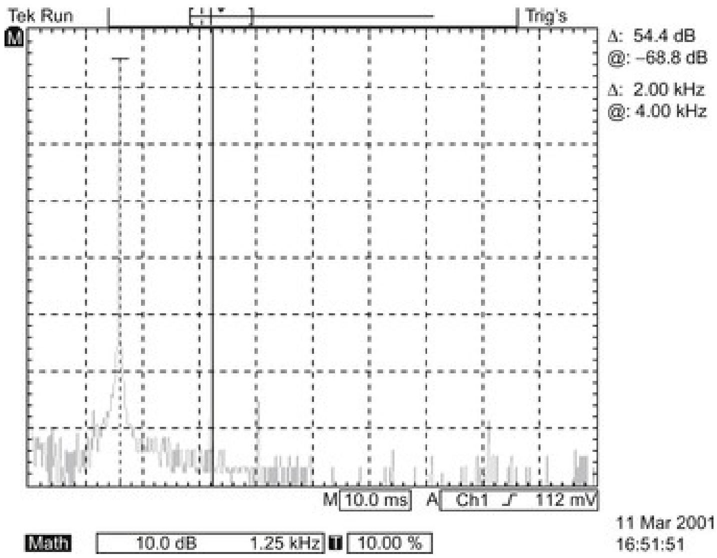


Figure 7.41 Typical distortion spectra of EC8010/5842/417A family at +18dBu.

The distortion is dominated by second harmonic at −44 dB (0.65%), and the fourth harmonic is 54 dB below this at an entirely negligible −98 dB. Because distortion for a triode is proportional to level, we can predict the distortion at the proposed operating level. The nominal input sensitivity is required to be 2.5 mV $_{RMS}$ for 5 cm/s, and we convert this to dBu:

$$\text{dBu} = 20\log\left(\frac{v_{(\text{mV})}}{775}\right) = 20\log\left(\frac{2.5}{775}\right) = -50\ \text{dBu}$$

But we know that programme peaks will be 12 dB higher than this, so the peaks reach −50 dBu+12 dB=−38 dBu. Using an EC8010, the stage had a measured gain of 32 dB, so programme peaks at the output of the stage reach −38 dBu+32 dB=−6 dBu. We tested distortion at +18 dBu, which is 24 dB higher than −6 dBu, so the distortion at −6 dBu will be 24 dB better than that measured at +18 dBu. Thus, the distortion at −6 dBu will be −44 dB − 24 dB=−68 dB=0.04%, which is perfectly satisfactory. If we like impressive numbers, we could instead quote the distortion at the nominal 5 cm/s level, which reduces it to 0.01%.

We should next check the input capacitance of the stage. For the EC8010, Siemens specifies C_{ag} =1.4 pF, but this will be multiplied by (1+ A_v) to give a Miller capacitance of 57 pF. C_{in} =7 pF, so the total input capacitance is 64 pF. Since the author already had the stage set up on the bench, it was easy to check this value.

Adding a resistor in series with the oscillator output produces a low-pass filter in conjunction with C_{input} . The resistor value is not critical, so long as its exact value is precisely known. The filter $f_{-3\ \mathrm{dB}}$ point can be found by adjusting oscillator frequency until the amplitude at the output of the test stage drops by 3 dB or its phase (relative to input) changes to 135° (180°−45°). Using an 18-kΩ resistor, $f_{-3\ \mathrm{dB}}$ =46.9 kHz.

$$C_{\mathrm{input}} = \frac{1}{2\pi f_{-3\ \mathrm{dB}} R} = \frac{1}{2 \times \pi \times 46{,}900 \times 18{,}000} = 189\ \mathrm{pF}$$

This is a long way away from the expected value. Since we know the gain A_v of the stage, C_{ag} can be determined using the Miller equation in reverse:

$$C_{ag} = \frac{C_{\mathrm{input}} - C}{1 + A_v} = \frac{189\ \mathrm{pF} - 9\ \mathrm{pF}}{1 + 38} = 4.6\ \mathrm{pF}$$

C_{in} is the capacitance from grid to all other electrodes, as specified by the manufacturer, plus a small allowance for strays – perhaps 2–5 pF, depending on test circuit layout.

Since the manufacturer claims C_{ag} =1.4 pF, the value of 4.6 pF came as quite a surprise, but a direct measurement of C_{ag} on the component bridge gave a value of ≈4.8 pF. All bridges have trouble measuring small capacitances, and the Marconi TF2700 used for this measurement was no exception. Nevertheless, the manufacturer's claimed value for C_{ag} is clearly hopelessly optimistic at audio frequencies.

Optimising the Input Transformer

Unfortunately, 190 pF is a large shunt capacitance for the input transformer and initial square wave tests with the Sowter 8055 were very disappointing, but a Zobel network across the transformer secondary greatly improved matters. The required value of Zobel capacitance depends on cartridge DC resistance, as shown in Table 7.6 .

Table 7.6 Zobel Capacitance for Sowter 8055 Moving-Coil Input Transformer when Loaded with 190 pF//6k8

Cartridge R_{DC}	4 Ω	6 Ω	8 Ω	10 Ω
C_{Zobel}	1.5 nF	1 nF	910 pF	680 pF

Alternatively, the Jensen JT-346-AX transformer can be used, but it is a little noisier (higher DC winding resistance). The Jensen transformer was designed for 3 Ω or 5 Ω cartridges when set to 1:12 step-up ratio, and the manufacturer's data sheet gives appropriate Zobel values (assuming zero load capacitance). Experimentation revealed that 680 pF and 2k4 were optimum Zobel values for 11 Ω source resistance and 190 pF load capacitance.

Ultimately, the choice between transformers will probably be determined by your country of residence because buying transformers abroad is expensive. Thus, UK readers will probably opt for the Sowter, and North American readers the Jensen, whereas European readers will probably investigate Lundahl.

The Second Stage

This stage has the highest amplitude signals and therefore can be expected to produce the most distortion. As expected, when tested, gain stages with active loads such as the μ -follower and β -follower produced significantly lower distortion than a simple common cathode triode amplifier, with the further bonus of a reduced output resistance.

The μ -follower proved to be an excellent test bed for determining the irreducible distortion of a valve. The huge expansion of Internet trading means that there is now a world market for New Old Stock (NOS) valves, and almost any valve that was ever made is available somewhere. The second stage needed a valve with $\mu \approx 16$, so any likely candidate was tested – together with some unlikely ones (full details in Chapter 3).

Surprisingly, given its good reputation, the 76 did not measure particularly well. Although it produced the lowest second harmonic distortion of all, its distortion was *not* proportional to level, and the higher-order harmonics were at a comparatively high level. Since single-ended design relies on distortion falling with level, this valve was reluctantly eliminated.

One triode was significantly better than all others. It might be packaged as a single or a dual triode, and it might have an Octal or a Loctal base, with a 6.3 V or a 12.6 V heater, but internally the valve is the same. Perhaps unsurprisingly, this valve is the 6SN7, 12SN7, 7N7, 14N7 or 6J5. The selection was further whittled down by the discovery that the metal envelope variants produced measurably higher distortion, probably due to outgassing from the hot envelope causing increased grid ion current.

Further tests revealed that the blackened glass variants such as the CV1988 (UK military 6SN7) consistently produced the lowest distortion, but that the far cheaper Pinnacle 6J5GT was very good, and selected examples equalled the CV1988. The manufacturers claim a usefully reduced C_{ag} (3 pF compared to 3.9 pF) for the Loctals, but the main reason for choosing a Loctal would be to avoid the leakage of a phenolic base that potentially increases noise. Any of these variations on the theme would be suitable, and the final choice would probably be decided by more prosaic matters such as heater supplies, convenience of single versus dual triodes, availability of valve bases or whether

you have any.
The μ -follower and β -follower were extensively tested, and the β -follower was very good, but the μ -follower had the slight advantage that it is more flexible about HT voltage, and the author wanted to ease regulator design by operating the second and third stages at the same HT voltage.
For the Pinnacle 6J5GT μ -follower, typical distortion at +28 dBu was 0.25% or −52 dB. Using this valve, programme peaks at the output of the second stage reach +12 dBu, which is 16 dB lower, so the distortion can be expected to be −52 dB−16 dB=−68 dB=0.04%, which is the same as the input stage.
Once distortion due to the lower valve in a μ -follower has been minimised, the choice of upper valve slightly affects distortion. Various valves were tried, such as triode-strapped D3a, 6C45π and Pinnacle 6J5GT. The difference in distortion between the various types was small, but the Pinnacle 6J5GT was marginally better at 8 mA than the other valves, so it was chosen.
Having found C_{ag} of the EC8010 to be higher than expected, the Pinnacle 6J5GT was also tested. Two different measurement methods gave $C_{ag} \approx 5.4$ pF, slightly higher than the generic value of 4 pF.

The Output Stage

The third stage has very similar output level requirements to the second stage, so another μ -follower was indicated. However, far less gain (and input capacitance) is required, so a 6J5GT is not suitable. There are very few low- μ small triodes available – the 6BX7, 6AH4 (μ =8) and 12B4-A (μ =6) are obvious choices. These valves were designed for television field scan or series regulator use, so linearity is not guaranteed. The author had considerable qualms about the decision, but eventually decided to select from his stock of 12B4-As to find a pair of low distortion samples for the output stage.
All of these low- μ valves require significant $-V_{gk}$ to set optimum operating conditions, so LED bias becomes less practical, and battery bias is required if bias shift is to be avoided after overload. Again, a Pinnacle 6J5GT was chosen for the upper valve.
If the 12B4-A cannot be selected for low distortion, one possible alternative is to use an NOS 37 (μ =9) for both the second stage and the output stage. Even though a test of nine samples indicated that this valve produces double the distortion of a Pinnacle 6J5GT, its distortion was far more consistent between samples than the 12B4-A, so final performance could be better than if a pair of unselected 12B4-As had to be used.

Refining Valve Choice by Heaters

If the 12B4-A has its heaters strapped in parallel for 6.3 V (I_h =0.6 A), a stereo pair requires I_h =1.2 A. If an *SN7/*N7 is shared between the stereo channels for the second stage, it requires I_h =0.6 A. Together with the EC8010 (I_h =0.28 A), a total of 2.08 A is then required from the 6.3 V regulator. This is achievable, but a little awkward, and a pair of series 300 mA heater chains would be more convenient. Shunting the EC8010 by a 315 Ω resistance allows it to be used in a 300 mA chain; the 300 mA 6J5GT is directly suitable, and the 12B4 A can be used as a 12 V 300 mA heater, so the final line-up is EC8010, Pinnacle 6J5GT and 12B4 A.

Apart from the relaxed requirements of the heater regulator, a series heater chain has other advantages, which are detailed in Chapter 4 , not least of which is reduced sensitivity to RF noise.

Choosing the Implementation of RIAA Equalisation

The EC8010 first stage has a gain of 38, resulting in an output of only 95 mV_{RMS} at 1 kHz 5 cm/s, which is well below the 200 mV_{RMS} decision point between split and 'all in one go' equalisation. Consideration of the second stage's noise requirements if 'all in one go' equalisation were to be applied to a 95 mV_{RMS} signal will quickly reinforce this decision.

We know that we will use split passive RIAA equalisation and the topology of individual gain stages. We must now choose impedances for the equalisers that give the best balance between distortion due to loading or grid current and equalisation errors due to stray capacitances and non-zero source resistances.

Grid Current Distortion and RIAA Equaliser Series Resistances

All valves source some grid current. When a valve is fed from a non-zero source impedance, its grid current develops a voltage across that impedance. Unfortunately, this voltage (which is in series with the wanted signal) is usually distorted, so it adds distortion to the wanted signal.

Passive RIAA stages must include series resistance to form their equalisers, so this provides a mechanism for grid current to introduce additional distortion. Sadly, reducing series resistance in order to reduce grid current distortion has snags:

- At frequencies when an equaliser provides maximum attenuation, the driving stage must drive a load equal to the series resistance. Reducing a stage's load resistance steepens its loadline and increases distortion. Stages including a cathode follower, such as the proposed μ -followers, are more tolerant of loading, but caution is still needed.

- The required capacitances for the RIAA equalisers become rather large. Fortunately, 1% polypropylene capacitors are now available, but their restricted range of values and voltage ratings means that a certain amount of juggling is necessary.

From the input stage to the second stage, a 20 kΩ series resistor would have been ideal from the point of view of grid current distortion, but this loading would have reduced the gain and increased the distortion of the EC8010 input stage. On test, 47 kΩ series resistance was a suitable compromise that minimised distortion due to the two effects. Happily, the 6J5GT/6J5GT μ -follower second stage could comfortably drive 20 kΩ, making grid current distortion due to the third stage 12B4-A/6J5GT μ -follower invisible.

3180 µs, 318 µs Pairing Errors due to Miller Capacitance

In our example RIAA stage, we argued that the only logical third stage was a cathode follower because it allowed DC coupling which eliminated interaction and errors to the 3,180 µs, 318 µs pairing, and its low input capacitance avoided high frequency loss. If we could tolerate interaction, and had a means of predicting and solving the problem, then this would allow a little more freedom of design choice.

If we want to achieve levels from vinyl comparable to those from digital sources, we *must* increase the gain of the RIAA stage. Increasing the μ of the valve at the input or second stage causes Miller capacitance problems, so the only practical way of substantially increasing gain (without increasing the number of valves producing distortion) is to substitute a common cathode amplifier for the final cathode follower, which immediately introduces two new problems:

- The final stage must have its input AC coupled, causing interaction between the new low frequency roll-off introduced by the coupling capacitor and the 3,180 µs time constant, producing low frequency response errors.

- Because the new final stage has a gain >1, Miller capacitance becomes significant, and the equalisation network will be loaded by a far larger stray capacitance than before, causing high frequency response errors.

The 75 µs Problem

Whenever possible, extended foil polystyrene capacitors are desirable for equalisation networks since this form of construction significantly reduces ESR and series inductance. Unfortunately, commercially available types have a voltage rating of only 63 V_{DC}, so the inter-stage coupling capacitor between the first and second stages has been forced to revert to its more traditional position, ensuring interaction with the 75 µs equalisation.

Additionally, the grid-leak resistor has been moved so that it is no longer near the grid but discharges the grid via the series resistor of the RIAA network, thus eliminating the potential divider that caused 1.6 dB loss in the basic pre-amplifier. To the author's knowledge, the first use of this cunning trick was in Arthur Loesch's transformerless RIAA MC stage [9] (see Figure 7.42).

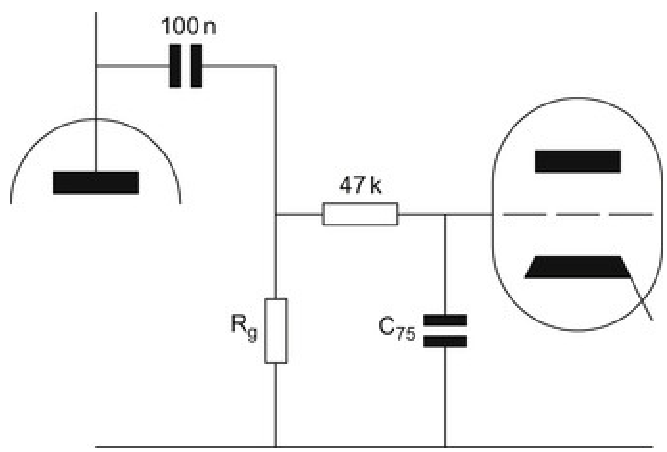


Figure 7.42 Modification to 75 µs implementation that eliminates unnecessary loss.

The Computer Aided Design (CAD) Solution

The various interaction problems can be solved by iterative CAD AC analysis. We start by calculating values in the normal way (assuming no interaction), and then use CAD to predict the effects of interaction on frequency response using a sweep between 2 Hz and 200 kHz. Once a problem is revealed, we adjust individual component values to seek improvement. Although this sounds laborious, it can actually be quite quick, provided that we think about how, and where, we make our adjustments.

We have five variables that must be juggled to produce the correct result, so some simplification is needed. We could best start by analysing a design that does not have interaction and gently modifying it, gradually introducing interactions until we reach our final design. Alternatively, the 3,180 µs, 318 µs pairing is most affected by interaction, so we could change these components first then determine the remaining values.

3180 µs, 318 µs Pairing Manipulation

- The shelving loss at frequencies <20 Hz caused by adding the inter-stage coupling capacitor can be cured by reducing the value of the upper resistor in the potential divider.
- A mid-range shelved response (where frequencies above 1 kHz are at a constant, but different, level to those below 250 Hz) can be cured by changing the lower resistor value in the potential divider. If the higher frequencies are at too high a level, this is because the potential divider has insufficient attenuation, so the lower resistor must be reduced in value, and vice versa.
- A peak in the response centred near 500 Hz can be cured by increasing the capacitor value, whereas a dip can be cured by reducing capacitor value. This result is not quite so easily deduced, but a larger capacitor would increase the time constant, lowering the frequency at which the potential divider takes effect so that attenuation begins earlier than it should, resulting in a dip in the final response.

The last two adjustments are highly interactive, and an increase in one immediately requires a proportionate decrease (have a calculator handy) in the other to maintain the correct time constant. It is usually easiest to optimise the resistor first. The model should be tested down to 2 Hz, and the low frequency roll-off adjusted to emulate a simple 6 dB/octave filter, then optimised for minimum amplitude deviation from 20 Hz to 20 kHz.

75 µs/3.18 µs Manipulation

Although RIAA record equalisation is *specified* with only three time constants (3,180 µs, 318 µs and 75 µs), this would imply a 6 dB/octave rising response at the fragile cutting head. Wright [10] pointed out that at the time of cutting, RIAA pre-emphasis cannot continue indefinitely and that a final time constant of ≈3.18 µs is commonly added to prevent excessive amplitude at ultrasonic frequencies from damaging the (probably Neuman) cutting head. Unfortunately, the value of this time constant varies between cutting head manufacturers, and the less common Ortofon heads use a time constant nearer to 3.5 µs. Nevertheless, it seems reasonable to accept that an electrical 3.18 µs time constant has been deliberately added at the cutting stage in addition to the inevitable mechanical losses within the cutting heads themselves. The new replay equation is therefore:

$$G_s = \frac{(1 + 318 \times 10^{-6} \times s)(1 + 3.18 \times 10^{-6} \times s)}{(1 + 3180 \times 10^{-6} \times s)(1 + 75 \times 10^{-6} \times s)}$$

where

$$s = j\omega \quad \text{and} \quad \omega = 2\pi f.$$

This is an even more unpleasant equation than the original RIAA equation. Briefly, the effect is that instead of tending towards a 6 dB/octave low-pass filter, it tends towards ≈27.5 dB attenuation that is constant with frequency. Within the audio band, the new equaliser corrects a 0.64 dB loss at 20 kHz.

The justification for adding a 3.18 μs time constant to the replay equalisation has little to do with amplitude response, but more to do with group delay and transient response. Uncorrected, the 3.18 μs time constant changes the phase of frequencies above 5 kHz so that they no longer arrive at the same time as lower frequencies (unequal group delay), and this distorts the transient response. We cannot compensate for the cutter, and we probably don't have the data to compensate for the cartridge response, but we can compensate for the hidden 3.18 μs time constant.

However, Yaniger convincingly argues in his 'His Master's Noise' [11] article that as moving coil cartridges typically have a fierce ultrasonic tip mass resonance (which will certainly compensate for the missing time constant) and that moving-magnet cartridges don't, only an RIAA stage intended for moving magnet cartridges should include the 3.18 μs time constant. Either way, the final time constant of 3.18 μs is physically easily included by adding a resistor in series with the capacitor producing the 75 μs time constant (see Figure 7.43).

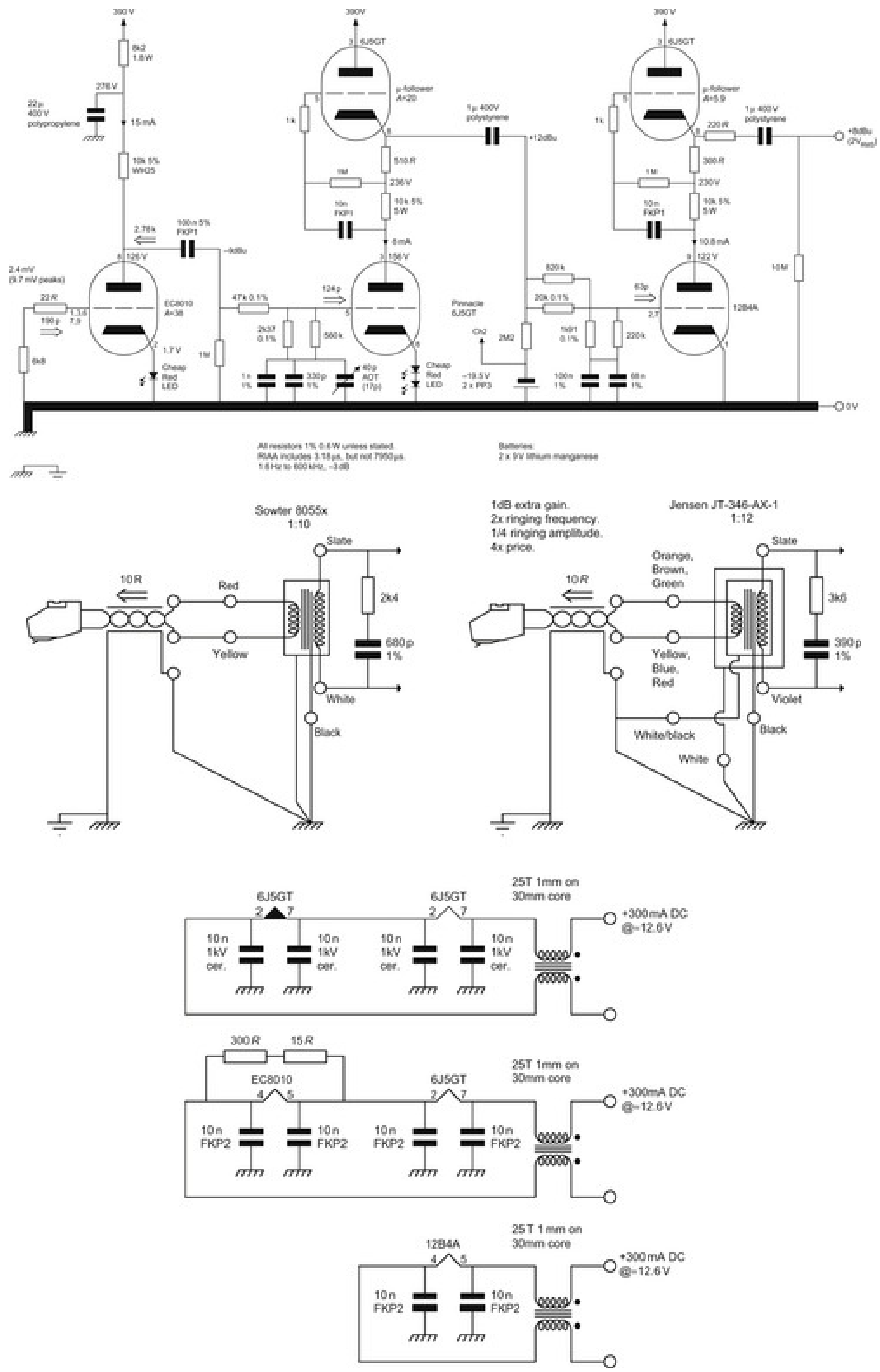


Figure 7.43 Final design of EC8010 μ-follower RIAA pre-amplifier.

Sadly, setting the exact value of the resistor is considerably more difficult because there are so many other high frequency roll-offs within the pre-amplifier, usually dominated by the output stage loading of the 3,180-μs, 318-μs pairing. Usually, only the additional resistor needs adjustment, but minor adjustments of the 75 µs capacitor are likely. The model should be tested up to at least 300 kHz, and finally adjusted for optimum group delay, then checked for deviations between 20 Hz and 20 kHz. It may be even necessary to make minor changes to the 3,180 µs, 318 µs pairing.

Practical RIAA Considerations

Setting the precise practical value of capacitance for the 75 µs, 3.18 µs pairing is awkward, so an Adjust On Test trimmer (AOT) is included. There are various alternatives for setting the AOT trimmer:

- Set the vanes almost half open (≈17 pF), and assume correct values for the other capacitors.
- Measure the other 75 µs, 3.18 µs capacitors on a bridge, and set the trimmer to give a predicted total capacitance of 1.35 nF, or connect them all in parallel and adjust the trimmer to give 1.35 nF.
- Measure RIAA frequency response accuracy (with a 3.18 µs capable instrument), and set the trimmer for correct response.

Note that the capacitance of all capacitors falls with frequency, it's just that the defects of the better dielectrics are less visible, so although the bridge methods are indirect, they are likely to be the most accurate provided that the capacitors are polypropylene or better.

RIAA Direct Measurement Problems

Given a well-equipped laboratory, direct measurement of RIAA equalisation errors seems simple. Unfortunately, RIAA equalisation ranges from ≈+20 dB at 0 Hz to ≈−25 dB at >50 kHz, making precise measurement quite difficult.

If a constant level is applied to the RIAA stage, its level must be chosen so as to avoid overload and the measuring amplifier must accommodate the ≈45 dB range without any error. Conversely, setting the output level to be constant requires that the oscillator can set exact levels over a ≈45 dB range that can be measured precisely. Depending on the test equipment, this is either a conversion problem between the analogue and digital domains, or an analogue attenuator problem. Either way, guaranteeing attenuator error ≤0.02 dB and *simultaneously* a flat frequency response over a 45 dB range is not trivial and costs money.

A popular alternative is to feed the RIAA stage via a passive RIAA pre-emphasis network and measure the combined frequency response. A theoretically ideal perfect RIAA pre-emphasis network would have an output that rose indefinitely at a rate of 6 dB/octave from ≈5 kHz, but practical passive networks *must* have a final time constant – even if it isn't 3.18 µs.

RIAA pre-emphasis networks are quite tricky to design, and even a perfectly designed and constructed RIAA pre-emphasis network has problems because it is sensitive to source and load impedances, which are generally considered to be constant during its design. Sadly, the carefully optimised real-world loading required by a moving magnet cartridge or a moving coil transformer disturbs the load impedance, and an incorrect oscillator source resistance would compound the problem. RIAA pre-emphasis networks should use polystyrene capacitors to avoid the slight frequency dependence of polypropylene capacitance (see Figure 7.44).

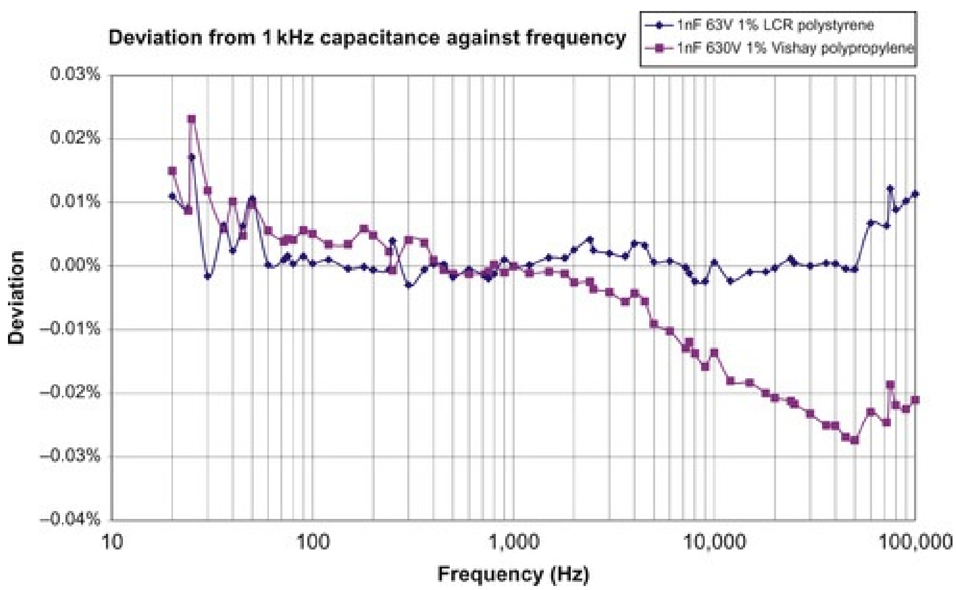


Figure 7.44 Deviation from 1kHz capacitance against frequency for polystyrene and polypropylene capacitors.

Summing up, keeping measurement errors below RIAA stage design errors is difficult.

Production Tolerances and Component Selection

Once we have optimised component values, we can test the effects of errors due to component tolerances. There is little point in specifying close tolerance components in one position if others with looser tolerances are able to upset performance.

The computer predicted the 20 Hz to 20 kHz frequency response 10,000 times, each time with random changes in all component values within their manufacturer's tolerance. This technique is known as Monte Carlo analysis, and provided that sufficient runs are used, it predicts a likely worst case spread of frequency response. The predicted error spread for the EC8010 pre-amplifier was ±0.25 dB using the specified standard component values and without deliberate pre-selection to obtain optimum values other than setting the 75 µs, 3.18 µs trimmer capacitor to its nominal value of 17 pF.

Although RIAA errors are awkward to measure, the problem can be side-stepped by pre-selecting capacitors using a component bridge, whilst a 4½ digit DVM might even allow selection of 0.1% resistors. Even without component pre-selection, the error with new valves is likely to be well within ±0.25 dB, and pre-selection could further reduce errors.

RIAA Equalisation Errors due to Valve Tolerances

Even when a design deliberately sets out to minimise the effects of valve tolerances, the valves still dominate RIAA errors because passive components can now be so precise.

Unfortunately, r_{out} of the input stage is a significant proportion of the series resistance that determines the 75 µs, 3.18 µs pairing. Despite this, the computer predicted an HF shelf loss of only 0.15 dB due to g_m of the EC8010 falling to ⅔ of its nominal value.

As g_m falls, r_a rises, which reduces gain and Miller capacitance. In this design, the valves expected to affect RIAA accuracy due to changes in Miller capacitance are the second and final valves, but as they are operated as μ-followers, changes in r_a do not affect gain ($R_L \approx \infty$), so this mechanism is not significant.

Because r_{out} for the μ-follower is such a small proportion of the resistance that determines the RIAA time constants, a tired valve in the upper stage of a μ-follower does not significantly affect RIAA accuracy.

This pre-amplifier's weakness is its sensitivity to variations of C_{ag} in the lower 6J5GT of the second μ-follower. If C_{ag} rises by 50%, an HF shelving loss of 0.32 dB is predicted, whereas if it falls by 50%, an HF shelving boost of 0.34 dB is predicted. Happily, the pre-amplifier is immune to ±50% variations in C_{ag} for the 12B4-A because the 20 kΩ series resistor chosen for low distortion forces low impedances in the 3,180 µs, 318 µs pairing.

Some pre-amplifiers using passive equalisation with high-μ valves, such as ECC83, have been found to sound audibly different with different makes of valve, giving rise to the belief that a Siemens ECC83 is better (or worse) than a

Mullard, when it was actually differing r_a and C_{gk} causing clear errors in RIAA equalisation.

The Balanced Hybrid RIAA Stage

The Denon DL103 moving-coil cartridge offers excellent performance for its price, so the challenge was to equalise and amplify its signal to the digital 2-V $_{RMS}$ standard with a similar price/performance ratio.

No Step-Up Transformers

The DL103 is an idiosyncratic cartridge to use – perhaps explaining its mixed reviews. Its low compliance means that it needs a very high mass arm having bearings that will not rattle, and its abnormally high coil resistance in combination with step-up transformer leakage inductance forms a significant low-pass filter, which means that it suffers high frequency loss unless partnered with step-up transformers each costing more than the cartridge. Having rejected step-up transformers, we now have the problem of amplifying 0.3 mV $_{RMS}$ at 1 kHz at 5 cm/s from a 40 Ω source with negligible hum and noise. If we ran a triode-connected E810F at full tilt, we could obtain g_m =50 mA/V, resulting in an equivalent noise resistance of 50 Ω – comparable with the source. But E810Fs are expensive, and we would need 35 mA per channel of even more expensive anode current. We need a cheaper form of low-noise amplification.

Semiconductors to the Rescue

The way to obtain the required g_m and therefore low noise is to use a semiconductor input device. We could make an entire semiconductor gain stage, then couple it to a valve stage, but that would be wasteful of components and involve unnecessary AC coupling. A more elegant solution is to make a hybrid cascode with the semiconductor as the lower device and the valve the upper (see Figure 7.45).

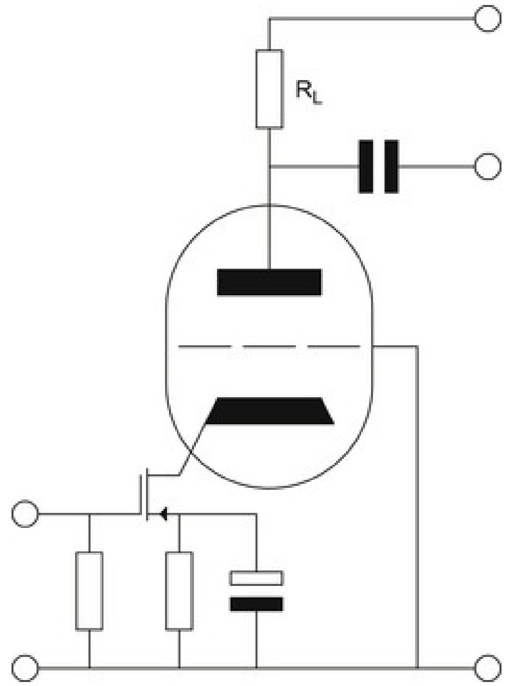


Figure 7.45 Hybrid FET/triode cascode.

The gain of a cascode is:

$$A_v \simeq g_m \cdot R_L$$

To a working approximation (neglecting semiconductor output resistance), the semiconductor sees as its load the triode's anode load R_L divided by (μ +1), which implies that we can adjust the balance of the cascode's gain structure between upper and lower devices by changing μ. A high value of μ puts more of the gain in the valve, whereas a low value shifts it towards the semiconductor. Unfortunately, high-μ valves such as the ECC83 and 7F7 probably can't pass enough current for the semiconductor to achieve the noise performance we require. Although we could drive additional current from elsewhere, this would be tantamount to injecting noise directly into the input stage, and no designer likes doing that. All of the semiconductor's required current must therefore come through the valve, and that means we need the Loctal *N7 (at these signal levels, we can't tolerate the leakage currents in the phenolic base of an* SN7). If we assume that we will need ≈3 mA and that the *N7 has r_a ≈6 kΩ and can tolerate a 33 kΩ load, we now know that our semiconductor sees a load of:

$$R_{L(\text{semiconductor})} = \frac{R_{L\ (\text{valve})} + r_a}{\mu + 1} = \frac{33 + 6}{20 + 1} \simeq 1.9\ \text{k}\Omega$$

This is an acceptable load and suggests that it is worth continuing with the design.

Miller Capacitance

A typical LSK170C JFET can achieve $g_m \approx 15$ mA/V at $I_{ds} = 3$ mA, so to a first approximation, the cascode gain will be:

$$A_v = g_m \cdot R_{L(\text{valve})} = 15 \times 33 \simeq 500$$

More significantly, the gain to the valve's cathode will be:

$$A_v = g_m \cdot R_{L(\text{semiconductor})} = 15 \times 1.9 = 28.5$$

This is a much higher value than normal within a cascode and is a direct consequence of deliberately choosing a semiconductor having high g_m. In effect, minimising noise by maximising semiconductor g_m always results in high gain to the valve's cathode, and this means that the semiconductor's Miller capacitance (the capacitance seen looking into its input) must rise:

$$C_{\text{Miller}} = (A_v + 1) \cdot C_{\text{reverse}}$$

Worse, the capacitance of the depletion region at a reverse biassed semiconductor junction is voltage dependent:

$$C_{\text{reverse}} \propto \frac{1}{\sqrt{V_{\text{reverse}}}}$$

If we want to reduce the signal dependency of our Miller capacitance, we need a reasonably large DC voltage (10 V or more) across the semiconductor. Increasing V_{reverse} reduces the problem of C_{reverse} in two ways:

- A high V_{reverse} reduces C_{reverse} (making it less of a problem).
- Signal swing becomes a smaller proportion of V_{reverse}, reducing modulation of C_{reverse}.

You will undoubtedly have seen FET/triode cascodes where the valve's grid is grounded, resulting in only a volt or two across the FET, implying a large and signal-dependent Miller capacitance at the FET's gate. This might not be ideal, but it is done because the designer is terrified of injecting noise into the input stage by lifting the valve's grid from ground.

DC Stabilisation and Consequent Gain Reduction

As soon as we include R_s to add DC feedback to stabilise DC conditions, we have also added AC feedback and reduced cascode gain. If we need to drop 100 mV whilst passing 3 mA, $R_s = 33$ Ω, and we can calculate the effect on cascode gain if this resistor is left unbypassed:

$$\beta = \frac{R_s}{R_{L(valve)}} = \frac{33}{33{,}000} = 0.001$$

Using the feedback equation, the cascode gain becomes:

$$A = \frac{A_0}{1 + \beta A_0} = \frac{500}{1 + 0.001 \times 500} \geq 330$$

Remembering that we want to amplify a 0.3 mV $_{RMS}$ signal, this translates to 100 mV $_{RMS}$ at the output of the cascode, which is satisfactory. We would normally fret about the effect on r_a of leaving such a capacitor unbypassed, but the cascode already has near-infinite r_a compared to its R_L, so making it higher doesn't make matters any worse. Nevertheless, we will have to address the cascode's high r_a problem later.

JFET Noise

The LSK170C is specified as producing 0.9 nV/√Hz at 1 kHz, so we can quickly estimate the theoretical signal-to-noise ratio of our cascode. We start by finding the equivalent noise of the LSK170C over the audio bandwidth:

$$V_{n(nV)} = 0.9 \times \sqrt{20{,}000 - 20} = 127\ \mathrm{nV_{RMS}}$$

We know that our cartridge produces 0.3 mV $_{RMS}$, so we divide the one by the other to obtain our signal-to-noise ratio:

$$\mathrm{S/N} = 20 \log\left(\frac{0.3\ \mathrm{mV}}{127\ \mathrm{nV}}\right) = 67.5\ \mathrm{dB}$$

But we saw earlier that RIAA equalisation confers a 3.4-dB noise advantage, so the final signal-to-noise ratio is 69 dB. This quick calculation has neglected the noise generated by the cartridge's 40 Ω winding resistance, but it's useful because Tim de Paravacini's rule of thumb is that the practical limit for any RIAA stage's signal-to-noise ratio is ≈68 dB ref. 5 cm/s. Thus, the calculation suggests that our hybrid cascode input stage can't be too far away from the practical limit.

Unfortunately, we have also neglected 1/*f* noise, and JFETs tend to have a 1/*f* noise corner somewhere between 10 Hz and 100 Hz – just where RIAA has plenty of gain. In short, 1/*f* noise is *not* negligible in this application and tends to disqualify JFETs from being used to amplify moving coil cartridges – a low-noise BJT is needed.

BJT Noise

As we saw in Chapter 3 , all amplifying devices have a voltage noise source and a current noise source. Unless we're considering capacitor microphone capsules or other capacitive sources, we can neglect current noise in FETs and valves, hence the simple 0.9 nV/√Hz JFET noise calculation. Sadly, bipolar junction transistors are not so simple, and we must calculate then sum both sources of noise.

The SSM2210 is Analog Devices' version of the MAT02/LM394 supermatch transistor and has v_n ≈0.8 nV/√Hz and i_n ≈2 pA/√Hz at 3 mA, but more significantly its 1/ *f* noise corner is ≈1.5 Hz – below the audio band and much lower than any contemporary JFET. We calculate the voltage noise as before and obtain a figure of 113 nV_{RMS} , but we must also calculate the noise current:

$$i_{n(pA)} = 2 \times \sqrt{20{,}000 - 20} = 283\ pA_{RMS}$$

We then use Ohm's law to find the voltage this develops across the 40 Ω source resistance of the cartridge:

$$v_n = IR = 284\ pA \times 40\ \Omega = 11\ nV_{RMS}$$

This noise voltage is uncorrelated with the transistor's noise voltage, so we must add noise powers:

$$v_n = \sqrt{v_{n1}^2 + v_{n2}^2} = \sqrt{113^2 + 11^2} = 114\ nV_{RMS}$$

In this particular instance, adding the noise due to the current source barely changed the final noise, but a higher source resistance such as a moving magnet cartridge would make a significant difference, biassing the noise argument firmly in favour of the JFET.

We divide the 114 nV_{RMS} noise into our 0.3 mV_{RMS} signal to find our signal-to-noise ratio as before, resulting in a figure of 70 dB once we include the 3.4 dB RIAA advantage. This calculated value of 70 dB doesn't violate the 68 dB rule of thumb because we have neglected all other noise sources in the RIAA stage. Nevertheless, this 70 dB figure is far more likely to be credible than the JFET's 71 dB because of the SSM2210's low 1/ *f* corner frequency. Irritatingly, the SSM2210 is now obsolete and has been replaced by the electrically identical SSM2212 that isn't available in eight-pin DIP. Surface Mount Devices (SMDs) *can* be soldered by hand, but you need a very fine tip and <0.5 mm diameter silver-loaded solder.

Choosing between the BJT and JFET: Equalisation, Distortion and HT Power

The BJT has far higher g_m than the JFET:

$$g_m = 35I_C = 35 \times 3 = 105\ \text{mA/V}$$

Therefore, cascode gain will be much higher:

$$A_v \geq g_m \cdot R_L \geq 105 \times 33 \geq 3{,}500$$

This implies that there will be ≈1 V_{RMS} leaving the first stage, rather than the 100 mV_{RMS} calculated for the JFET version, so the immediate implication is that the JFET requires split equalisation over three stages, whereas the BJT version must use 'all in one go' equalisation and might only need two stages.

So far, we have not considered distortion, but cascodes are not especially linear because the lower device faces the steep loadline of the resistance looking into the valve's cathode. Moreover, the BJT version will suffer 10 times as much distortion simply because with a fixed input signal, 10 times the gain translates into 10 times the output voltage.

Steep loadlines tend to increase second harmonic distortion, which can be cancelled by push–pull action. Thus, one way to deal with the greater distortion of the BJT version would be to configure a pair of cascodes as a differential pair. Not only would this halve the voltage on each cascode (halving the distortion), but it would also cancel the predominant second harmonic distortion. The SSM2210 comprises two NPN transistors having the defect h_{fe} matched to 'about 0.5%' [12], making it particularly suitable for differential pairs.

Thus, the decision between JFET or BJT will be determined not so much by marginal noise considerations but by single-ended versus balanced considerations.

The EC8010 RIAA stage demonstrated that it is possible to obtain very low distortion in a single-ended topology, but that one part of the price is a high HT voltage (390 V) because the μ -follower is a series amplifier. 100 V of the voltage drop is due to the 10 kΩ 5 W resistors, and a β -follower could avoid that drop, bringing the HT down from ≈400 V to ≈300 V. Turning to the balanced topology, it's very easy to not only double the component count, but also double the HT power requirement – and that's expensive.

If a balanced topology is to be chosen, it must use the same HT power (or less) as an equivalent single-ended design and have another advantage to justify the potentially higher component count. Having become accustomed to hum-free balanced working from moving coil cartridges, the author was not keen to return to the hum problems associated with unbalanced working.

Reconciling the Balanced Decision with Practicalities

Having made the decision to adopt a balanced topology and therefore choose the SSM2210 BJT input device, the design needs to be made practical. The EC8010 RIAA stage requires 26 W from its HT, and almost half of that power is due to the need to sink 15 mA through each EC8010 input stage in order to secure low noise. Conversely, the SSM2210 produces its lowest voltage noise at I_c=3 mA, and because differential pairs are inherently low distortion, there is no need to squander HT voltage across series amplifiers. Thus, one design aim should be to reduce HT voltage and keep the current down, minimising HT power.

We have already seen that the higher gain of the BJT/triode cascode requires 'all in one go' equalisation, so another design aim should be to only need two gain stages.

Considering the second stage in particular, although the differential pair cancels even harmonic distortion, it would be better to minimise distortion before cancellation, and that implies maximising its ratio of R_L to r_a. But we need to be able to drive a cable or sound card, so a unity gain cable driver is also required. We are now able to draw a block diagram (see Figure 7.46).

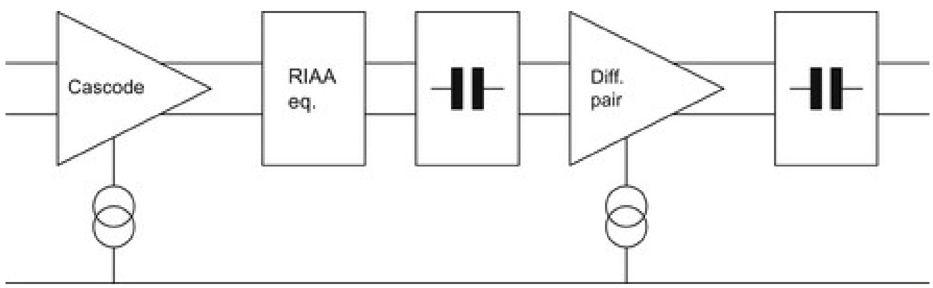


Figure 7.46 Block diagram of balanced hybrid RIAA stage.

Implications of the Block Diagram

In our implementation, the emitters of the SSM2210 are tied together and its bases are tied to ground, so both its transistors have the same V_{BE}, forcing their base currents to be the same, and if their defect h_{FE} is matched, then their collector currents must also be matched. If the triodes draw no grid current, then $I_k = I_a$, and all the transistors' collector current must flow through the triodes' anode loads, so if the anode loads are matched, the matched collector currents must result in matched anode voltages.

The significance of the previous argument is that the block diagram usefully shows us that the 'all in one go' equalisation can be implemented before AC coupling to the second gain stage, minimising interaction between the 3,180 μs and 318 μs time constants and the AC coupling time constant, and also enabling the use of low-voltage close tolerance capacitors for the RIAA network.

However, the diagram also shows us that the Miller capacitance of the second gain stage is across the RIAA network, so this must be taken into account during calculation. More significantly, it should be minimised (to minimise errors when it varies), suggesting that E88CC (C_{ag} =1.4 pF) would be a better choice than *N7 (C_{ag} =3 pF).

Another requirement highlighted by the block diagram is that we need double the usual number of coupling capacitors, so they had better not be expensive.

Since the SSM2210 bases of the BJT/triode cascode differential pair must be at 0 V (to avoid coupling capacitors to the cartridge), the emitters must be at −0.7 V, necessitating a negative supply for the tail's CCS. Given that a subsidiary supply is already required, the ideal CCS is a cascode BJT design, which would work well from a −15 V regulated supply. An identical CCS will do very nicely for the second gain stage's tail.

The Unity-Gain Cable Drivers

Unity-gain cable-driving ability immediately implies low distortion followers with constant current loads: either cathode followers or source followers. Further, we need AC coupling somewhere between the second gain stage and the output terminals. The alternatives are presented in Table 7.7 .

Table 7.7 Comparison of Follower Alternatives

	AC coupling before followers	**AC coupling after followers**
Cathode followers	High V_{ak} (and therefore P_a) or subsidiary supply needed. Needs bias servo to force cathodes to 0 V	Easy biassing, ideal V_{ak} , but elevated heaters and large coupling capacitors needed
Source followers	As above, but subsidiary supply could be much lower voltage	As above, but no heater supply

As Table 7.7 shows, the advantage of AC coupling before the followers is a smaller coupling capacitor, but this is won at considerable expense. Traditionally, the expense would have been considered worthwhile because a 2.2-μF capacitor would have been large, expensive and polyethylene terephthalate. However, low voltage metallised polypropylene capacitors are now readily available, and a batch of 2.2 μF 400 V Vishay MKP1840s measured so well (low D , low ESR, constancy of C with f and high ratio of imaginary-to-real capacitance) that the author had no hesitation in choosing to AC couple after the followers.

From the point of view of sensitivity to induced interference, the ideal source resistance is zero because it equalises the (opposing) interference voltages developed by equal induced interference currents in each cable leg, enabling them to be nulled, irrespective of whether the source is balanced or not. Thus, the increased source impedance (≈1.5 kΩ as opposed to ≈100 Ω) at line frequency (50 Hz or 60 Hz) due to the series reactance of a 2.2 μF capacitor degrades hum rejection by ≈24 dB (1,500 Ω/100 Ω) in an unbalanced system. For a balanced system, any degradation is due to imbalance impedance, not absolute, so if we take the extreme of one 2.2 μF coupling capacitor being +5% and the other −5%, we have an imbalance reactance of only 76 Ω, implying that AC coupling at the output of the followers will not degrade interference suppression. Note that this argument assumes low output resistance from the followers at all frequencies.

Source followers have very slightly better distortion than cathode followers, but their main advantage is the avoidance of elevated heaters, so the author chose source followers.

Deciding the HT Voltage

The Analog Devices data sheet tells us that we need to operate the SSM2210 at 3 mA to minimise voltage noise, and this single requirement pretty well determines the rest of the entire design. We already know that we need to minimise valve Miller capacitance, so we prefer the E88CC over the 7N7, and we know that for the E88CC, V_{ak} ≈90 V is an operating point with good linearity. If we set R_L =33 kΩ, then 3 mA will drop 99 V across it, and we will need an HT voltage of ≈195 V (189 V plus an allowance for V_{gk}).

195 V is a very significant voltage because it is comfortably within the bounds of the statistical regulator introduced in Chapter 5 . This is important because a cascode's high r_a means that it has no rejection of power supply noise. Choosing a BJT lower device rather than a JFET increased the gain, and therefore the signal at the anode, reducing the effect of power supply noise, but the biggest improvement (>40 dB) came from choosing a differential topology. The combination of a simple, extremely quiet HT supply and gain stages having good common-mode rejection ratio (CMRR) enables all the stages to share a common supply, greatly simplifying the design.

We now have enough information to be able to draw a full circuit diagram (see Figure 7.47).

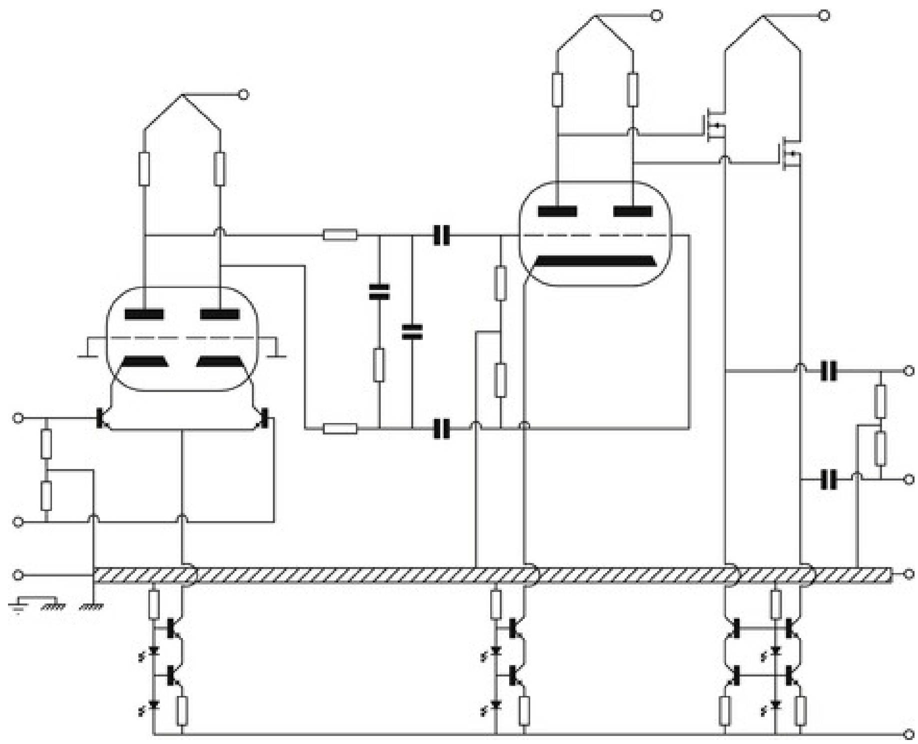


Figure 7.47 Circuit diagram of balanced hybrid RIAA stage.

Input Stage BJT Miller Capacitance

It was mentioned earlier that the semiconductor in a hybrid cascode suffers large and signal-dependent Miller capacitance unless the valve's grid is elevated, but that doing so risks injecting noise directly into the input stage. There are two reasons why we could safely elevate the differential pair's grids:

- Any injected noise would be common mode and rejected by the differential pair's CMRR of >40 dB.
- The statistical regulator produces very low noise and hum, minimising injected noise.

Having deemed it safe to elevate the grids, we must decide whether we need to, and if so, by how much.
Looking up into the cathode, we see:

$$r_k = \frac{R_L + r_a}{\mu + 1} = \frac{33 + 6}{30 + 1} = 1.3\ \mathrm{k}\Omega$$

At 3 mA, the transistor has g_m =105 mA/V, so the gain to the cathode is ≈130.

A good fit to the published SSM2210 curves of C_{cb} against V_{cb} can be obtained using:

$$C_{cb} = 28.206\, V_{cb}^{-0.3058}$$

Note that this equation has no physical significance – it just provides a good fit. Knowing C_{cb} and gain to the cathode, we can calculate Miller capacitance. From the E88CC curves, if $V_a = 90$ V and $I_a = 3$ mA, $V_{gk} \approx -2.6$ V, so if the grid is not elevated, $V_{cb} = 2.6$ V, resulting in $C_{in} = 2{,}760$ pF. Both halves of the differential pair have this capacitance to ground, but the cartridge is connected between the bases, so it sees these two capacitances in series, resulting in 1,380 pF seen by the cartridge. In combination with the DL103's measured 62 µH coil inductance, this results in a resonant frequency of 540 kHz – well above the audio band, and certainly not low enough to peak the audible response.

Knowing that we will use the statistical regulator, we could conveniently elevate the valves' grids in 5.6 V steps. If we elevate them to 11.2 V (two Zener drops), V_{cb} rises to 13.8 V and the cartridge sees 830 pF, which is 60% of the unelevated value, and a useful but not compelling reduction.

VCE and BJT Linearity

We know that if we don't elevate the triodes' grids, the collectors of the SSM2210 will be held at +2.6 V. Since their bases are at 0 V, their emitters are at −0.7 V, and $V_{CE} = 3.3$ V. Small-signal transistors such as the SSM2210 achieve constant current output characteristics at $I_C = 3$ mA once $V_{CE} > 100$ mV, and 1.3 kΩ is a comparatively shallow loadline, so we should expect to be able to swing 2 V_{pk-pk} without clipping at each collector. The gain of the E88CC section of the cascode would translate this to 100 V_{pk-pk} at each anode, or 71 V_{RMS} between the anodes, 37 dB higher than the nominal level at this point. In short, $V_{CE} = 3.3$ V is perfectly adequate, there's no need to increase it, and we can justifiably leave the triodes' grids at 0 V.

In practice, the E88CC draws grid current before the SSM2210 runs out of collector swing, further reinforcing the previous argument. Nevertheless, on test, with the triodes' grids held at 0 V, the stage managed a very creditable maximum output of 85.61 V_{RMS} between the anodes at 3.2% pure third harmonic distortion, just before grid current introduced even harmonic distortion (see Figure 7.48).

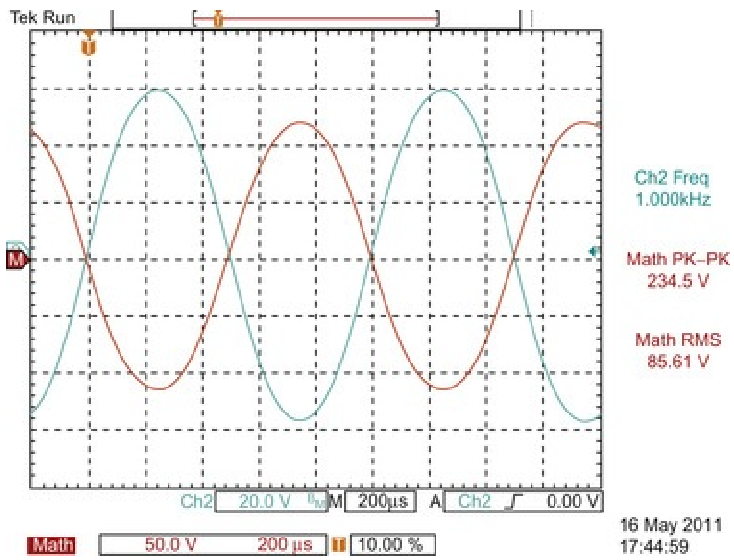


Figure 7.48 Maximum output swing of differential hybrid cascode stage.

Backing off to 1% distortion allowed 48.17 V $_{RMS}$ between the anodes with an input of 13.76 mV $_{RMS}$, corresponding to a 33 dB overload margin and a gain of 3,500.

Input Resistance and Bias Current

The individual amplitude against frequency response graph that comes with each DL103 shows that it was tested into a load of 1 kΩ, so this is the input resistance we should aim for. This makes sense from noise considerations as the potential divider formed by a source resistance of 40 Ω and load resistance of 1 kΩ causes an almost negligible S/N degradation of 0.34 dB, whereas the 100 Ω or greater recommended load resistance in the DL103 pamphlet would cause a ruinous 2.9 dB degradation at 100 Ω.

Unlike a valve or JFET, a BJT draws a significant bias current due to its forward biassed PN junction, and its small-signal input resistance h_{ie} loads the cartridge. The SSM2210 data sheet includes a graph of h_{ie} against I_C, and at $I_C = 3$ mA, $h_{ie} = 4.4$ kΩ. We need a DC path to 0 V from each base to allow base bias current to flow, so this resistance will be in parallel with h_{ie}. Note that because the transistors are matched and the cartridge is connected between their bases, no current flows through the cartridge.

If we want the cartridge to see 1 kΩ:

$$R_{\text{base bias}} = \frac{1}{(1/1{,}000\ \Omega) - (1/4{,}400\ \Omega)} = 1{,}130\ \Omega$$

But this is the total resistance formed by the two base bias resistors in series, so each one needs to be half this value at 565 Ω, and the E24 standard value of 560 Ω will be fine.

But the base bias current flows through these resistors, and although their exact value is not critical, they should be matched to avoid generating an offset voltage that would be multiplied by the (×3,500) gain of the amplifier. Thus, 562 Ω 0.1% tolerance is a better choice.

Input Stage Noise

Although not explicitly stated, the previous noise calculations assumed a single-ended input stage with a single source of voltage and current noise, but each transistor in the differential pair is a noise source.

We saw earlier that with a source resistance of 40 Ω, the current noise source was negligible in comparison to the voltage source, so we need only consider the two voltage noise sources which are in series and equal. The signal voltage remains the same, and we add noise powers, resulting in the differential pair having a signal-to-noise ratio 3 dB worse than the equivalent single-ended stage. This is a deliberate design choice; we have bought a >40 dB hum reduction due to the differential pair's CMRR at the expense of a 3 dB rise in random noise.

To enable direct comparison of theory and measurement, noise was measured at the output of the input stage before RIAA. With the input terminated by a 43 Ω resistor (to simulate 40 Ω DL103 source resistance plus 3 Ω arm and cable resistance), noise was measured at −61.5 dBu ± 0.5 dB (22 Hz to 22 kHz bandwidth, RMS rectifier). Referred to the 1 V_{RMS} signal at this point, this is equivalent to −63.7 dB, and adding the 3.4 dB RIAA advantage, we have an entirely respectable S/N ratio of 67 dB. Given that the single-ended S/N ratio using the BJT was predicted to be 70 dB, and that this was expected to be degraded by 3 dB by the differential pair to 67 dB, the agreement is excellent, confirming that practical semiconductor noise matches manufacturers' data sheets sufficiently well to enable reliable noise predictions over the audio bandwidth.

Summarising, whilst the noise performance might not be quite at the practical limit, it is satisfyingly close, and a >40 dB rejection of hum pick-up on the cabling from the cartridge should render any reasonable turntable entirely hum-free, so the author considers this to be a very worthwhile trade. Note that magnetic fields coupled directly into cartridge coils from adjacent motors or transformers are differential mode and therefore cannot be rejected.

RIAA Calculations

The equations we need are the Lipshitz equations we saw earlier:

$$R_1C_1 = 2187\ \mu s$$
$$R_1C_2 = 750\ \mu s$$
$$R_2C_1 = 318\ \mu s$$
$$\frac{C_1}{C_2} = 2.916$$

As with split equalisation, before we can apply these equations, we must first determine our source resistance, load resistance and load capacitance.

We will treat the problem as single-ended, and as values are determined, convert them to balanced values.

To a first approximation, the output resistance of a cascode is equal to its load resistance. Strictly, we need to determine the triode's r_a:

$$r'_a = r_a + (\mu + 1) \cdot R_k$$

Thus, we first need to find R_k, which for this hybrid cascode is the small-signal resistance seen looking into the SSM2210's collector, $1/h_{oe}$. Fortunately, the SSM2210 data sheet gives a graph of small-signal output conductance (h_{oe}) in terms of µA/V against I_C, and at $I_C = 3$ mA, so we simply invert the value from the graph and find that the small-signal resistance seen looking into the collector is approximately 18 kΩ at 3 mA. At the chosen operating point, $\mu \approx 30$, so:

$$r'_a = r_a + (\mu + 1) \cdot R_k = 6 + (30 + 1) \times 18 \simeq 550\ \text{k}\Omega$$

In parallel with the 33 kΩ load resistance, this becomes 31.13 kΩ. Note that unlike the split equalisation examples seen earlier, it is R_L that dominates output resistance.

The input resistance of the following stage is equal to its grid-leak resistance, which we generally arbitrarily set to 1 MΩ, but the author found that 1M2 gave more convenient capacitor values. Note that although 1M2 exceeds the Mullard data sheet's $R_{gk(max)}$ value of 1 MΩ, this is permissible because anode current runaway is prevented by the differential pair's CCS.

There seemed no reason to choose different DC conditions for the second stage, so with $R_L = 33$ kΩ and $I_a = 3$ mA, $A = 25$. From the Mullard E88CC data sheet, $C_{ag} = 1.4$ pF, so:

$$C_{\text{Miller}} = (1 + A) \cdot C_{ag} = (1 + 25) \times 1.4 = 36.4\ \text{pF}$$

This is in parallel with $C_{in} = 3.3$ pF, making a round 40 pF.

The four Lipshitz equations allow us to start anywhere, but a good place is by setting the capacitance across the input of the second stage. Remembering that we are quite restricted in capacitor values, the author chose 1 nF//150 pF between the two grids, which is equivalent to 2,300 pF from each grid to ground. The 40 pF input capacitance of each valve is in parallel with this, so the final value for the purposes of calculation is 2,340 pF. In Lipshitz's notation, this is C_1, so:

$$C_1 = 2.916 \times C_2 = 2.916 \times 2{,}340 \times 10^{-12} = 6.823 \text{ nF}$$

But this is the value of capacitance that would be needed for an unbalanced network, and we want balanced, so we divide by 2 to give 3.411 nF.

Using the balanced value of C_1, we find the balanced value of R_2 directly:

$$R_2 = \frac{318\ \mu\text{s}}{C_1} = \frac{318 \times 10^{-6}}{3{,}411 \times 10^{-12}} = 93.22 \text{ k}\Omega$$

Finally, we need R_1:

$$R_1 = \frac{750\ \mu\text{s}}{C_2} = \frac{750 \times 10^{-6}}{2{,}340 \times 10^{-12}} = 320.55 \text{ k}\Omega$$

The calculated value of R_1 is the effective series resistance seen in each leg by the equalisation network, which has the 1M2 grid-leak in parallel, so we must account for this:

$$R_1' = \frac{1}{(1/R_1) - (1/R_{\text{grid-leak}})} = \frac{1}{(1/320.55) - (1/1{,}200)} = 437.39 \text{ k}\Omega$$

But this pure series resistance includes the 31.13 kΩ output resistance of the preceding stage, so the actual series resistor we need in each leg is:

$$R_1'' = 437.39 \text{ k}\Omega - 31.13 \text{ k}\Omega = 406.25 \text{ k}\Omega$$

Thus, we have calculated all the component values for our RIAA network, and if we were able to DC couple to the second stage, these are the values we would use. However, we know that we need to AC couple to the 1M2 grid-leak resistors, and this must have a slight effect. The solution, as always, is to drop our calculated RIAA equaliser values, coupling capacitance, source resistance, load resistance and load capacitance into a CAD package together with the RIAA equation and iteratively adjust them until we get a flat response.

The author had some 56 nF 500 V Soviet PTFE capacitors that he wanted to use, but you might not have any, so Table 7.8 also shows values calculated for 100 nF coupling capacitors.

Table 7.8 Practical RIAA Equaliser Component Values

Table 7.5 Practical RIAA Equaliser Component Values

	100 nF	56 nF
R_1	412 k 0.1%+750 Ω	412 k 0.1%+5 k 11%
R_2	93k1 0.1%+160 Ω	
C_1	3n3 1%+110 pF 1%	
C_2	1 nF 1%+150 pF 1%	

As can be seen, only R_1 changed significantly with the inclusion of the coupling capacitor.

The Source Followers

The author chose FQP1N50P because he had previously bought lots of them and because earlier tests showed that they produced lower distortion than 6C45π cathode followers. We don’t expect to drive 20 m cables, so 6 mA of quiescent current per follower will be fine.

The Constant Current Sinks

The CCSs for the differential pairs’ tails are standard, using BC549C (the ‘C’ variant has a guaranteed h_{fe} >420) to maximise their output resistance.

The source follower CCSs must dissipate 0.6 W with 100 V across them, and in the absence of CRT video driver transistors (lower output capacitance), MJE340 will have to do for the outer device. These CCSs are identical to those in the Bulwer-Lytton power amplifier, and share the LED reference chain not simply for economy, but also because it renders any noise due to the reference chain common mode, which can therefore be rejected by the next stage. Given the voltage at the followers’ sources, the CCSs could have returned their current to 0 V, rather than −15 V. However, not only would this have dirtied the 0 V, but the 14 mA required by the LEDs would have been supplied by the HT, greatly increasing required HT power.

We can now draw a full circuit diagram with component values (see Figure 7.49).

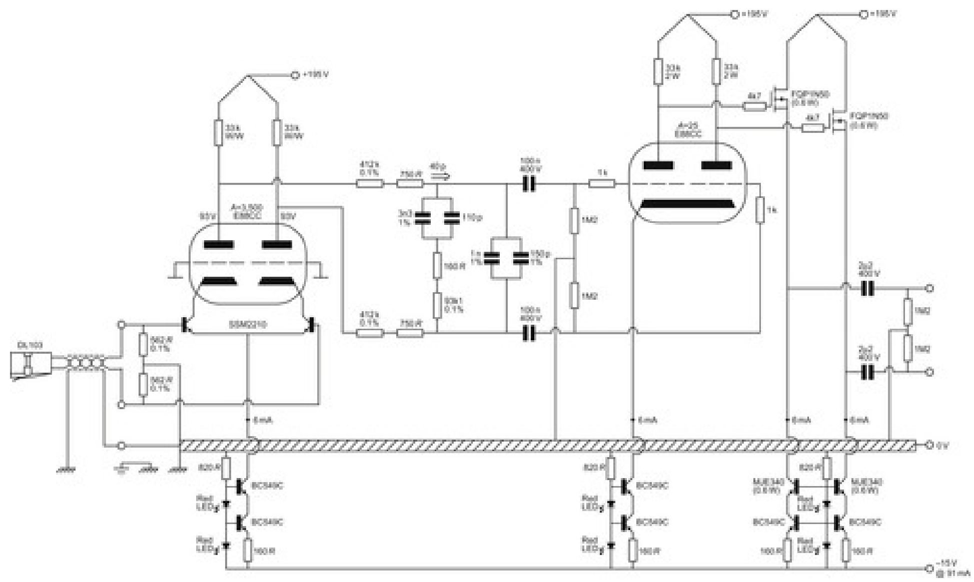


Figure 7.49 Circuit diagram of balanced hybrid RIAA stage with component values.

The output coupling capacitors have 1M2 resistors to ground not because such a precise value is required, but because it is convenient to use the same value as the second differential pair's grid-leak resistors.

The HT Supply

We stated earlier that we needed 195 V, and we find that we need 48 mA of current, so we have achieved our design aim of using significantly less HT power than the EC8010 RIAA stage. We can use a single statistical regulator to power all the stages because (being balanced) each stage draws negligible signal current from the HT supply, minimising the signal voltage developed across its output impedance. To minimise interference entering the RIAA stage, the statistical regulator should be within the RIAA stage, not in the power supply. Further, for optimum LF RF rejection, the 10 μF 400 V polypropylene capacitor across the Zener chain should have a Kelvin connection (these capacitors are available from Suppression Devices of Clitheroe, UK) (see Figure 7.50).

Figure 7.50 Four wire Kelvin connected capacitor.

Total Gain and Channel Balance

As configured, the RIAA stage has 5.1 dB gain in hand to match the 4 V_{RMS} balanced digital standard from 0.3 mV_{RMS}, and this is to allow for low-amplitude recordings.

Cartridges do not always produce matched outputs even when perfectly aligned. If required, the current programming resistor in each input stage's CCS can be finely adjusted to correct channel balance. This works because $g_m = 35 \cdot I_C$ and $A = g_m \cdot R_L$, so a change in I_C directly changes gain. If such an adjustment is implemented, use fixed resistors rather than a trimmer potentiometer to avoid contact noise degrading the RIAA stage's S/N ratio.

Summary

The balanced hybrid RIAA stage produces almost three times as much distortion as the EC8010 RIAA stage and it is third harmonic rather than second, but it costs rather less than a third of the price and has a comparable S/N ratio.

Nothing beats a well-balanced transformer for rejecting cable interference −100 dB rejection is commonly achieved. Conversely, achieving >40 dB CMRR from a practical differential pair over the 20 Hz to 20 kHz audio bandwidth requires attention to construction detail, and the rejection only holds for small interference voltages. In short, the balanced hybrid RIAA stage cannot rescue poor screening or earthing strategy, but it can put the final polish on good practice and allows the DL103 to give of its best.

The Bulwer-Lytton power amplifier in Chapter 6 included a volume control but needed no more because it used its associated Digital to Analogue Converter (DACs') digital input selector to select between computer server, CD player and broadcast receiver/recorder. However, we may need rather more analogue flexibility, so in this chapter we will investigate the hotch-potch of functions that have been traditionally allocated the term 'pre-amplifier'.

[1] Radiolympia view. Wireless World 1949; November: 438.
[2] Self, D, *Small signal audio design* . (2010)Focal ; 48–51 .
[3] James EJ. Simple tone control circuit. Bass and treble, cut and lift. Wireless World 1949; February: 48–50.
[4] Baxandall PJ. Negative-feedback tone control. Wireless World 1952; October: 402–5.
[5] Self, D, *Small signal audio design* . (2010)Focal ; pp. 259–67 .
[6] Lipshitz, SP, On RIAA equalization networks , *J Audio Eng Soc* **27** (June (6)) (1979) 458 – P481 .
[7] Terman, FE, *Electronic and radio engineering* . 4th ed. (1955)McGraw-Hill ; 438 .
[8] Van der Ziel, A, *Noise: sources, characterization, measurement* . (1970)Prentice-Hall .
[9] Few stones unturned. An article by Herbert Reichert featuring Arthur Loesch's RIAA MC stage. Sound Practices Early 1993; 23–28.
[10] Wright A. The tube pre-amp cookbook. 2nd ed; 1997.
[11] Yaniger S. His master's noise: a thoroughly modern tube phono preamp. Available at the 'Articles' section of diyAudio.com
[12] Audio dual matched NPN transistor SSM2210. Analog devices data sheet; 2003.

References

Recommended Further Reading

- Self, D, *Small signal audio design* . (2010)Focal Oxford ; Even if you think ‘solid state’ is pronounced ‘squalid state’ you owe it to yourself to own this book. You might not agree with all the author’s views but he doesn’t shirk from providing supporting evidence, and as he’s worked on both sides of the design fence (recording and reproduction), you have to take his opinions seriously .
- Wright A. The tube pre-amp cookbook. 2nd ed; 1997. This self-published work lacks the production quality of a professional publication, but at least you know all the money is going to the author. You might not agree with Allen’s dogmatic approach but his designs have stood the test of time and are ignored at your peril.
- Vogel, B, *The sound of silence* . (2008)Springer Berlin ; This entire book is about how to design low-noise RIAA stages, either with valves or solid state. Its strength is its thorough mathematical treatment of the problems illustrated by copious worked examples. Sadly, it’s expensive and English is not the author’s mother tongue .
- Gayford, M, *Microphone engineering handbook* . (1994)Focal Oxford ; The two hardest analogue audio challenges are microphone amplifiers and RIAA stages because both require low noise and wide dynamic range, so there is much to be learned from studying microphones and their amplifiers. Be warned that this expensive book is far more about microphones than their amplifiers, but it does give valuable insights .

Appendix

Valve Data

Obtaining detailed valve data used to be extremely difficult, and the author considered himself to be very fortunate in having a large collection of printed Mullard valve data sheets. Fortunately, there are now many excellent sources of scanned valve data sheets, and full manufacturers' data sheets can now be found at various sites, but the definitive site has to be: http://tubedata.org.

Frank's site is a service to the world, and the author would be lost without it.

Sadly, website addresses change rapidly (in terms of book lifetimes), but many valve vendors provide updated links to useful sites. Keep searching.

European Pro-Electron Valve Codes

The Pro-Electron system of valve codes gives significant information about a valve.

	Heater Type (1st Letter)	Valve Type (2nd, and Subsequent Letters)		Base Type (1st Digit of Serial Number)
A	4 V	Small-signal diode	**1**	Use 2nd digit
B	180 mA	Double small-signal diode	**2**	B8B
C	200 mA	Small-signal triode	**3**	Octal
D	0.5–1.5 V	Power triode	**4**	B8A
E	6.3 V	Small-signal tetrode	**5**	B9D
F	13 V	Small-signal pentode	**8**	B9A
G	5 V		**9**	B7G
H	150 mA	Hexode or heptode		
K	2 V	Heptode or octode		
L		Power tetrode or pentode		
M		Fluorescent indicator		
N		Thyratron		
P	300 mA			
Q		Nonode		
X		Gas-filled rectifier		
Y		Half-wave rectifier		
Z		Full-wave rectifier		

As an example, the ECC88 has a 6.3 V heater, with two small-signal triodes on a B9A base. European manufacturers such as Mullard/Philips reversed the order of digits and letters after the heater letter to signify a Special Quality type, such as E88CC, so although this is a superior version of ECC88 and a plug-in replacement, its heater is specified as 365 mA, although only Mullard/Philips/Amperex adhered to this, and most manufacturers actually used

300 mA heaters. Brimar specified 365 mA, but the author tested 10 Brimar ECC88/6DJ8s, and found that they were all 300 mA. Sadly, coding reversal usually implies entirely different valves (EL81≠E81L, EF80≠E80F).

It would be nice to say that the Pro-Electron coding system was rigidly observed, but it's rather like French verbs – there are some irregular types, and the 6.3 V heater, dual triode ECC99 valve actually uses a B9A base rather than the B7G implied by its code.

It was common for a single company to use different brand names in different markets. Thus, Osram or Marconi-Osram was used for the British Commonwealth market and GEC for the rest of the world, yet the valves and their type numbers were identical. STC was the British arm of Western Electric, but used different type numbers for industrial valves, so the WE437A became the 3A/167M, and the brand 'Brimar' (*Bri*tish made for the American *mar*ket) was used for consumer valves.

The EF86 was a popular small-signal pentode and many companies made equivalents. Mullard's Special Quality plug-in equivalent was the M8195, whereas GEC's very quiet plug-in equivalent used the British military designation CV4085 (CV=Common Valve – common across the military services). Although Telefunken made a long-life plug-in equivalent (EF806s), they also made long-life electrical equivalents with different pin-outs: EF804, EF804s.

Since the valves in any piece of equipment were consumable items, claiming that a particular valve had no equivalents ensured that the manufacturer had also cornered the replacement market. Thus, most manufacturers used proprietary codes in an attempt to disguise their valves, and some deliberately invented proprietary bases (such as the notorious Mazda octal) to prevent electrical equivalents being inserted. The significance of this market manipulation is that lesser-known equivalents should be cheaper than the more generic code.

Ediswan/Mazda applied their proprietary code in a startlingly Humpty Dumpty fashion ("When *I* use a word, it means just what I choose it to mean – neither more nor less"), sometimes they conformed to their code, sometimes they didn't.

	Heater Type 1st Number		**Valve Type 1st and Subsequent Letters**
6	6.3 V	**C**	Frequency changer
10	100 mA	**D**	Small-signal diode
20	200 mA	**F**	Small-signal pentode
30	300 mA	**K**	Thyratron
		L	Triode
		M	Fluorescent indicator
		P	Pentode or beam tetrode
			Half-wave rectifier

		U	Half-wave rectifier
		UU	Full-wave rectifier

Thus, the Mazda 6F13 is a small-signal pentode with a 6.3 V heater, and a 6/30L2 is a small-signal triode having a 6.3 V 300 mA heater, but you would need the data sheet or an inspired guess to realise that it is actually a *double* triode.

American Valve Codes

Like the Pro-Electron system, the American RETMA number/letter system sought to introduce some logic to consumer valve codes. Unfortunately, this utopian ideal was soon overturned, and a non-systematic four digit industrial code was used in tandem. Nevertheless, the number/letter system offers some clues as to a valve's internals.

1st Digit	Letters	2nd Digit
Approximate heater voltage (however, 7 or 14 means Loctal base)	Purely random – relates to individual valve design	Number of electrodes (including heater and metal envelope)

Example: 6SN7 has a 6.3 V heater and seven electrodes (two individual triodes plus one heater adds up to seven). Unfortunately, this coding was not always strictly obeyed – the 6BQ5 is actually a pentode (direct equivalent to EL84).
In addition, the following American Octal suffixes were also used:

Suffix	Meaning	Comment
None	Metal envelope	Introduced in 1935, although the envelope provides a useful screen, envelope outgassing and the consequent grid gas current worsens noise and distortion performance.
G	Glass	Early valves tended to use the ST14 (Shouldered Tube) envelope that looks like a soft drink bottle.
GT	Glass, tubular	Later glass envelopes were of a shorter, tubular construction.
GT/G	Interchangeable	Usable with equipment specified for either G or GT.

A suffix following an oblique signified a development of a basic type, and they were stated to be reverse-compatible with the original type. Thus, for the 6SN7:

	6SN7GT	6SN7GTA	6SN7GTB
$P_{a(max)}$ per triode	3.5 W[a]	5 W	
$P_{a(max)}$ total	5 W	7.5 W	
$V_{a(max)}$	300	450	
Controlled warm-up time	–	–	Yes

[a]Depends on date of manufacture and manufacturer. Early valves tend to be 2.5 W, later ones 3.5 W.

The increased voltage rating was usually achieved by additional perforations in the supporting micas to lengthen leakage paths between anode and grid. The increased power rating was usually achieved by changing the anode from single to double sutures which increased the radiating area and allowed better cooling.
For many valves, the 'B' suffix indicates controlled warm-up time. Ideally,

series heater chains should be driven by constant current sources, but in practice they were more likely to be connected directly across the mains supply (a constant voltage source) via either a tapped resistor or varistor.
Since most valves were intended for use in radios, capacitances and screening were important. As an alternative to the metal envelope, early octal glass envelope valves added a pressed metal ring (often connected to pin 1) around their base that acted as a crude guard to reduce capacitances to other components. The Loctal and B9G guard rings added a flange under the base that encircled valve pins and was therefore much more effective because it reduced inter-electrode leakage and capacitance.

Pin Connections

Unlike ICs, valves are numbered, viewed from *underneath* counting clockwise.

Thermionic Emission

The Richardson/Dushmann equation for emitted cathode current (A) per unit area (m^2) is:

$$I = AT^2 \, e^{-q_e \Phi / kT}$$

where

$A = \left(\dfrac{4\pi m_e q_e k^2}{h^3}\right) \geq 1.204 \times 10^6 \ \mathrm{A/m^2/K^2}$, but see note

Φ=work function of the cathode surface (≈4.55 eV for tungsten)

k=Boltzmann's constant (≈1.381×10^{-23} J/K)

T=absolute temperature (°C +273.16)

e=base of natural logarithms (≈2.718)

m_e=electron rest mass (≈9.109×10^{-31} kg)

q_e=electronic charge (≈1.602×10^{-19} C)

h=Planck's constant (≈6.626×10^{-34} J s).

NB: Although the theoretical value for A≈1.202×10^6, Spangenberg [1] noted that the experimental value was typically half this value, but Van der Ziel [2] explained this discrepancy by pointing out that the problem lay not in a deviation from the theoretical value of “ *A*”, but in the temperature coefficient of the work function “Φ” and recommended using the theoretical equation for “ *A*” as a means of determining the true value of “Φ” at the temperature of interest.

Standard Component Values

The following series of components covers one decade; other values are obtained by multiplying or dividing by factors of 10.

E6											
1	1.5	2.2	3.3	4.7	6.8						
E12											
1	1.2	1.5	1.8	2.2	2.7	3.3	3.9	4.7	5.6	6.8	8.2
E24											
1	1.1	1.2	1.3	1.5	1.6	1.8	2	2.2	2.4	2.7	3
3.3	3.6	3.9	4.3	4.7	5.1	5.6	6.2	6.8	7.5	8.2	9.1
E96											
1	1.02	1.05	1.07	**1.1**	1.13	1.15	1.18	1.21	1.24	1.27	**1.3**
1.33	1.37	1.4	1.43	1.47	**1.5**	1.54	1.58	1.62	1.65	1.69	1.74
1.78	1.82	1.87	1.91	1.96	**2**	2.05	2.1	2.15	2.21	2.26	2.32
2.37	2.43	2.49	2.55	2.61	2.67	2.74	2.8	2.87	2.94	3.01	3.09
3.16	3.24	3.32	3.4	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	5.9	6.04	6.19	6.34	6.49	6.65	6.81	6.98	7.15	7.32
7.5	7.68	7.87	8.06	8.25	8.45	8.66	8.87	9.09	9.31	9.53	9.76
E96 values in bold are common to the E24 series.											

ResCalc

We frequently need a value that is *not* a standard value. The problem is to find the best combination of two standard values that would give the required value with minimum error. Do an Internet search (or go directly to Pete Millett’s or Duncan Munro’s site), and you will find that Mark Lovell and the author wrote a piece of freeware called ResCalc that solves the problem. Choose your available resistor series, tell ResCalc the required value, and it gives you the question to the answer. Even nicer, it displays the resistor with the correct colour code.

Resistor Colour Code

Most resistors are marked with their value in the form of a colour code consisting of four or six concentric bands of paint on the body of the component which are read from left to right (see Figure 1).

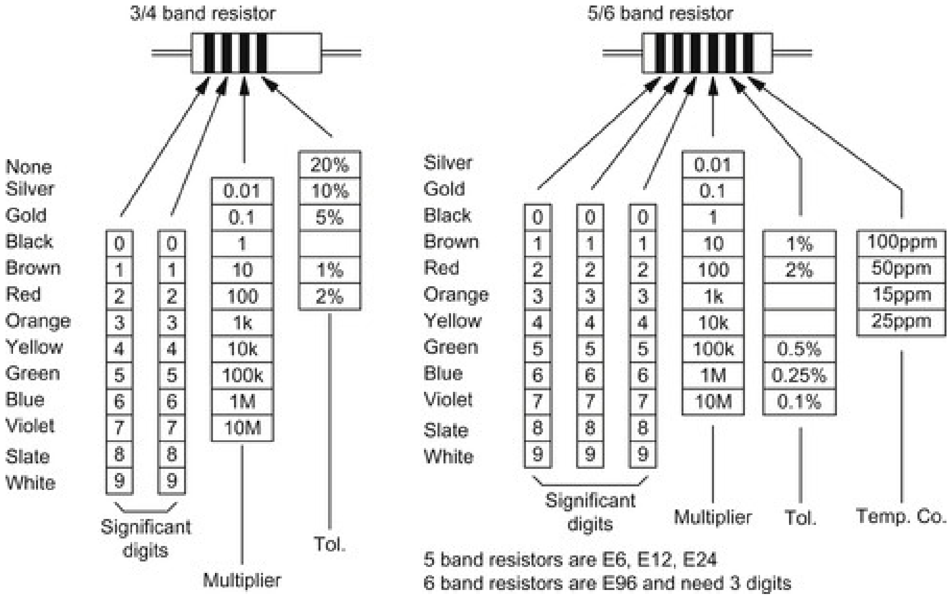


Figure 1 Resistor colour codes.

Four-Band Resistors

The first two bands denote the two significant digits of the value.

The third band is the *multiplier*, whose value is 10^{x}, where x is the value of the band. Gold used as a multiplier means 10^{-1}=0.1, and silver means 10^{-2}=0.01.

The fourth band is the tolerance, which will usually be 1% (brown) or perhaps 2% (red). On very old equipment (or carbon resistors), you will see gold (5%) and silver (10%); the use of these colours as tolerances dates from the days when 5% was considered to be close tolerance. If there is no fourth band, the tolerance is 20%.

Six-Band Resistors

The first three bands denote the significant digits, and the fourth band is the multiplier.

The fifth band is the tolerance; note that six-band resistors imply greater precision and so 5%, or worse, tolerance will not be seen.

Examples

Yellow, violet, yellow, red=470 kΩ 2%

Yellow, violet, black, orange, brown, red=470 kΩ 1% 50 ppm

Red, red, red, red=2.2 kΩ 2%

Red, red, black, brown, brown, red=2.2 kΩ 1% 50 ppm

Brown, black, black, red=10 Ω 2%

Brown, black, black, gold, brown, red=10 Ω 1% 50 ppm

Note that because the six-band resistors have an extra significant digit, their multiplier is always one level lower than for the same value in a four-band component.

Sometimes it can be difficult to decide which end of the resistor is which, and the value makes sense either way round:

Brown, orange, yellow, red=130 kΩ, 2%, but read the other way round=24 kΩ 1%

If in doubt, measure the resistor with a digital multimeter; it is far easier to change the component now, than when it has been soldered into place.

Plastic Capacitor Coding

The first of the three letters refers to the plates:

F=foil

M=metallised

The second letter ‘K’ is for Kunststoff – German for plastic.

The last letter denotes the dielectric:

S=PS, polystyrene

P=PP, polypropylene

C=PC, polycarbonate

T=PETP, polyethylene terephthalate

N=PEN, polyethylene naphthalate

I=PPS, polyphenylene sulphide

Thus, FKP is foil polypropylene, whilst MKT is metallised polyethylene terephthalate.

Many small (and particularly ceramic) capacitors use pF as their base units. Some colourful European capacitors used the resistor colour code to designate value, but it is now more common to use a three-digit code whereby the first two significant digits are followed by a multiplier signifying the number of zeros after the first two digits. Thus, 103=10,000 pF=10 nF.

Cable

Coaxial Cable Capacitance

Audio coaxial cable has a typical capacitance of ≈120 pF/m, but RF coaxial cable tends to have a lower capacitance (perhaps ≈70 pF/m) dictated by the required characteristic impedance. The capacitance per unit length of a coaxial capacitor is:

$$C = \varepsilon_0 \varepsilon_r \log \frac{d}{D}$$

where

ε_0=permittivity of free space (≈8.854×10^{-12} F/m)

ε_r=relative permittivity (compared to free space)

d=inner diameter

D=outer diameter.

Unfortunately, the logarithmic relationship implies that low-capacitance cable requires a very small inner diameter. Oscilloscope probe leads need low capacitance and solve the resulting fragility problem using a steel core plated with copper then silver (to reduce resistance) and have a capacitance of <2 pF/m. The cable is made by the Surprenant company and fits all the standard miniature coax connectors such as SMA, SMB, MCX.

American Wire Gauge (AWG)

The author found the following equation in a post by psychokids at diyAudio and cannot resist repeating it:

$$D = 0.005 \times (\sqrt[39]{92})^{(n-36)}$$

where

D=diameter in inches

n=AWG number.

Quite what was in the mind of the originator of this equation is unknown, but it works.

Square Wave Sag and Low Frequency $f_{-3\ \mathrm{dB}}$

A square wave with LF sag is a decaying exponential, whose instantaneous voltage at any time ‘ *t*’ may be found using:

$$v = V_0 \, e^{-t/\tau}$$

Rearranging, and solving for *τ*:

$$\tau = \frac{-t}{\ln(v/V_0)}$$

where ‘ *t*’ is the time allowed for the decay across the bar top, but for a square wave with equal positive and negative durations, it is half of the periodic time *T*:

$$T = 2t$$

But *T* is the reciprocal of frequency:

$$f = \frac{1}{T}$$

So:

$$t = \frac{1}{2f}$$

Substituting:

$$\tau = \frac{-1}{sf \, \ln(v/V_0)}$$

From the frequency domain, a CR filter has a −3 dB cut-off frequency:

$$f_{-3\ \mathrm{dB}} = \frac{1}{2\pi\,\mathrm{CR}}$$

But CR= *τ* and *τ*= *L*/ *R*, so a universal equation, valid for both CR and LR, is:

$$f_{-3\ \mathrm{dB}} = \frac{1}{2\pi\tau}$$

Rearranging:

$$\tau = \frac{1}{2\pi f_{-3\,\mathrm{dB}}}$$

We now have two formulae for *τ*, which can be equated as:

$$\frac{1}{2\pi_{-3\,\mathrm{dB}}} = \frac{-1}{2f \, \ln(v/V_0)}$$

Solving for the ratio $f/f_{-3\ \mathrm{dB}}$:

$$\frac{f}{f_{-3\,\mathrm{dB}}} = \frac{-\pi}{\ln(v/V_0)}$$

Sag is the percentage of peak-to-peak level by which the horizontal bar has sagged in level. 10% sag is easily measured on an oscilloscope, and means that the level has fallen from 100% to 90%, so:

$$10\%\ \text{sag} \equiv \frac{v}{V_0} = 0.9$$

Applying 10% sag to the $f/\,f_{-3\text{ dB}}$ formula:

$$\frac{f}{f_{-3\text{ dB}}} \approx 30$$

So, 10% sag means that the applied square wave frequency is 30 times higher than $f_{-3\text{dB}}$

Sag Observed Using a Square Wave of Frequency "f" (%)	$f/\,f_{-3\text{ dB}}$
10	30
5	60
1	300

Playing 78s

You have just inherited a collection of 78s that appear to be in superb condition, some of them are original recordings of legendary performers, and you are desperate to play them. There are four main problems.

Correct Speed

Although colloquially known as '78s', referring to their speed, very early 78s were recorded on rather crude lathes and the actual recorded speed was somewhat variable. The first requirement for replay is therefore a turntable that will not only rotate at 78 rpm, but also has varispeed, so the Garrard 301 and 401, and Thorens TD124 are obvious contenders. The BBC modified the Technics SP10 direct drive turntable to give varispeed, added a pick-up arm/cartridge plus elaborate electronics, and called the whole confection an RP2/10 (ReProducer 2, version 10). Oddly, the disco-oriented Technics SL1200 (production terminated 2010) was capable of playing 78s – provoking the startling thought of nightclubs belting out Alma Cogan at 110 dB(A) from a 78.

Groove Size

The 78 has a coarse groove, and was traditionally played with a crude steel 'needle'. The LP stylus of a modern cartridge is far too small, so a dedicated large-diameter stylus is required. Traditionally, only broadcast cartridges like the Shure SC35 and Ortofon OM Pro were offered with 78 styli by their manufacturers, but the situation has now changed. Vinyl has recovered to become a niche market, and the cartridge manufacturers must respond to the demands of that niche, with the result that almost all manufacturers now offer a mono cartridge, either for mono vinyl, or with a larger tip for 78s. Grado goes one step further and offers a mono 78 cartridge with four different size tips to allow quick determination of the best compromise between worn grooves and surface damage for a given record. A final alternative could be to ask a specialist re-tipping concern to fit 78 tip to your cartridge.

Because different 78s are likely to need a different size tip, moving magnet cartridges are more suitable for playing 78s because removable styli are the norm (Audio Technica OC9 notwithstanding). Even so, the Lyra Helikon Mono *moving coil* cartridge could presumably be fitted with a 78 tip, and if you had an arm with a replaceable headshell or arm assembly, different cartridges could have different tips. You would need an awful lot of priceless 78s to justify the expense, and probably only a national library that needed to transcribe rare recordings could afford it.

Incidentally, the 78 is the only recording medium that is provably robust over decades. Magnetic tape went through a sticky patch in the 1970s with unstable binders. 'Perfect sound for ever' turned out not to be true if the CD in question had been pressed at Blackburn. The UK plant had been converted from pressing Laservision video discs to CDs, and used silver rather than aluminium as the reflective coating, which wasn't a problem, but they then had a problem with imperfect lacquer sealing at the periphery. Even that wouldn't have been too bad, except that the paper liners had an unusually high sulphur content, which promptly reacted with the silver to produce yellow silver sulphide, which isn't so reflective, and causes data errors. 'Laser rot' was nonsense. Any recording format based on videotape (CDs were originally mastered on U-matic) becomes unplayable not just because the medium deteriorates, but because working machines are unavailable. Two-inch quadruplex videotape is only 50 years old, and was in common use for almost 30 years, yet there are now very few of even the final generation machines in working order.

As a species, we are becoming increasingly sloppy about safeguarding our data. The Rosetta stone might have been difficult to read, but at least that was largely a linguistic problem, rather than deterioration of the medium.

Pick-Up Arm Mechanics

Playing a 78 drives considerable vibration into the pick-up arm, and loose bearings cause rattles and mistracking. At a more subtle level, a stylus traversing an imperfection, or speck of dust, produces a mechanical impulse and excites arm resonances, which greatly magnifies the subjective nuisance. Paradoxically, the inferior medium needs a good arm to replay it adequately – so a modern arm such as the gimballed Rega RB251 arm and its derivatives seem a minimum requirement, although unipivots can be even better at rejecting clicks and bangs.

Equalisation

Analogue Disc

It took some time before the manufacturers of 78s and LPs standardised their equalisation. The following table gives the electrical time constants used by major organisations, and therefore an indication of the likely equalisation required [3].
(All time constants are specified according to Lipshitz's notation [4]).

		Time Constants (μs)		
		t_3	t_4	t_5
78	'Standard'	–	636	–
	Decca 'ffrr'/European	–	636	25
	AES	–	400	63.6
	Pre-1954 DG	–	450	50
	BBC	–	450	25
	International	–	**450**	**50**
LP	Pre-1954 DG	1590	450	50
	Pre-1954 Decca	1590	318	50
	Columbia/EMI	1590	318	100
	European	2230	318	50
	NAB	3180	318	100
	RCA New Orthophonic	3180	318	75
	RIAA	**3180**	**318**	**75**

Errors in t_4 and t_5 cause sharp peaks and troughs in the critical mid-band, and despite popular belief, cannot possibly be corrected by tone controls or graphic equalisers.
If only the later 78s (when an international standard had been fixed) are to be played in addition to modern LPs, then t_4 need only be switchable between 318 μs and 450 μs, and t_5 between 75 μs and 50 μs.
Langford Smith [5] stated that the Decca (London) LP recording characteristic had the following response (January 1951):

Frequency	30 Hz	50 Hz	100 Hz	300 Hz	1 kHz	10 kHz	15 kHz
Level	−17.5 dB	−14 dB	−9 dB	−3 dB	0 dB	+14 dB	+16 dB

Lipshitz time constants that produce a response passing within ±0.1 dB of these points are:

τ_3=9.6 ms

τ_4=735 μs

τ_5=110 μs

τ_6=10.2 μs.

Modern Analogue 'Microgroove'

	τ_2 (μs)	τ_3 (μs)	τ_4 (μs)	τ_5 (μs)
RIAA	–	3180	318	75
IEC	7950	3180	318	75

Note that the 7,950-μs IEC time constant is *replay only*, and is a 20-Hz high-pass filter intended to remove rumble produced by turntables; see Chapter 7 for details on why to ignore it.

The RIAA record equalisation implies a 6 dB/octave rising response with frequency. If a 3.18-μs time constant is added to the head power amplifier to protect the fragile cutting head from ultrasonic energy, the Lipshitz replay equation becomes:

$$G_s = \frac{(1+\tau_4 \cdot s)\cdot(1+\tau_6 \cdot s)}{(1+\tau_3 \cdot s)\cdot(1+\tau_5 \cdot s)}$$

where

s= *jω*, *ω*=2 *πf*

τ_3=3180 μs

τ_4=318 μs

τ_5=75 μs

τ_6=3.18 μs.

Thus:

$$G_s = \frac{(1+318\times10^{-6}\times s)(1+3.18\times10^{-6}\times s)}{(1+3.18\times10^{-6}\times s)(1+75\times10^{-6}\times s)}$$

Alternatively, the following equation [6] (modified to include 3.18 μs) can be used:

$$G_\omega = 10\log\left[1+\frac{1}{(\omega\tau_4)^2}\right] - 10\log\left[1+(\omega\tau_5)^2\right] - 10\log\left[1+\frac{1}{(\omega\tau_3)^2}\right] + 10\log\left[1+(\omega\tau_6)^2\right]$$

The first (fundamental) equation requires considerable manipulation, but allows phase to be found, whereas the second is far easier to compute if only gain is required. The two equations produce exactly the same gain results.

Note that including the 3.18-µs time constant changes the extreme HF response from a 6 dB/octave low-pass filter to a final attenuation of 27.5 dB that is constant with frequency.

CD

0 µs, 15 µs

This equalisation was used on a few very early CDs. A sub-code flag told the player to apply the equalisation that is commonly implemented digitally in the oversampling filter with numerical precision, but that rather negates the intended noise advantage of pre-emphasis after the Digital to Aalogue Convertor, so precisely implemented analogue equalisation would theoretically be better.

Sourcing Components: Bargains and Dealing Directly

New Parts

The mainstream electronics distributors stock general electronic components, but do not necessarily stock specialist audio components such as large polypropylene capacitors. It is well worth shopping around for specialist components, as some stockists have imaginative pricing policies.

If you and your friends are able to club together to generate a large order it can be worth approaching manufacturers directly; after all, the worst they can do is to laugh at you. Transformers and capacitors are cottage industry components, so they can often be made specially to order, and the author has had transformers made by Sowter Transformers and capacitors by Suppression Devices. If you choose to follow this course, specify in proper engineering terms as completely as possible what you need, and remember that every additional complication adds to the finished price.

Companies at Hi-Fi shows often give a ‘show’ discount on their goods, so it may be worth timing your order to coincide with a show. They may even be prepared to negotiate a further discount on the last day of the show.

New/Second-Hand Parts

Electronics surplus shops are excellent places for picking up bargains, provided that you know what to look for. It is a good idea to take a digital multimeter (for spotting open-circuit transformers), ESR meter (for spotting faulty electrolytic capacitors), tape measure and calculator with you. If you can afford (modern) or carry (old) a specialist resistance meter that applies a high voltage (known generically as a ‘Megger’ in the UK, but ‘hi-pot tester’ in the USA), you can leave the DVM at home and also spot failed insulation (rather common).

Amateur radio and vintage audio fairs are often fruitful sources of components, but bear in mind that most of the stands are occupied by traders who circulate from one fair to another, so they are unlikely to offer many goods at ‘giveaway’ prices. Nevertheless, bargains can be had, and otherwise awkward-to-source components can be found. In addition to components, there are usually many larger items on sale – the author was delighted to pick up a scruffy Garrard 401 for £30, but there are also boat anchors. (‘Boat anchor’ is the charming amateur radio term for equipment whose main attribute is mass.) When large companies close down a site or laboratory, they often auction equipment, and this *can* be a source of carefully maintained test equipment. It is your responsibility to check the condition of what you buy, and you have no comeback afterwards. Don't get carried away at the auction, and remember that tax will be added to your bid price.

Beware of the second-hand test equipment that has not come from an entire laboratory being closed down. Many large companies undertake a rolling replacement of test equipment, but rather than upgrading an entire laboratory’s oscilloscopes at once (very expensive), one or two might be bought per year. There are now older oscilloscopes which the accountants think are surplus (probably because the engineers claimed they were worn out). However, real engineers feel that there’s no such thing as too much test equipment, so they hide the good kit and only release the faulty stuff. Thus, the equipment they intend to release is the stuff that:

- works exactly as the manufacturer intended, but is horrible;
- has a definite fault that nobody in that laboratory could/would fix;
- and worst, has a nasty little intermittent fault/feature that means nobody trusts it anymore.

In case you think that this means that second-hand equipment is invariably doom and gloom, there is the recent possibility that some fool will decide to ‘clear valuable space’ by emptying a room ‘full of junk I’ve never seen anyone use’. Sadly, the only time such fools display efficiency is when they empty store rooms of irreplaceable (but rarely used) equipment, which is then sold for peanuts; these clearances tend to be handled by well-established industrial re-sellers rather than eBay.

Nevertheless, eBay can be a useful source, and many sellers are entirely honest in their descriptions. However, the following descriptions (and their translations) may be helpful:

> *" I don't have any electronics knowledge and haven't been able to test it, but I'm sure it's excellent."*

Translation: "I am an unprincipled cheat trying to duck responsibility for the fact that the junk you're bidding for doesn't work."

> *" Rare."*

Translation: "I didn't see any others for sale at the instant I placed this item."

> *" Excellent condition."*

Translation: "Not visibly broken."

> *" Good condition for its age."*

Translation: "Ruined."

An honest and knowledgeable seller will have uploaded sufficient photographs (correctly lit and in focus) to enable you to accurately gauge the condition of their item and will be happy to answer (sensible) questions. The very best seller may require you to prove that you understand what you're buying and that you will look after it.

If you are not careful, Internet purchases can cost more than you thought. Carriage is always extra (and some eBay sellers inflate postage to reduce their auction fees). If it is a foreign purchase, once it arrives in your country, it will be liable for import duty, and you may need to insure against unreliable postal systems. This all adds to costs, so do your sums before bidding. Nevertheless, the Internet is an extremely useful world wide source for otherwise unobtainable items.

Even within your own country, second-hand goods can be expensive, because some sellers have wildly inflated ideas about the value of their goods, so bargains are often best found through your circle of friends, who are aware of your hobby. If you buy second-hand equipment from private advertisements, remember to add the cost of your return journey to inspect the goods, plus whatever cost is required to refurbish. Just like buying a second-hand vehicle, always be prepared to walk away from a deal.

References

[1] Spangenberg, K, *Vacuum tubes*. 1st ed (1948)McGraw-Hill; p. 31.

[2] van der Ziel, A, *Solid state physical electronics*. 2nd ed (1968)Prentice-Hall; p. 140.

[3] QC11 owner's manual. Acoustical Manufacturing Company; 1953–1967.

[4] Lipshitz, SP, June*On RIAA equalisation networks*. (1977)JAES; 456-81.

[5] Smith, L, *Radio designers' handbook*. 4th ed (1953)Iliffe; p. 730.

[6] Digest of Hi-Fi. Part Four: Record replay. Hi-Fi Answers; August 1990.

0-9, and Symbols

300 pF air-spaced capacitors

317 IC voltage regulator

A

volume control

C

13E1 conditions

334Z constant current sink

D

volume controls

E

E55L pentode

F

FFT

G

H

6SN7 valve family

I

J

K

L

volume controls

M

N

volume control values

O

P

volume controls

volume control plus unity-gain line driver

Q

R

volume controls

S

T

volume control

U

V

volume controls

W

Index

Z